

ISBN 978-89-294-XXXX-X-XXXX

국가R&D 데이터 품질평가와 개선방안  
(Evaluation of National R&D Data Quality and its  
Direction for Improvement)

2016. 8



# 머리말

최근 데이터가 기업과 기관 활동의 중요한 자산으로서 부각되면서, 많은 기업과 기관들이 데이터 품질관리와 품질평가의 중요성을 재인식하고 있습니다. 본 서는 2007년 데이터 품질관리체계 구축 방안의 데이터 품질관리 후속으로 국가R&D 데이터의 품질평가를 다루고 있습니다.

한국과학기술정보연구원(KISTI) NTIS센터는 과제, 인력, 성과, 시설·장비 정보 등 국가R&D를 데이터 17개 부처·청의 대표전문기관으로부터 정보연계를 통해 수집하고 있으며, 대표전문기관의 데이터 품질제고 노력에 대한 인센티브를 부여하여 위해 2008년부터 NTIS 데이터 품질평가를 지원하고 있습니다. 정보연계를 통해 수집된 데이터가 NTIS를 통해 서비스하여 실시간 연계 및 유사과제검색이 가능한 수준으로 매년 품질점검 기준 개선하여 품질평가를 하여 데이터품질을 크게 개선시키고 있습니다. 본 서는 이러한 데이터 품질평가 사례 및 개선점을 살펴보고 있습니다.

마지막으로 본 서에 수록된 내용은 한국과학기술정보연구원의 공식의견이 아님을 밝혀두고자 합니다.

2016년 9월

한국과학기술정보연구원

# 목 차

제1장 서론 .....	6
1. 데이터 품질관리 개요 .....	7
2. 데이터 품질평가 개요 .....	11
제2장 데이터 품질평가의 기존 연구와 사례 .....	55
1. 기존 연구 .....	56
2. 국내·외 데이터 품질관리 사례 .....	58
제3장 국가 R&D 데이터품질 평가 .....	18
1. 국가 R&D 데이터 품질평가 개요 .....	19
2. 국가 R&D 데이터 품질평가 효과 .....	27
제4장 국가 R&D 데이터 품질평가 개선방향 .....	74
제5장 결론 .....	74
[부록] 중·장기적 접근이 요구되는 국가 R&D 데이터 품질평가 지표 .....	77
참고문헌 .....	85



# 제1장

## 서론

---

1. 데이터 품질관리 개요
2. 데이터 품질평가 개요

---

# 1. 데이터 품질관리의 개요

광의(廣義)의 측면에서, 데이터란 조직의 전략과 목적을 달성하기 위해 구축되고, 운영되는 정보시스템과 관련된 모든 정보자료와 산출물을 의미한다. 데이터는 데이터베이스 내부에 저장되어 있는 데이터 값 이외에, 데이터 모델이나 표준 데이터와 같은 구조 정보와 문서 형태의 산출물을 포함한다. 협의(狹義)의 측면에서, 데이터는 정보 시스템에 탑재된 디지털 데이터를 의미하며 실제 품질 진단의 대상이 된다(한국데이터베이스진흥원, 2014).

데이터의 품질이란 기업과 사용자의 목적 달성을 위해 데이터와 데이터서비스를 통해 사용자의 기대수준을 지속적으로 충족시키는 것이다(Larry P. English, 1999). 한편, 데이터 품질관리란 기관이나 조직 내·외부의 정보시스템과 데이터베이스 사용자의 기대를 충족시키기 위해 지속적으로 수행하는 데이터 관리 및 개선활동을 말한다(한국데이터베이스진흥센터, 2006). 공공분야에서는 데이터베이스의 품질을 데이터베이스의 최신성, 정확성, 상호연계성 등을 확보하여 사용자에게 유용한 가치를 전달할 수 있는 수준으로 정의하고 있다. 또한, 데이터베이스 품질관리를 데이터베이스 품질을 확보하기 위한 목표 설정과 진단, 개선 등 일련의 활동을 말한다(행정안전부, 2011).

데이터 품질관리 모형은 미시적 관점, 거시적 관점, 부가가치적 관점으로 구분할 수 있다. 미시적 관점에서는 데이터 품질관리의 각 요소를 확인하고, 요소별 데이터 품질 향상 방안을 도출한다. 거시적 관점에서는 전사 조직 측면에서, 데이터 관리의 성숙도 단계를 정의하고 각 조직의 성숙도를 측정하여 상위 단계로 발전하도록 유도한다. 부가가치적 관점에서는 데이터품질관리의 비용·효과·위험 모형을 개발하여 각 조직의 상황에 맞는 데이터 품질관리 방안을 제시한다.

데이터 품질관리 요소는 데이터 값(Data Value), 데이터 서비스(Data Service), 데이터 구조(Data Hierarchy), 데이터 관리 프로세스(Data Management Process) 등으로 구분이 가능하다. 이러한 요소들은 서로 연계되어 조직 데이터

---

품질에 영향을 주고 있으므로, 종합적·체계적인 품질 관리 노력이 요구된다. 그 결과 <표 1>와 같은 데이터 품질관리 프레임워크가 수립되었다. 여기서 데이터 품질관리 프레임워크는 데이터 품질관리의 대상이 되는 구성요소와 요소간의 관계를 정의한 데이터 품질관리의 기본 개념 틀을 의미한다.

<표 1> 데이터 품질관리 프레임워크

조직 \ 대상	데이터 값	데이터 구조	데이터관리 프로세스
CIO* / EDA** (개괄적 관점)	데이터 관리 정책		
DA*** (개념적 관점)	표준 데이터	개념 데이터 모델 데이터 참조 모델	데이터 표준관리 요구사항 관리
Modeler (논리적 관점)	모델 데이터	논리 데이터 모델	데이터 모델 관리 데이터 흐름 관리
DBA**** (물리적 관점)	관리 데이터	물리 데이터 모델 데이터베이스	DB 관리 DB 보안 관리
User (운영적 관점)	업무 데이터	사용자 View	데이터 활용 관리

\* CIO: Chief Information Officer  
\*\* EDA: Equipment Data Associates

\*\*\* DA: Data Administrator  
\*\*\*\* DBA: Database Administrator

출처: 한국데이터베이스진흥센터 (2006)

### 가. 데이터의 이해

표준 데이터는 정보시스템에서 사용하는 용어, 도메인, 코드와 기타 데이터의 관련 요소에 대해, 동일한 형식과 내용으로 정의하여 사용하는 표준 관련 데이터를 의미한다. 이러한 데이터는 정보시스템의 데이터 품질을 확보하기 위한 필수 요소로써, 이를 정의·관리하여 데이터 간의 불일치와 데이터의 오류를 방지하고 데이터에 대한 이해도를 높일 수 있다. 또한 기관이나 기업 전사 차원에서 단일화·표준화된 정보시스템을 구현하여 데이터 통합을 효율적으로 수행할 수 있도록 한다.

모델 데이터는 데이터 모델을 운영하고 관리하는데 필요한 데이터를 의미한다. 모델 데이터는 데이터 참조 모델과 개념 데이터 모델, 논리 데이터 모델, 물리



---

데이터 모델에 대한 메타 데이터, DBMS 객체정보를 포함한다. 데이터 모델에 대한 메타 데이터를 관리하여 데이터 구조의 최신 정보를 유지하고 전사 차원의 데이터 모델을 공유하고 재사용을 극대화하며, 체계적인 데이터 모델의 변경 관리를 가능토록 한다.

관리 데이터는 DB를 효과적으로 운영하고 관리하는데 필요한 데이터를 의미한다. 관리 데이터에는 사용 관리 데이터, 장애·보안 관리 데이터, 성능 관리 데이터, 흐름 관리 데이터, 품질관리 데이터 등이 포함된다. DB의 사용과 성능 관리, 데이터 흐름 관리, 품질관리에 대한 데이터를 조직 내에서 정의한 목표 값을 기준으로 수시 또는 정해진 시간에 따라 관리하여 DB를 효율적으로 운영·활용할 수 있다.

업무 데이터는 기관이나 기업의 업무와 비즈니스를 수행하는데 필요한 데이터를 의미한다. 업무 데이터는 데이터 흐름에 따라 소스 데이터, 운영 데이터, 분석 데이터로 구분한다. 정보시스템 정보의 근간이 되는 데이터를 소스 데이터, 운영 데이터, 분석 데이터와 같이 각각의 특성에 따라 분할·관리하여 데이터 관리와 DB 구축 업무의 효율성을 도모할 수 있다.

## 나. 데이터 구조의 이해

개념 데이터 모델은 업무 요건을 충족하는 데이터의 주제영역과 핵심 데이터 집합을 정의하고 상호 간의 관계를 정의한 모델이다. 개념 데이터 모델은 구축하고자 하는 업무 모델의 핵심 데이터 구조를 설계하여, 전체 업무에 대한 큰 윤곽을 잡고 세부적인 단계로 진전토록 한다. 즉, 개념 데이터 모델은 데이터의 기본구조를 확실히 정의하여 업무 요건 변경에 취약하지 않은 데이터 구조를 구축할 수 있게 한다. 또한, 기관이나 기업의 업무 특성에 적합한 주제영역과 핵심 데이터 집합과의 관계를 정의하여 상세 논리 데이터 모델과 물리 데이터 모델과의 데이터 구조적 연결을 지원한다.

데이터 참조 모델은 데이터 아키텍처의 구축과 유지관리, 조직에서 사용하고 있는 데이터 모델의 상호 운영과 타 조직 데이터모델의 참조와 재사용을 목적으로 업무영역별, 주제영역별로 표준 데이터 집합과 관리 항목들을 정의한 모델을 말한다. 데이터 참조 모델을 통해, 새로이 데이터 모델링을 할 경우

---

사전에 기본 정보를 확보함으로써 전체 모델링에 소요되는 시간 절약이 가능하다. 새로운 데이터 모델링 시 데이터 참조 모델을 활용하여 정보의 누락을 예방할 수 있고, 기존에 검증된 데이터 참조 모델을 이용하여 자사 데이터 모델의 오류를 확인하거나 보완할 수 있다. 또한 중앙 기관이나 부서에서 양질의 데이터 참조 모델을 채택하고 활용하여 단기간에 하부 기관이나 부서 데이터 모델의 품질 개선이 가능하다.

논리 데이터 모델은 개념 데이터 모델을 상세히 하여 논리적인 데이터 집합, 관리 항목, 관계를 정의한 모델이다. 논리 데이터 모델은 전체 데이터 구조에서 가장 핵심이 되는 모델로, 전체 업무 범위와 업무 구성 요소를 확인할 수 있다. 논리 데이터 모델 단계에서 데이터를 상세히 정의하고 관리하여 불필요한 데이터 중복과 데이터의 불일치를 미연에 방지할 수 있다.

물리 데이터 모델은 DBMS의 특성과 성능을 고려하여 논리 데이터 모델을 구체화시킨 모델이다. 물리 데이터 모델은 DBMS 선정한 다음 정의하며 해당 DBMS에서 최상의 성능을 보장하고 논리 데이터 모델에서 저장하는 데이터의 물리적 특성을 최대한 반영하여 설계하고 관리한다. 물리 데이터 모델의 설계 단계에서 샘플 데이터를 이용하여 논리데이터 모델의 정합성을 재검증할 수 있다. 또한, 물리 데이터 모델의 설계를 위해 업무 요건과 필요에 따라 사용자 화면이 완성되어야 한다. 따라서 물리 데이터 모델은 사용자 애플리케이션과 상호검증하에 설계될 수 있다.

DB는 물리 모델을 구현한 결과물이고, 구축된 실제 데이터가 저장되는 데이터 저장소를 의미한다. DB 저장소인 테이블과 접근 속도를 보장하기 위한 인덱스, 업무규칙이 반영된 제약 사항과 기타 DB 관련 객체를 정의하고 관리하여 DB를 효과적으로 운영할 수 있다.

사용자 View는 데이터를 제공하는 정보시스템 상의 화면이나 출력물을 말한다. 사용자 View는 데이터 품질관리 전반에 걸쳐 수행한 작업의 결과물이며 사용자에게 제공되는 최종 산출물이므로, 데이터에 대한 만족도를 극대화하기 위해 사용자가 요구하는 수준에 따라 개발되고 관리되어야 한다. 화면, 출력물과 시스템의 구조적 관계를 정의하여 관리하면 사용자 View의 개선과 관리를 위해 필요한 데이터 모델이나 SQL 등에 대한 일련의 변경 작업을 신속·정확하게 수행할 수 있다.

---

## 다. 데이터 관리 프로세스의 이해

데이터 관리 프로세스는 고품질 데이터의 운영·관리를 위한 데이터 관리 정책 수립에서부터 데이터 표준 관리, 데이터 모델 관리, 데이터 흐름 관리, DB관리, DB 보안관리, 데이터 활용 관리, 요구사항 관리에 이르는 일련의 데이터 관리 활동들을 체계적인 흐름으로 표현한 프로세스를 의미한다. 고품질의 데이터 관리를 위한 체계적인 프로세스를 이해시켜, 기관이나 기업의 조직 환경에 가장 적합한 형태로 데이터의 정의, 평가, 변경과 같은 일련의 프로세스를 수립할 수 있도록 지원하는 역할을 한다.

데이터 관리 정책은 기업의 비전·목표 달성을 위해 필요한 데이터 확보 계획과 확보한 데이터를 효과적으로 유지하고 관리하기 위한 체계와 계획을 정의하는 작업이다. 기업은 데이터 관리 정책을 수립하여 기관의 비전·목표에 알맞은 데이터를 확보하고, 확보한 데이터를 사용자가 원하는 시간에 원하는 형태로 안정적으로 서비스할 수 있는 기본 원칙의 수립과 관리 체계의 구축이 가능하다.

데이터 표준 관리는 데이터 표준화 원칙에 따라 정의된 표준 단어 사전과 도메인 사전, 표준 용어 사전, 표준 코드, 데이터 관련 요소 표준 등을 기관에 적합한 형태로 정의, 변경, 관리하고, 데이터 표준의 준수 여부 체크 등을 통한 데이터 정제와 개선 활동 등을 의미한다. DB의 설계와 개발을 지원하고 전사적인 데이터 표준의 사용과 재사용을 통해 시스템 간의 상호 운용성, 데이터 공유, 시스템 통합, 비즈니스 프로세스 개선 등을 지원한다. 전사적으로 공통된 표준을 사용하게 하여 데이터의 일관성과 정합성을 유지할 수 있다.

데이터 모델 관리는 데이터 요구사항 관리에 의해 변경되는 데이터 구조를 모델에 반영하는 작업 절차와 DB 시스템 구조와 동일하게 데이터 모델을 유지하도록 하는 작업 절차를 말한다. 데이터 모델은 기관의 비즈니스 목적에 맞는 최적화된 데이터 서비스를 제공하고 DB를 구성하고 유지하기 위해 체계적으로 관리되어야 한다.

데이터 흐름 관리는 소스 데이터(문서, Text, DB 등)를 수기로 생성하거나 추출, 변환, 적재를 통해 생성하여 타깃 DB에 저장하고 가공하는 것을 관리하는

---

것을 의미한다. 각 기관이 관리 중인 데이터가 생성, 변경되고 활용되는 생애 주기를 관리하여 전사 데이터에 대한 현황을 파악하고 최적화된 형태로 활용되고 있는가의 여부를 확인할 수 있다.

DB 관리는 원활한 데이터 서비스를 위해 필요한 DB를 안정적으로 운영, 관리하는데 필요한 작업을 체계화하는 것을 말한다. DB 관리에는 백업과 보안, 튜닝, 모니터링 등의 작업이 포함된다. DB 관리 작업은 DB와 DB에 저장된 데이터가 오류와 훼손 없이 안정적으로 서비스될 수 있도록, DB에 대한 생성과 변경, 보안, 성능 개선, 백업 관리를 연속적으로 수행할 수 있도록 체계화하는 작업이다.

DB 보안 관리는 DB와 적재된 데이터를 미허가된 일련의 행위로부터 안전하게 보호하는 것을 목표로, 인적, 정책적, 물리적, 논리적 대응체계가 포함된다. DB 보안 관리 작업은 DB 보안 관리의 기준을 제시하여 내·외부인의 불법적이고 미인가된 위협으로부터 데이터의 기밀성, 무결성, 가용성 확보를 목적으로 한다.

데이터 활용 관리는 데이터의 활용 여부를 점검하거나 활용도를 증진시키기 위해, 측정 대상 데이터와 품질 지표를 선정하고 품질을 측정·분석하여, 품질을 충족시키지 못하는 경우, 원인을 분석하여 담당자가 조치하도록 하는 작업을 의미한다. 애플리케이션에서 미활용 되는 데이터를 점검하여 DB의 사용 환경을 개선하고, 업무적 중요도가 높은 데이터에 대한 품질의 평가와 개선으로 데이터의 활용도를 높인다.

요구사항 관리는 데이터를 비롯하여 관련 애플리케이션 및 시스템 전반에 걸친 사용자의 요구를 수집·분류하여 반영하는 작업을 말한다. 사용자의 정보 요구 사항을 종합적으로 검토하고 확인하여 요건에 적합하도록 시스템을 개선하고 반영함으로써 사용자의 만족도를 높이고 고품질의 서비스를 가능하게 한다.

---

## 2. 데이터 품질평가의 개요

### 가. 데이터 품질평가의 정의와 종류

데이터 품질평가는 해당 조직이 운영·관리 중인 정보시스템에 저장된 정형·비정형 데이터의 품질을 측정하여 현재 수준을 평가하고 품질 저하 요인을 분석하여 개선 사항을 제안하는 절차이다.

데이터 품질평가의 종류는 데이터 값 진단, 데이터 구조 진단, 데이터 관리 프로세스 진단 등으로 구분한다. 정형 데이터에 대한 데이터 값 진단은 운영 데이터베이스의 테이블·컬럼·코드·관계·업무규칙을 기준으로 데이터의 값에 대한 현상을 분석한다. 또한 데이터 값과 관련된 품질 기준을 적용하여, 오류 내역을 산출하고 주요 원인을 분석하여 개선 사항을 제안한다. 비정형 데이터에 대한 데이터 값 진단은 비정형 콘텐츠 자체의 상태와 메타데이터에 대한 데이터 품질진단으로 이루어진다. 데이터 구조 진단은 데이터 모델링 관점에서 데이터 품질을 진단한다. 중요 업무 데이터베이스의 역 모델링(Reverse Modelling)을 통해 논리 모델을 작성하고, 현행 데이터베이스의 구조 무결성·데이터 구조 표준화·관리 수준·변경 관리 등의 현황을 진단한다. 데이터 관리 프로세스 진단은 정형 데이터와 비정형 데이터에 대한 현행 데이터 관리 프로세스를 분석하여 문제점을 도출하고, 이를 개선할 수 있는 핵심 업무 프로세스를 표준화하여 재설계한다.

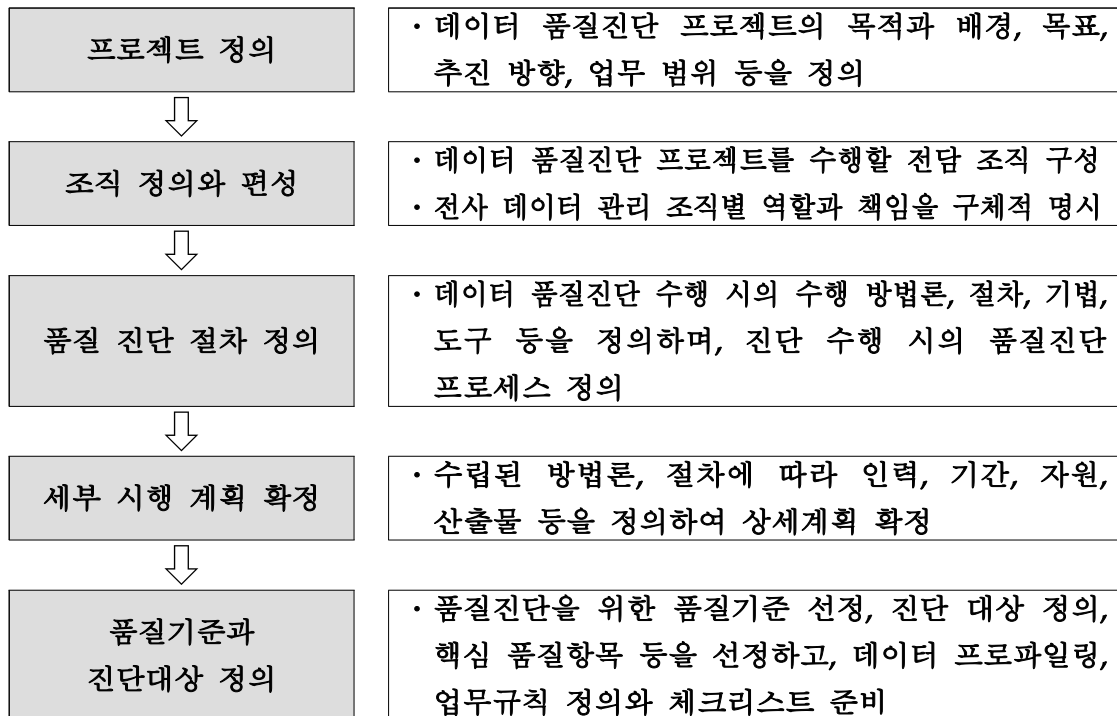
### 나. 데이터 품질평가의 절차

데이터 품질평가의 절차를 품질진단 계획수립, 품질기준과 진단 대상 정의, 품질측정, 품질측정 결과 분석, 데이터 품질 개선 등 5단계로 구분 가능하다.

데이터 품질진단 계획수립 단계는 두 개 과정으로 구분할 수 있다. 첫 번째 과정은 데이터 품질진단을 수행하기에 앞서 실시할 품질진단의 개요와 정의,

데이터 품질과 관련 이용자의 요구사항 파악, 품질진단 수행 조직의 정의, 세부 수행절차 확립 등의 계획을 수립하는 과정이다. 두 번째 과정은 실제 품질측정을 수행하기 위한 품질기준과 대상 선정, 데이터 프로파일링, 업무규칙 정의 및 체크리스트 준비 등이 이루어지는 과정이다. 품질진단 계획수립의 전반적 절차는 아래의 <표 2>와 같다.

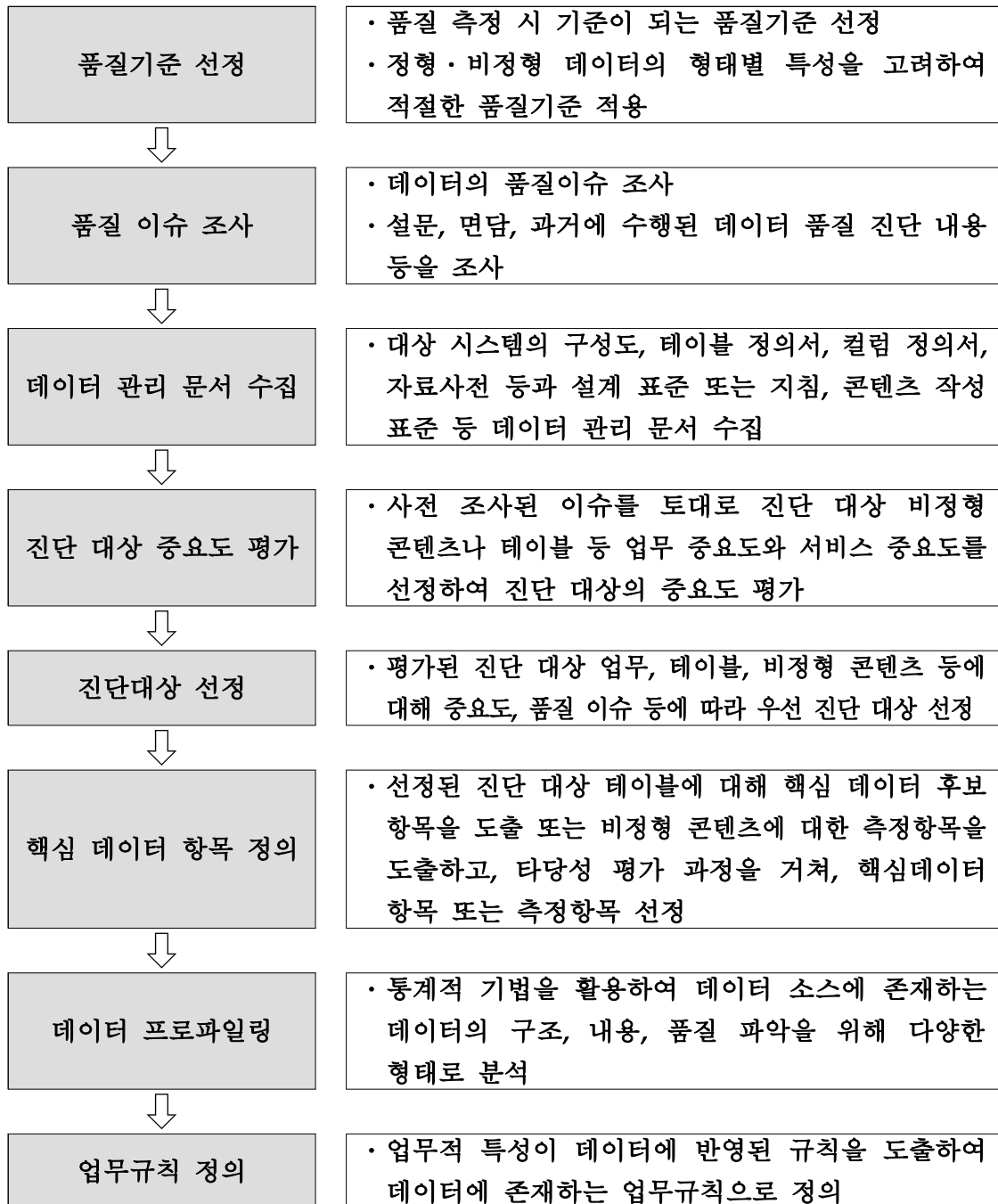
<표 2> 데이터 품질진단 계획수립 절차



출처: 한국데이터베이스진흥센터 (2009)

데이터 품질기준과 진단 대상 정의 단계는 실제로 품질측정을 수행하기 위한 품질기준과 대상 선정, 데이터 프로파일링, 업무규칙 정의와 체크리스트 준비 등이 이루어지는 과정이다. 데이터품질진단 수행 시 적용할 품질측정의 기준을 사전에 정의하고, 데이터의 형태(정형, 비정형)에 따라 데이터 품질진단을 수행할 대상 정보 시스템의 테이블과 컬럼 등을 정의한다. 또는 진단 대상이 되는 멀티미디어 콘텐츠와 해당 메타데이터를 선정하여, 데이터 유형별 특성에 따른 데이터 프로파일링과 업무규칙 정의, 체크리스트 준비 등을 수행한다. 품질기준과 진단 대상 정의의 전반적 절차는 아래의 <표 3>와 같다.

<표 3> 데이터 품질기준과 진단 대상 정의 절차



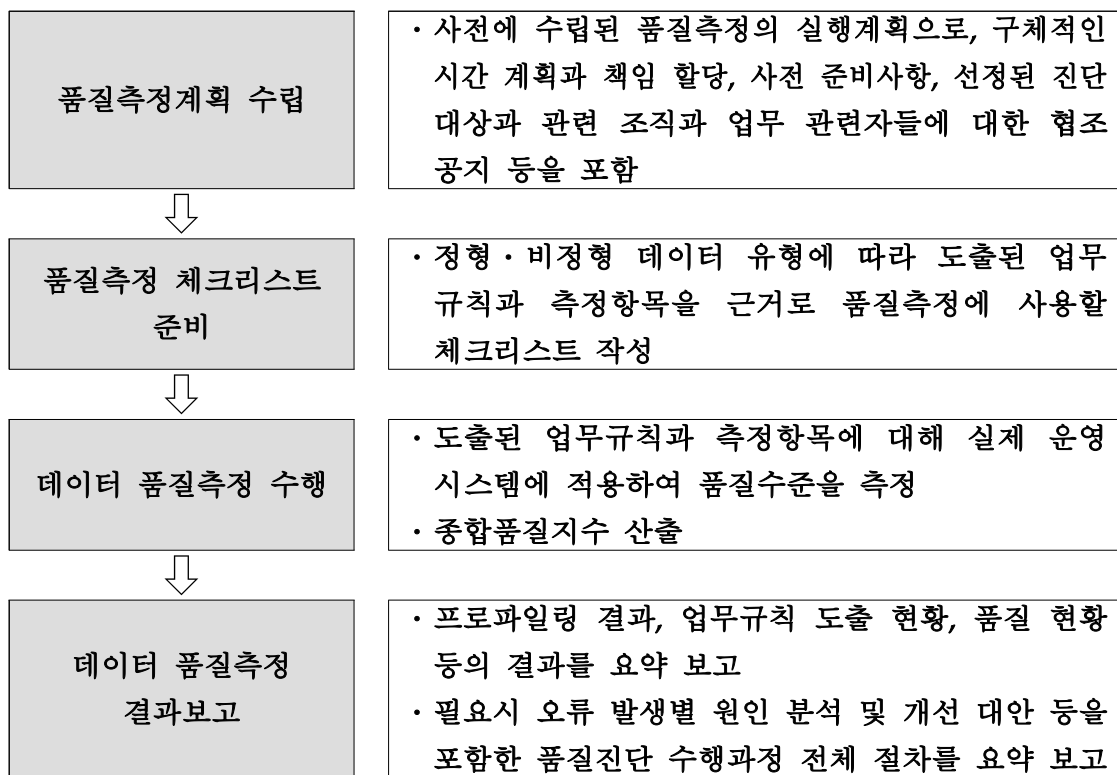
출처: 한국데이터베이스진흥센터 (2009)

정형 텍스트 데이터에 대한 일반적 데이터의 품질기준을 완전성, 유일성, 유효성, 일관성, 정확성의 5개 품질기준으로 나눌 수 있다. 완전성은 필수항목에

누락이 없어야 함을 의미한다. 유일성은 데이터 항목은 유일해야 하며 중복이 있어서는 안 됨을 말한다. 유효성은 데이터 항목이 정해진 데이터 유효범위와 도메인을 충족하여야 함을 의미하며, 일관성은 데이터가 지켜야 할 구조, 값, 표현되는 형태가 일관되게 정의되고, 서로 일치하여야 함을 말한다. 정확성은 실제 존재하는 객체의 표현 값이 정확하게 반영되어야 함을 의미한다.

데이터 품질측정 단계는 도출된 업무규칙을 핵심데이터에 적용하여, 오류데이터를 추출하고 그 현황을 파악하는 오류데이터 검증 단계와 그 결과를 요약하여 품질 현황을 보고하는 단계로 구성된다. 데이터 품질측정 절차는 아래의 <표 4>와 같다.

<표 4> 데이터 품질측정 절차

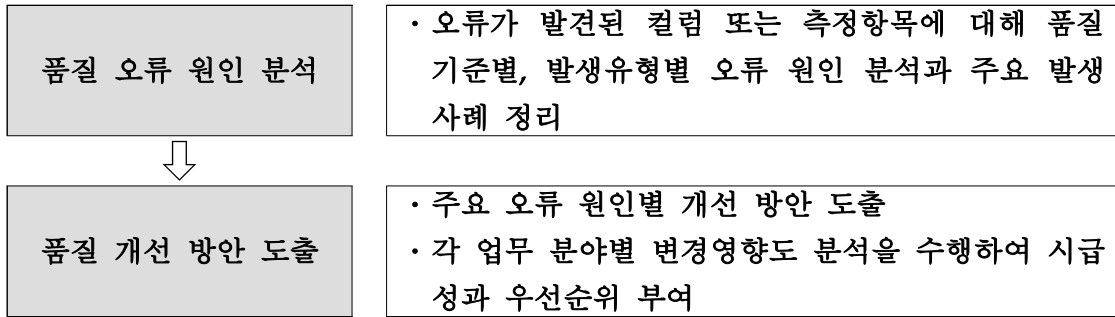


출처: 한국데이터베이스진흥센터 (2009)

데이터 품질측정 결과 분석 단계는 데이터 품질측정이 완료되면 오류 유형에 따른 발생 원인을 분석하고 개선방안을 도출하는 단계이다. 본 단계의 자세한 절차는 아래의 <표 5>와 같다.



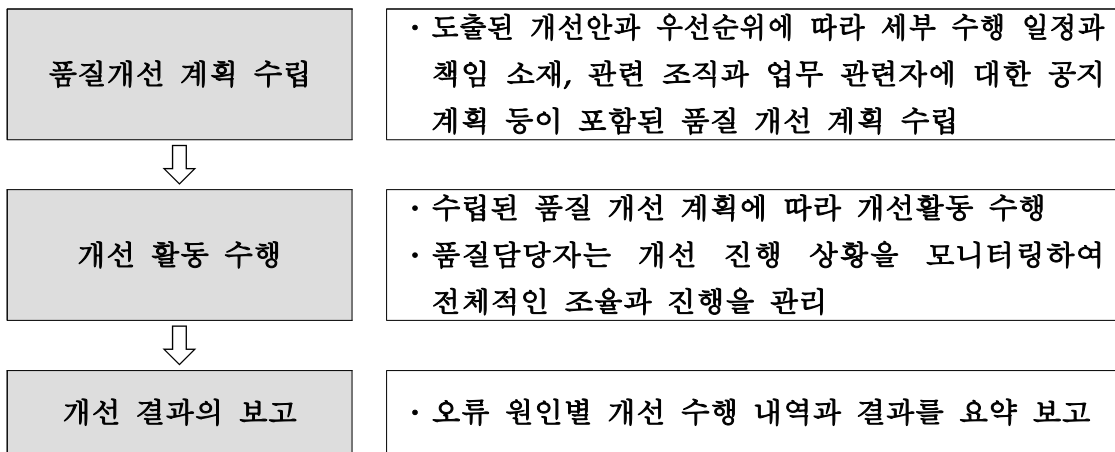
<표 5> 데이터 품질측정 결과 분석 절차



출처: 한국데이터베이스진흥센터 (2009)

데이터 품질개선 단계는 오류유형과 각 업무 분야에 해당되는 개선방안의 시급성과 우선순위에 따라 데이터 품질 개선계획을 수립하고 개선활동을 수행한다. 개선 활동이 완료되면 개선 결과를 검토하고 요약하여 보고하는 단계이다. 아래의 <표 6>은 데이터 품질개선의 세부 절차를 나타낸다.

<표 6> 데이터 품질개선 절차



출처: 한국데이터베이스진흥센터 (2009)

---

## 제2장

### 데이터 품질평가의 기존 연구와 사례

---

1. 기존 연구
2. 국내·외 데이터 품질관리의 사례

---

# 1. 기존 연구

## 가. NTIS에 대한 기존 연구

NTIS에 대한 연구는 한국과학기술정보연구원을 중심으로 다양하게 수행되고 있다. NTIS의 소개와 포괄적 설명을 담은 연구로, 김재수 (2009)는 NTIS의 개요와 추진 체계, 주요 서비스 내용, R&D정보 서비스 확대 방안을 설명하였다. 양명석 · 최광남 · 정옥남 · 김재수 (2013)는 NTIS의 추진경과를 개발 단계, 고도화 단계, 차세대 단계 등으로 나누어 각 단계별 개방방안과 방향, 특성, 구조 등에 대해 설명하고 향후 NTIS의 발전 방향을 제시하였다.

NTIS의 정보연계에 관한 연구로, 손강렬 · 임종태 (2009)는 NTIS의 국가R&D 참여인력/과제/성과정보의 정보연계 방식과 데이터 정제 프로세스에 대해 알아보고, 정보연계 과정에서 발생 가능한 데이터 품질문제 요인을 분석하고 해결책을 제시하였다. 신성호 · 윤영준 · 양명석 · 김진만 · 손강렬 (2011)는 분산되어 있는 데이터베이스를 통합하여 공동 활용을 위한 데이터베이스 통합 모형을 개발하고, 통합 데이터베이스 내 데이터 품질 향상 방안을 제안하였다. 그 결과, 정제된 데이터가 정제하지 않은 데이터에 비해 품질이 향상되었음을 밝혔다. 최희석 · 김재수 (2012)는 정보연계방식의 특성과 장 · 단점을 비교하고 사례를 분석하였다. 또한 NTIS의 정보연계 현황을 분석하고, 연계서버 자원활용의 효율성 향상과 데이터 수집의 즉시성을 확보할 수 있는 방식을 적용한 정보연계 모델을 제안하였다.

NTIS의 데이터를 대상으로 한 품질평가 방법에 관한 연구로, 김진만 · 신성호 · 윤영준 · 양명석 · 손강렬 (2010)은 데이터 품질관리의 중요성을 강조하며, 데이터 값에 대한 정량적 측정 및 평가를 위한 방법을 소개하였다. 정량적 측정방법으로는 대상 데이터의 오류율에 서비스 가중치를 반영한 방법을 제시하였다.

NTIS의 IT서비스관리 측면에서 논의한 연구로, 최희석 · 정옥남 · 김윤정 · 김태현 · 임철수 · 이병희 · 김재수 (2009)는 NTIS의 효율적인 기획과 업무운영을

---

---

위한 NTIS 표준운영절차(Standard Operating Procedures, SOP)의 수립하고 SOP 공유시스템을 소개하였다. 이병희·임철수·신동구·정옥남·김재수 (2009)는 NTIS의 IT서비스관리 국제 표준인증인 ISO/IEC 20000과 정보보호 국제 표준인증인 ISO 27001 획득 사례를 소개하며, 인증을 통해 서비스 품질 제고와 객관적 신뢰도 확보를 기대사항으로 제시하였다.

NTIS 사용자의 고객가치에 관한 연구로, 서상혁·이선영·이병희 (2012)는 NTIS와 국가과학기술정보센터(National Digital Science Library, NDSL)의 사용자를 대상으로 정보마케팅관점에서 정보고객가치와 만족도 영향요인을 실증 분석하였다. 분석결과를 근거로 한 인과모형에 따르면 정보고객가치에 가장 큰 영향을 미치는 변수는 정보품질, 서비스품질 순으로 나타났다. 또한, 정보마케팅적 관점에서 정보고객행동을 분석하고 고객지향적 서비스를 벤치마킹해야 한다고 주장하였다. 서상혁·이선영·이병희 (2014)는 NTIS를 이용하는 연구관리자를 대상으로 정보서비스 영향요인을 분석하였다. 그 결과, 정보서비스품질은 정보고객가치와 정보원인식에 정(+)의 영향을 미치며, 정보서비스품을 우선하기 위해서는 고객지향 정보서비스 전략이 수립되어야 한다고 주장하였다.

위 선행연구들을 종합하면, 정보서비스의 품질 향상 노력으로 NTIS에 대한 사용자의 신뢰도와 만족도를 제고할 수 있다. 이를 위해, NTIS는 IT서비스관리와 같은 전문 경영 시스템을 도입하고, 데이터 품질향상을 위한 다양한 정보연계 체계 구축 방법에 관한 연구들을 수행하였음을 알 수 있다. 다만, 품질향상을 위한 연구가 IT, 시스템적으로 접근한 연구가 다수이고, 데이터에 대한 품질을 측정·평가한 연구가 많이 이루어지지 않았다는 점에서 객관적인 품질평가 방법에 관한 연구가 수행되어야 할 필요가 있다.

<표 7> NTIS에 대한 기존 연구

구 분	저 자	내 용
NTIS의 소개	김재수 (2009)	· NTIS의 개요와 추진 체계, 주요 서비스 내용, R&D 정보 서비스 확대 방안 소개
	양명석 · 최광남 · 정옥남 · 김재수 (2013)	· NTIS의 추진경과를 개발 단계, 고도화 단계, 차세대 단계 등으로 나누어 각 단계별 개방방안과 방향, 특성, 구조 등에 대해 설명 · 향후 NTIS의 발전 방향을 제시
NTIS의 정보연계	손강렬 · 임종태 (2009)	· NTIS의 국가R&D참여인력/과제/성과정보 연계 방식과 데이터 정제 프로세스 소개 · 정보연계 과정에서 발생 가능한 데이터 품질문제 요인을 분석하고 해결책 제시
	신성호 · 윤영준 · 양명석 · 김진만 · 손강렬 (2011)	· 분산되어 있는 데이터베이스를 통합하여 공동 활용을 위한 데이터베이스 통합 모형 개발 · 통합 데이터베이스 내 데이터 품질 향상 방안을 제안 · 정제된 데이터가 정제하지 않은 데이터에 비해 품질이 향상되었음을 밝힘
	최희석 · 김재수 (2012)	· 정보연계방식의 특성과 장단점을 비교하고 사례 분석 · NTIS의 정보연계 현황을 분석 · 연계서버 자원활용의 효율성 향상과 데이터 수집의 즉시성을 확보할 수 있는 방식을 적용한 정보연계 모델을 제안
NTIS의 품질평가	김진만 · 신성호 · 윤영준 · 양명석 · 손강렬 (2010)	· 데이터 품질관리의 중요성을 강조 · 데이터 값에 대한 정량적 측정 및 평가를 위한 방법을 소개 · 정량적 측정방법으로는 대상 데이터의 오류율에 서비스 가중치를 반영한 방법을 제시하였다.
NTIS의 IT서비스관리	최희석 · 정옥남 · 김윤정 · 김태현 · 임철수 · 이병희 · 김재수 (2009)	· NTIS의 효율적인 기획과 업무운영을 위한 NTIS 표준운영절차(Standard Operating Procedures, SOP)의 수립하고 SOP 공유 시스템을 소개

구분	저자	내용
NTIS의 IT서비스관리	이병희 · 임철수 · 신동구 · 정옥남 · 김재수 (2009)	<ul style="list-style-type: none"> <li>· NTIS의 IT서비스관리 국제 표준인증인 ISO/IEC 20000과 정보보호 국제 표준인증인 ISO 27001 획득 사례를 소개</li> <li>· 인증을 통해 서비스 품질 제고와 객관적 신뢰도 확보를 기대사항으로 제시</li> </ul>
NTIS 사용자의 고객가치	서상혁 · 이선영 · 이병희 (2012)	<ul style="list-style-type: none"> <li>· NTIS와 국가과학기술정보센터(National Digital Science Library, NDSL)의 사용자를 대상으로 정보마케팅관점에서 정보고객 가치와 만족도 영향요인을 실증분석</li> <li>· 분석결과를 근거로 한 인과모형에 따르면 정보고객가치에 가장 큰 영향을 미치는 변수는 정보품질, 서비스품질 순</li> <li>· 정보마케팅적 관점에서 정보고객행동을 분석하고 고객지향적 서비스를 벤치마킹해야 한다고 주장</li> </ul>
	서상혁 · 이선영 · 이병희 (2014)	<ul style="list-style-type: none"> <li>· NTIS를 이용하는 연구관리자를 대상으로 정보서비스 영향요인을 분석</li> <li>· 정보서비스품질은 정보고객가치와 정보원 인식에 정(+)의 영향을 미치며, 정보서비스 품질을 우선하기 위해서는 고객지향 정보 서비스 전략이 수립되어야 한다고 주장</li> </ul>

#### 나. 데이터 품질평가에 대한 기존 연구

김문영 (2005)은 데이터 품질의 평가 요소를 크게 데이터의 내용, 데이터의 구조, 데이터의 이동과 흐름관점으로 파악하였다. 데이터의 내용은 데이터베이스에 저장되어 있는 데이터의 값을 의미하며, 데이터의 내용을 대상으로 하는 품질 평가는 데이터의 값이 업무적으로 유효하고 사실과의 일치여부, 상호 모순 여부를 평가한다. 주요 기준은 데이터 내용의 완전성, 유효성, 정확성, 일관성으로 구분하였다. 데이터의 구조에 대한 품질평가는 기업의 경영 목표 달성을 위해 필요한 데이터들이 사용자와 전사적 관점에서 인식, 분석하고 정의하였는가를 평가한다. 데이터의 이동과 흐름에 관한 평가는 데이터를 최종 사용자에게

---

유의미한 정보나 지식으로의 변환하기 위한 과정으로, 적시성과 편리성을 평가 요소로 선정하였다.

정혜정(2007)은 국제 표준인 소프트웨어 품질평가 모델 ISO/IEC 25012을 중심으로 데이터의 품질평가를 위한 품질 특성을 기능성, 사용성, 효율성, 신뢰성, 이식성, 준수성으로 구분하여 총 19개의 평가항목들을 제시하였다. 다만 6가지 품질 특성에 적용할 가중치 문제는 소프트웨어 제품의 분류 기준의 불명확성으로 인해 해결하지 못하였다. 19개 평가항목 이외에도 객관성, 업무 관련성, 비용 효율성 등에 관한 평가가 데이터 품질평가지 고려되어야 한다고 주장하였다.

최승준·박제원·김종배·최재현(2014)는 빅데이터 분산처리 지원 시스템에 대한 품질평가모형을 제안하였다. 제안 모델은 기능성, 신뢰성, 효율성, 유지 보수성, 충분성 등 5가지 특성을 이용하여 빅데이터 처리성능 또는 사용자의 빅데이터 활용 정도와 같은 빅데이터 자원 분석 측면에서의 시스템에 대한 품질 측정이 가능함을 보였다.

정영기(2015)는 공공 개방데이터의 품질을 진단하는 방법을 제시하며, 품질기준 별 중점 품질진단 항목을 분류하고 실제 공공 개방데이터를 대상으로 품질을 평가하였다. 데이터의 품질기준은 완전성, 유효성, 일관성, 정확성, 적시성, 보안성으로 나누어지며, 프로파일링(값, 구조 진단), 업무규칙 진단, 체크리스트(인터뷰, 설문) 등 품질진단기법에 따라 각각 달리 적용하였다. 데이터 표준화 지침 수립과 관리를 위한 프로세스 정립, 품질평가 자동화 도구 도입, 빅데이터에 대한 이해가 수반된 품질관리의 요구를 주장하였다.

호진원·이미영(2013)는 데이터베이스를 활용한 감사에서 데이터품질이 감사 결과에 미치는 영향을 감사 사례를 통해 알아보았다. 주요 데이터 품질관리 문제로는 필수 데이터 미등록, 데이터 오입력 등 데이터 입력 소홀 문제, 서로 다른 시스템간 정보공유 상에서 발생하는 공통 데이터의 불일치 문제, 입력 데이터의 신뢰성과 정확성 검증 미흡 문제 등이 있다. 중앙행정기관의 경우, 데이터 품질관리를 기관의 성과지표 목표로 설정하고, 공공기관의 경우, 공공기관 평가지표에 데이터 품질관리 지표를 추가하여 품질관리의 의무화를 주장하였다. 또한, 피감기관의 데이터 품질 향상에 초점을 맞춘 새로운 감사접근방법 모색을 강조하였다.

<표 8> 데이터 품질평가에 대한 기존 연구

저 자	내 용
김문영 (2005)	<ul style="list-style-type: none"> <li>· 데이터 품질의 평가 요소를 크게 데이터의 내용, 데이터의 구조, 데이터의 이동과 흐름관점으로 파악</li> <li>· 데이터의 내용을 대상으로 하는 품질평가는 데이터의 값이 업무적으로 유효하고 사실과의 일치여부, 상호 모순 여부를 평가. 주요 기준은 데이터 내용의 완전성, 유효성, 정확성, 일관성으로 구분</li> <li>· 데이터의 구조에 대한 품질평가는 기업의 경영 목표 달성을 위해 필요한 데이터들이 사용자와 전사적 관점에서 인식, 분석하고 정의하였는가를 평가</li> <li>· 데이터의 이동과 흐름에 관한 평가는 데이터를 최종 사용자에게 유의미한 정보나 지식으로의 변환하기 위한 과정으로, 적시성과 편리성을 평가 요소로 선정</li> </ul>
정혜정 (2007)	<ul style="list-style-type: none"> <li>· 국제 표준인 소프트웨어 품질평가 모델 ISO/IEC 25012을 중심으로 데이터의 품질평가를 위한 품질 특성을 기능성, 사용성, 효율성, 신뢰성, 이식성, 준수성으로 구분하여 총 19개의 평가항목들을 제시</li> <li>· 19개 평가항목 이외에도 객관성, 업무 관련성, 비용효율성 등에 관한 평가가 데이터 품질평가지 고려되어야 한다고 주장</li> </ul>
최승준 · 박제원 · 김종배 · 최재현 (2014)	<ul style="list-style-type: none"> <li>· 빅데이터 분산처리 지원 시스템에 대한 품질평가모델을 제안</li> <li>· 제안 모델은 기능성, 신뢰성, 효율성, 유지보수성, 충분성 등 5가지 특성을 이용하여 빅데이터 처리성능 또는 사용자의 빅데이터 활용 정도와 같은 빅데이터 자원 분석 측면에서의 시스템에 대한 품질 측정이 가능함을 보임</li> </ul>
정영기 (2015)	<ul style="list-style-type: none"> <li>· 공공 개방데이터의 품질을 진단하는 방법을 제시하며, 품질기준 별 중점 품질진단 항목을 분류하고 실제 공공 개방데이터를 대상으로 품질을 평가</li> <li>· 데이터의 품질기준은 완전성, 유효성, 일관성, 정확성, 적시성, 보안성으로 나누어지며, 프로파일링(값, 구조 진단), 업무규칙 진단, 체크리스트(인터뷰, 설문) 등 품질진단 기법에 따라 각각 달리 적용</li> <li>· 데이터 표준화 지침 수립과 관리를 위한 프로세스 정립, 품질평가 자동화 도구 도입, 빅데이터에 대한 이해가 수반된 품질관리의 요구를 주장</li> </ul>



저 자	내 용
호진원 · 이미영 (2013)	<ul style="list-style-type: none"> <li>· 데이터베이스를 활용한 감사에서 데이터품질이 감사결과에 미치는 영향을 감사 사례를 통해 알아봄</li> <li>· 주요 데이터 품질관리 문제로는 필수 데이터 미등록, 데이터 오염력 등 데이터 입력 소홀 문제, 서로 다른 시스템간 정보공유 상에서 발생하는 공통 데이터의 불일치 문제, 입력 데이터의 신뢰성과 정확성 검증 미흡 문제</li> <li>· 중앙행정기관의 경우, 데이터 품질관리를 기관의 성과지표 목표로 설정하고, 공공기관의 경우, 공공기관 평가지표에 데이터 품질관리 지표를 추가하여 품질관리의 의무화를 주장. 또한, 피감기관의 데이터 품질 향상에 초점을 맞춘 새로운 감사접근방법 모색을 강조.</li> </ul>

## 2. 국내 · 외의 데이터 품질관리 사례

### 가. 관세청의 데이터 품질관리

관세청은 관세행정 데이터의 폭발적 증가, 신규 품목의 지속적 확대 등에 따른 신속·정확한 통관업무 지원을 위해 수출입 통관 및 화물 정보의 정확도 등 데이터 품질의 향상이 필요하였다. 데이터의 효율적 관리를 통한 고부가가치 창출을 위한 노력이 확산되고, 이를 위한 데이터 품질관리의 중요성이 대두되었다. 이에 관세청은 고품질 데이터 확보와 체계적인 품질관리 활동 수행을 위해 중장기 ISP 추진을 통해 데이터 품질관리 로드맵을 수립하고 단계적으로 과제를 추진하였다.



(그림 1) 관세청의 데이터 품질관리 로드맵

(출처: 한국정보화진흥원, 2012)

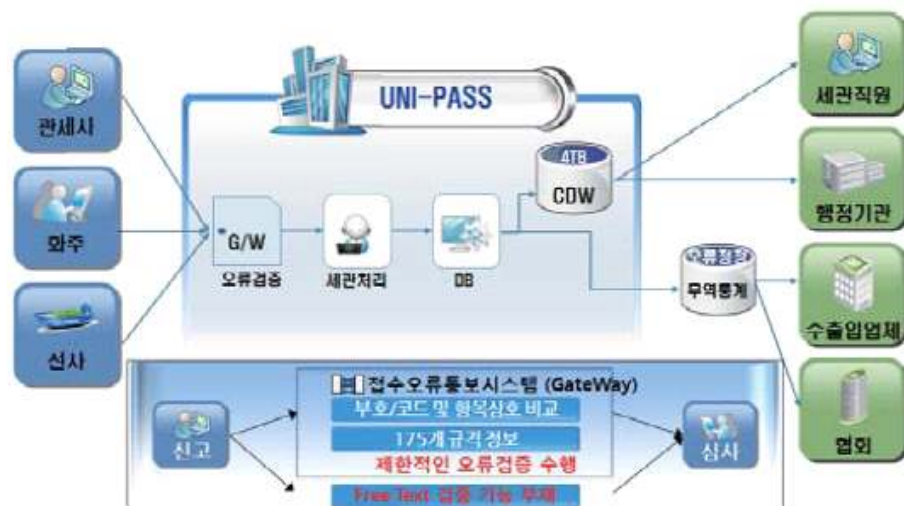
관세청은 2009년부터 본격적인 품질관리 활동을 수행하였으며 2012년까지 약 38.6억 원의 예산과 54명의 인력을 투입하였다.

<표 9> 관세청의 품질관리 투입 예산과 인력 규모 추이

	'09년	'10년	'11년	'12년	합 계
투자규모 (백만원)	1,000	1,040	960	860	3,860
인력규모 (명)	15	19	16	14	54

출처: 한국정보화진흥원 (2012)

관세청의 데이터는 생산 주체가 관세사, 화주, 선사 등의 수출입 관련 신고 민원 입력을 통해 최초 생성되는 특징을 보유하고 있어 저품질의 부정확한 데이터 입력이 주요 이슈로 대두되었다. 수출입 관련 신고 민원의 특성상 정형화할 수 없는 입력 중심의 텍스트 데이터가 많아, 저품질 데이터가 지속적으로 생산되었다. 저품질 데이터 입력은 무역통계 부정확으로 이어지고, 나아가 국가 경제정책 수립 저해와 관세청의 대내·외 신뢰도 저하를 야기하였다.



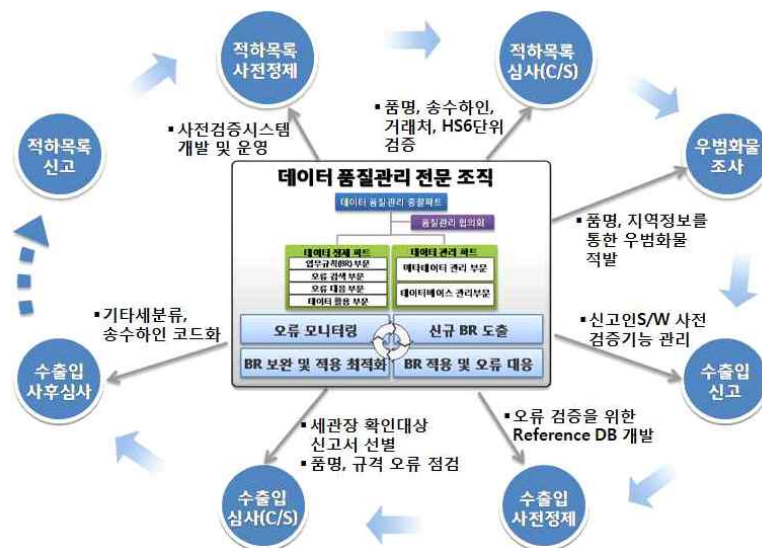
(그림 2) 관세청 데이터 흐름도

(출처: 한국정보화진흥원, 2012)

이러한 문제를 해결하고자, 관세청은 수출입 신고 등 외부 민원인의 입력을 통해 데이터가 최초 생산되는 데이터의 특성을 고려하여, 오류 데이터 생성을 사전에 방지하는 체계 중심으로 품질관리 활동을 추진하였다. 부정확한 데이터 입력의 사전 방지를 위해 S/W 개발업체를 대상으로 기본적이고, 구조적인 오류를 방지할 수 있는 설명회를 개최하였다. 또한, 데이터를 생산하는 관세사, 선사 등 약 1,000명을 대상으로 입력 오류 및 사전검증 체계 교육을 실시하고, 부실 데이터 입력에 대한 과태료 부과 등의 제도시행 추진을 병행하였다. 부정확한 입력 데이터의 사전검증을 위한 참조DB구축과 시범 적용으로 입력통제를 자동화하였다. 품명 입력 시 시스템에서 자동으로 오류여부 선별할 수 있는 오류검증 업무규칙 수립하고 적용하였다.

상기 활동과 같이 지속적인 품질개선 활동으로, 양질의 데이터를 확보하여 얻어지는 관세 수입의 증가, 업무 효율성 증가 등을 계량화하여 제시함으로써 지속적인 예산을 확보하고, 전 직원의 관심과 참여를 유도하였다.

데이터 품질관리 성과로는 중복되어 있고 부정확하게 등록되어 있는 해외 거래처 중복코드를 정제하여 거래처 정보의 정확도를 향상시켰다. 품명 검증을 통해 사전에 인지하지 못했던 추가적인 데이터 오류를 발견하고 개선하였다. S/W 기능 개선을 통한 오류를 원천 방지하였다. 또한, 품명과 수량정보를 붙여서 입력하는 오류를 방지하기 위해 품명과 수량정보를 자동으로 분리하여 전송되도록 하는 가이드라인을 마련하여 S/W 개발사에 제공하였다.



(그림 3) 관세청 데이터 품질관리 전담조직의 역할

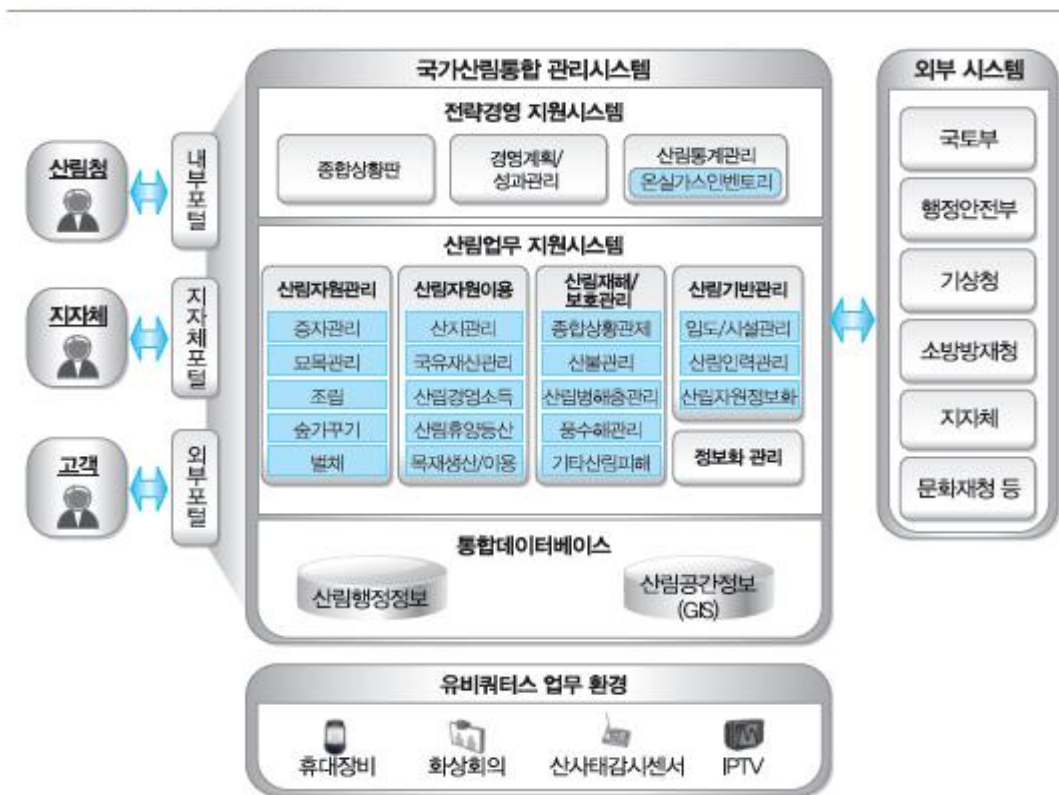
(출처: 한국정보화진흥원, 2012)

---

## 나. 산림청의 데이터 품질관리

2010년 산림청의 '국가생물종지식정보시스템'이 국내 정부부처 최초로 데이터 품질관리 부문 인증서를 취득하였다. 산림청은 산림을 효율적·과학적으로 관리·지원하기 위해 정보 인프라 구축에서 스마트 U-Forest 구현까지의 전 영역을 정보화하고 있다. 또한 정보화 효율성 제고를 위한 제도도입, 체계적인 정보화 사업추진을 위한 기반마련, 대국민과 산림공무원 관점에서 산림서비스 제공을 위해 산림정보화를 추진하고 있다. 산림정보화 기본계획(2010~2014년)은 산림청의 정보화비전 '녹색 산림행정을 선도하는 국가산림 통합 정보화구현'의 실현에 맞춰져 있다. 이를 위해 국가정보화 기본계획(행정안전부), 녹색성장 5개년 계획(녹색성장위원회)의 세부 추진과제와 연계해 수립한 국가산림통합관리체계 구축, 대국민 서비스 통합 및 강화, 유비쿼터스 기술을 활용한 현장업무 개선, 정보 통합 관리 및 전략적 활용 등 4개의 정보화 전략과 전략별 상세과제의 주요 내용을 담고 있다. 특히 산림 업무를 정보화 지원으로 효율성을 높일 수 있도록 현장업무를 산림자원관리, 산림자원이용, 산림재해보호관리, 산림기반관리업무로 통합 구축하였다. 또 각 분야에서 생성되는 데이터는 통합DB에 저장되며, 시스템으로 유기적으로 공유가 가능하도록 구축하였다. 궁극적으로 통합DB는 서로 적절히 연계·분석해 산림경영계획 수립 및 정책의사결정에 활용할 수 있도록 목표를 수립하였다.

산림청은 데이터 품질관리를 위해 품질관리를 표준부분, 구조부분, 품질부분으로 세분화해 진행했다. 데이터 표준부분은 용어 표준화를 통해 유사한 의미를 가진 단어를 통합해 표준화시킴으로써 용어의 오남용을 방지하였다. 기존의 유사용어는 표준 용어와의 매핑을 통해 동일한 의미로 사용되고 있음을 표시하고 향후 표준 용어(표준 단어 3,234개, 표준 용어 7,146개)만을 사용하도록 정의하고 관리하도록 하였다.



(그림 4) 산림정보화 목표시스템 구성  
(출처: <http://www.dqc.or.kr/>)

표준 도메인 사전은 각 개별 소스시스템별로 혼재되어 사용되고 있는 컬럼을 분류하고 분석해 컬럼명, 데이터 타입, 길이 등도 표준화하였다. 이를 위해 표준 도메인을 87개를 만들었으며, 코드 표준화 작업은 현행코드를 분석해 코드 관리지침을 수립하였다. 관리대상 코드들은 취합하여 표준화 대상을 선정하고 통합코드로 생성하였다(표준 코드 401개). 이후에는 기존코드와 통합코드간의 변환규칙을 정립하고 통합코드 분류체계를 정립해 통합코드 저장소로 관리할 계획이다. 표준 코드는 코드목록·값을 대상으로 통합관리 코드와 개별관리 코드로 구분해 GCC 표준 코드를 우선 적용하고 표준화 지침을 바탕으로 표준 코드를 생성하였다.

데이터 구조부분은 산림청의 현행데이터 구성도의 경우 주요 산림업무와 시스템을 기준으로 산림환경보호, 산림생태관리, 산림재해관리, 산림이용 등과 같이 9개 데이터 주제영역, 17개 데이터그룹으로 구성하였다. 현행 데이터 구성도의

---

데이터 그룹별로 개념 데이터 관계도를 작성해 17개의 개념 데이터 관계도를 작성하고 운용중인 모든 정보화시스템의 ERD를 작성해 논리데이터·물리데이터 모델 현행화를 완료하였다. 데이터 참조모델은 범정부DRM 및 범정부 데이터와 매핑해 산림청의 데이터 참조모델을 수립하였다.

또한 현행 데이터 구성도와 산림정보화기본계획을 반영해 향후 산림청의 목표 데이터 구성도를 산림정책을 비롯해 산림생태관리, 산림재해관리, 산림이용 등과 같이 9개 데이터 주제영역, 19개 그룹으로 구성하였다. 목표 데이터 구성도는 데이터 그룹별로 개념 데이터 관계도를 작성해 19개 개념 데이터 관계도를 작성하고 산림생태 및 재해분야 등의 업무지원시스템과 대국민서비스를 위한 논리 데이터 모델에 대해 작성을 완료하였다.

데이터 품질부분은 우선, 데이터 품질을 높이기 위한 방안을 산림청의 현재 정보화 조직과 예산 등을 고려해 수행 가능한 방안을 수립하였다. 수립된 주요 방안은 다음과 같다.

- 품질기준, 핵심품질항목, 업무규칙 정의 등 값의 품질관리를 위한 측정기준 수립
- 핵심품질항목 기준과 업무규칙 기준 등 값의 품질관리를 위한 품질기준 수립
- 품질진단관리, 결과분석 관리, 품질개선관리 등 값의 품질관리를 위한 절차 수립
- 품질관리 체계를 위한 기능구조를 정의하고 기능별로 필요 직무를 도출하여 역할의 분담하고, 데이터 품질관리 활동의 승인과 통제역할은 청에 의해 이루어지고 수행조직에서는 데이터 품질관리 그룹에 의해 시스템 운영 및 개발 활동에 이해 발생되는 품질활동이 통제 관리되도록 조직(안)을 정의
- 표준관리 프로세스, 구조관리 프로세스, 품질관리 프로세스 정의

산림청의 데이터 품질관리를 위해, 데이터 표준관리는 전청 차원의 관심과 역할별 오너십을 재정 의하고 운영프로세스를 체계화해 지속적으로 관리가 될 수 있도록 하였다. 표준화 이행을 위한 핵심성과지표 및 변화관리 실행을 위해서는 연차별로 마련한 이행계획을 지속적으로 수립하였다.

수립된 이행계획을 실행하기 위해 산림청 전청차원의 데이터 표준화 지침을

---

정의하고, 지침을 이행할 수 있도록 산림청 표준화 훈령을 제정해 공포하고, 정보화사업 제안요청서에 표준화 준수를 명시하였다. 표준지침의 실행점검을 위해 청 내에 표준화위원회를 운영하고, 신규시스템 구축 및 고도화 사업 수행시 용어, 도메인의 표준준수 검수를 의무화하고 데이터 표준 관리 시스템을 통한 컨트롤타워를 구축하였다.

#### 다. 특허청의 데이터 품질관리

특허청이 데이터 품질관리를 시행하는 추진배경은 국제적 심사협력, 데이터 교환 등 국제협력 프로젝트를 주도하고 위해 우리나라 특허정보의 오류를 정비하고 표준화 하는 등 데이터의 고급화 필요, 원천DB로서 보급용 특허정보 가공 DB를 다양화·고급화하여 민간의 특허정보DB 공유와 개방 확대의 요구 증가, 특허정보의 활용 확산 및 대중화를 위해 특허정보 생산·유통체계 전반에 대한 점검을 통해 종합적인 발전대책 수립·시행의 필요 등이다.

특허청은 데이터 품질관리를 위해 맞춤형 데이터 품질관리 체계 수립하고 고품질의 특허 데이터 확보를 위한 미래지향적 품질관리 방안을 수립하였다. 2006년 데이터 품질관리 1차 사업을 수행하여 데이터 품질관리의 기반을 마련하고, 데이터 품질 체계를 마련하였다. 2009년 데이터 품질관리 체계 고도화(2차 사업)를 추진하여 데이터 품질관리의 의무화 수준 강화를 위해 정부기관 최초로 「특허청 데이터품질관리 규정(훈령 제643호)」를 제정(2009.11.)하여, 데이터 품질관리를 제도화 하였다. 2010년 데이터 품질관리 안정화와 문화조성을 위해 「특허청 데이터베이스 표준안」을 제정하여 특허청 정보화 사업 추진시 데이터베이스 표준 준수를 의무화하고, 전사적인 데이터 품질관리 문화정착을 위해 현업부서의 데이터 오너십 제도를 시행하였다. 또한, DB모델 정규화 및 표준화, 오류코드 정제 등 고품질의 특허정보를 운영할 수 있는 특허넷 시스템 구축하였다. 2012년에는 업무규칙(BR)을 대폭 확충하고, 특허행정 절차에 대응하는 데이터의 흐름을 관리하는 정보시스템을 구축하여 「특허데이터 품질관리 체계 선진화」 중장기 정보화 전략을 수립하였다.

특허청의 특허정보관리 업무 흐름을 보면 데이터 입수, 생성, 구축, 가공, 보급, 서비스의 6개 단계를 거치며, 데이터관리센터 등 특허정보원의 4개 조직에 위탁 운영 중에 있다.





(그림 5) 특허정보관리 업무 흐름도  
(출처: 특허청, 2012)

특허 데이터는 국내 출원 건은 특허청의 전자출원시스템(특허로)을 통해, 외국 출원 건은 해외 특허청에서 31개국의 지식재산권 데이터를 항공우편 등으로 연간 1천만 건을 입수하고 있다. 데이터 생성은 출원인이 서면으로 제출하는 출원서 등 문서를 특허넷에서 활용할 수 있도록 연간 44만여 건을 전자화하고 있다. 데이터 구축은 입수한 데이터의 오류를 정제하여 검색DB에 7개국의 데이터를 탑재하고 있다. 특허 DB에 구축된 데이터는 총 2억 916만 건으로 국내 3,706만 건(18%), 국외 1억 7,210만 건(82.4%)이고, 특허 종류별로는 특허·실용신안이 1억 6,610만 건(79%), 상표 631만 건(3%), 디자인 3,675만 건(18%)으로 구성되어 있다.

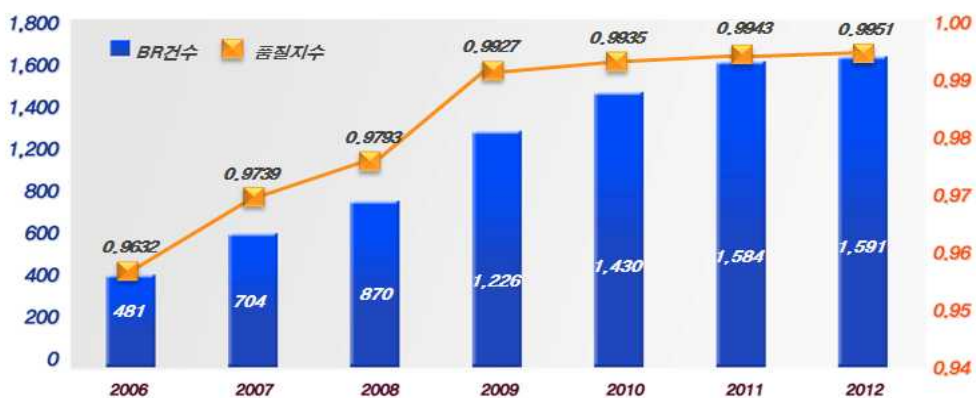
<표 10> 특허정보DB 구축 현황

('12. 8. 기준, 단위 : 만 건)

구분	특허·실용	상표	디자인	소계	비율
국내	702	631	2,373	3,706	18%
국외	15,908	-	1,302	17,210	82%
계	16,610	631	3,675	20,916	100%

출처: 특허청 (2012)

데이터의 정비는 정비담당자 및 정보시스템을 통해 오류데이터를 탐지·제거하고 있다. 특허넷에 존재하는 과거 공보데이터 오류를 수작업으로 정비하고 데이터오류 검증식(BR)을 통해 오류를 자동으로 검증하고 제거하고 있다. 데이터오류 검증식은 특허넷 시스템 데이터베이스(DB)에 저장된 데이터에 대한 특허 관련 법령, 규칙 및 절차 준수 여부를 전산 프로그램으로 검증하는 업무규칙을 말한다. 2002년부터 데이터관리센터를 통해 오류 정비 활동을 체계적으로 수행한 결과, 우리나라 특허정보의 품질이 3국 특허청 수준으로 개선되었다. 그 예로 산업재산권 공보데이터 오류율은 한국 특허청의 경우 0.013%, 미국 특허청의 경우 0.005%로 나타났다.



(그림 6) 데이터오류 검증식 확충 및 데이터품질 개선 현황

(출처: 특허청, 2012)

---

이렇게 DB로 구축되고 오류가 정제된 특허정보는 공보 발간과 영문초록 발간을 통해 가공된다. 공보 발간은 발명된 기술내용을 대중에게 공개하기 위해 공개공보, 등록공보 등 총 8종의 공보를 연평균 32만 건 발간하여 DB로 구축하고 있다. 영문초록 발간은 국내에 출원된 특허기술 정보를 해외에 알리기 위해 공개 및 등록 특허공보에 대해 영문초록을 발간하여 연간 13만 여 건의 DB를 구축하고 있다. 가공된 정보는 국내 지식재산서비스업체 및 외국 특허청 등에 보급되고 있으며, 민간에 KIPRIS<sup>1)</sup>로, 외국 특허청에 K-PION<sup>2)</sup>으로 서비스되고 있다. 그 결과 정보서비스 활용도가 지속적으로 증가하고 있으며, EPO의 한국특허문헌 인용순위도가 1998년 24위에서 2008년 9위로 상승하였다.

특허청의 데이터 품질관리의 성과로 다양한 프로그램, 절차 및 정책 등을 수립하여 전 직원의 참여 독려를 통해 전사적인 특허데이터에 대한 품질관리의식과 문화를 조성하였다. 지속적인 업무규칙(BR)을 도출·적용하고 오류 데이터 발생 원인을 차단하여 고품질의 특허데이터를 확보하고 오류 데이터를 예방하고 있다. 또한, 고품질 특허데이터 확보로 글로벌 데이터 표준화를 주도하고 특허정보를 제공하고 있다.

#### 라. 영국 British Telecom(BT)의 데이터 품질관리

British Telecom은 업무 처리에 대한 신속성이 강조되고, 정보를 신규 비즈니스 창출에 가장 중요한 자산으로 인식하기 시작하였다. 그러나 프로세스와 시스템 대응 미흡으로 정보 중심의 비즈니스로 변화하는데 여러 가지 문제점이 발생하였다. British Telecom의 프로세스는 조직 기능 중심으로 설계되어 있어, 프로세스가 원활하지 못하고 매우 복잡하여 전체적인 관점에서 파악이 불가능하였다. 전사적인 데이터 구조상 데이터의 생산, 사용, 관리를 위한 전사 차원의 표준이 부재하였다. 기존 IT 전략은 기존 시스템 개선을 통해 사용 효율을 높이는 것이나, 조직 중심의 개발로 시스템이 다양하고 매우 복잡하여 비즈니스 환경 변화에 신속하고 효과적인 대응이 미흡하였다. 조직 내 데이터의 흐름에 대한 이해 부족으로 저품질의 데이터가 입력되고, 이를 활용함으로써 업무처리 시

---

1) KIPRIS : 일반국민과 기업들에게 인터넷으로 특허정보검색서비스 제공

2) K-PION : 특허공보와 심사정보를 영어로 번역하여 외국특허청에 제공

---

---

문제가 발생하였다.

이러한 문제점의 해결책으로 British Telecom은 1998년 데이터 품질 프로젝트의 증가에 따라 이들 프로젝트를 감독하고 예산을 통제하기 위한 정보관리강화그룹(IMSG; Information Management Steering Group)을 구성하였다. 1999년에는 BT Wholesale(도매부문)이 정보관리강화그룹에 합류하면서 데이터 관리 조직이 확대되어 정보관리포럼(IMF; Information Management Forum)으로 전환되었다. 1999년, 정보 활용 극대화를 위한 정보화 정책서(BT Group Information Policy)를 발간하고 데이터 품질관리 수행 방법론을 개발하였다. 또한, 2000년에는 British Telecom의 CEO가 직접 주관하여 정보의 전략적 중요성을 공유하는 컨퍼런스를 개최하였다.

British Telecom의 데이터 품질관리에서 얻을 수 있는 시사점은 첫째, 업무 관점에서 품질 이슈를 공유함으로써, 데이터의 오너십을 갖는 협업 부서들이 적극 참여하고, 품질 문제와 관련된 이해관계자를 구분할 필요가 있다는 것이다. 둘째, 전사 차원의 품질관리를 위한 정보관리 포럼을 설립하여 정보관리 정책을 제시하고 품질관리 방법론 개발 등 일관성 있는 정책을 추진하여야 한다. 셋째, 데이터 품질관리를 상시적인 업무 프로세스로 이해시켜 전사차원의 데이터 품질의 지속적인 관리가 필요하다. 또한, 컨퍼런스나 브리핑 등 데이터 품질관리를 위한 지속적인 커뮤니케이션 활동 실시하고, 품질관리 관련 지식을 발굴하고 공유할 필요가 있다.

데이터 품질관리를 통해 얻은 성과로는 실패비용 절감, 재작업 감소로 인한 생산성 향상 등 주로 프로세스 개선을 통해 7년 동안 약 1.2조원 이상(7억 GBP)의 비즈니스 이익을 달성하였다. 데이터 정확도 향상에 따른 서비스 효율성 제고로 고객 만족도가 향상되었고, 자산 현행화, 제품 및 서비스 공급에 따른 청구서 발행 정확도도 향상되었다. 다양한 신규 채널을 통한 e-비즈니스 활성화, 프로세스 개선에 따른 업무의 효율성이 증대되었다.

---

# 제3장

## 국가 R&D 데이터 품질평가

---

1. 국가 R&D 데이터 품질평가 개요
2. 국가 R&D 데이터 품질평가 효과

---

# 1. 국가 R&D 데이터 품질평가의 개요

## 가. 국가R&D 데이터 품질평가의 필요성

최근 기관이나 기업내 IT(Information Technology)가 차지하는 역할과 비중이 증가하고 있다. 데이터 품질이 저하되면, 업무의 정확성과 효율성이 떨어지고 데이터 정제 등 추가 작업과 비용의 소요가 발생한다. 데이터 품질 저하로 발생하는 피해는 민간 기업의 경우 해당 기업이 책임을 지지만, 공공기관의 경우 국민이 피해를 받는다는 점에서 공공분야에서 데이터 품질관리가 더욱 중요시 된다.

이에 따라, 국가과학기술정보서비스(NTIS)도 고품질의 국가연구개발 데이터를 확보하기 위해 국가R&D데이터 품질관리체계를 수립하고 데이터의 품질기준과 품질진단, 관리 프로세스 등을 정의하고 있다. 국가R&D 인력/과제/성과정보 데이터 정제 지침 및 매뉴얼을 제정하여 NTIS 데이터 품질향상에 노력하고 있다 (손강렬·임종태, 2009). 또한, 연 2회 연계정보를 대상으로 데이터 품질평가를 실시 하여, 고품질의 데이터를 확보하고, 각 기관의 능동적인 업무 협조를 이끌어내고 있다.

## 나. 국가과학기술정보서비스의 정보 연계

국가과학기술정보서비스(NTIS)는 국가연구개발사업 관련 정보를 종합적으로 제공하는 세계 최초의 국가연구개발사업정보 지식포털이다. 2015년 현재, 사업·과제, 인력, 연구시설·장비, 성과 등으로 구성된 389개의 표준정보로 구성되어 있다. 국가연구개발사업을 관리하는 정부부처 등에서 수집한 표준 정보를 정제하고 가공하여 정부부처와 대학, 기업, 국가연구개발사업에 관심있는 대국민에게

다양한 서비스를 제공하고 있다(그림 7). NTIS가 서비스 중인 정보건수는 약 377.6만 건으로 국가연구개발사업의 연구과제정보 43만 건, 국가연구개발사업을 통해 창출된 성과물정보 310만 건, 국가연구개발사업에 참여한 연구인력정보 약 16만 명, 국가연구개발사업을 통해 취득한 취득가 3천만 원 이상의 연구장비정보 약 8.6만 점 등이 있다.



(그림 7) NTIS 개념도  
(출처: <http://www.ntis.go.kr>)

국가연구개발사업의 관리 등에 관한 규정(2016.07.01.)의 제25조(연구개발정보의 관리) 제4항에 따라, 연구관리 전문기관이 운영 중인 정보관리시스템을 NTIS과 연계하여야 한다. 연계 대상이 되는 정보는 추진 중인 국가연구개발사업의 연구개발과제 정보, 참여연구원 정보, 평가위원 및 평가 결과, 연구개발성과, 실패한 연구개발과제 정보 등 모든 연구개발정보에 해당한다. 동 규정 제25조 제5항에 따라, 연구시설·장비의 정보도 NTIS로 연계하여야 한다(미래창조과학부,

2016가). 422개 표준정보는 정보 분야를 기준으로 과제, 성과, 참여인력, 평가위원, 연구시설·장비로 나눌 수 있으며, 정보 분야별 상세 정보항목은 <표 11>와 같다.

<표 11> 국가연구개발정보표준 항목

정보 분야	정보항목	세부항목 수	정보 분야	정보항목	세부항목 수
과제 (150)	사업정보	13	참여 인력 (39)	경력사항	3
	기본정보	35		기본정보	6
	최종평가결과	10		논문실적	8
	평가위원	5		자격사항	3
	연구비정보	21		재직기관사항	6
	위탁과제/공동연구	18		지식재산권	3
	참여연구원정보	8		학력사항	10
	환수금 및 제재부 기금	27			
	연구개발공고	7			
	연차정보	6			
성과 (152)	연수지원	7	평가 위원 (11)	기본정보	1
	연수지원통계	4		소속학협회현황	3
	기술거래	15		저역서실적	4
	논문(게재성과)	19		훈포장사항	3
	사업화	11			
	지식재산권	13	연구 시설 · 장비 (70)	시설·장비일반	9
	인력양성(배출실적)	7		시설·장비취득정보	18
	생물자원(배출실적)	8		시설·장비활용	7
	생명정보(배출실적)	10		시설·장비관리	21
	신품중(배출실적)	16		시설·장비공동활용정보	4
	화합물(배출실적)	9		시설·장비연계정보	10
	소프트웨어(배출실적)	13		시설·장비심의	1
연구보고서(배출실적)	10				
기술요약정보	10				

출처: 미래창조과학부(2016나) 재구성

NTIS에서는 2008년부터 데이터품질 관리지표를 설정하여 지속적인 데이터품질 관리로 품질향상에 노력하고 있다. 범부처 국가연구개발표준정보 중 각 부처로부터 수집하는 항목을 2008년 248개에서 2016년 318개로, 70개 항목을 추가하였다. 연계 데이터의 목표 오류율을 2008년 4.6% 이하에서 2016년 0.5% 이하로 설정하여, 데이터의 품질을 높이고 있다. 또한 기관에서 연계 받는 데이터의 실시간 수집율을 2009년에는 60% 이상으로, 2016년에는 95% 이상으로 설정하였다(<표 12> 참조).



<표 12> NTIS의 데이터품질 관리지표

구분	'08	'09	'10	'11	'12	'13	'14	'15	'16
수집대상항목 (개)	248		247		263			284	318
데이터 오류율(목표)	4.6%	4.0%	2.3%	2.0%	1.7%	1.5%	1.2%	0.8%	0.5%
실시간 수집율(목표)	-	60%	65%	80%	85%	88%	90%	92%	95%
		이상	이상	이상	이상	이상	이상	이상	이상

#### 다. 국가 R&D 데이터 품질관리의 개관

NTIS는 17개 대표전문기관으로부터 국가연구개발정보표준 항목을 수집하고 있다. 그러나 각 대표전문기관마다 관리하는 데이터의 형태가 다르므로 공통된 표준으로 통합하고 높은 수준의 데이터 품질을 확보하는데 어려움이 있다. 또한 기관별 데이터 관리 정책이 다르고, 기관간 협력과 지원의 한계가 존재한다. 이러한 문제점을 극복하여 고품질의 데이터를 확보하기 위해 NTIS는 다양한 노력을 하여 왔다. 2007년과 2008년에는 데이터품질관리방안 마련을 목표로, 데이터품질관리체계 구축방안을 마련하고, NTIS 표준운영절차(SOP)를 수립하였다. 또한 데이터품질관리시스템(DQMS)과 콜센터, 데이터정제시스템 등을 도입하여 데이터 품질관리 성숙도 1레벨을 달성하였다. 2009년과 2010년에는 데이터품질관리체계의 정립을 목표로, 데이터품질관리체계를 정립하고 ISO27001 정보보호관리 인증을 받았다. 또한 데이터품질관리 교육을 실시하고 메타데이터관리시스템(MDMS)을 도입하였다. 이에 데이터 품질관리 성숙도 2레벨을 달성하였다. 2011년부터 2014년까지는 데이터품질관리체계 확산을 목표로, 메타데이터관리시스템을 통한 데이터의 표준과 구조, 흐름을 관리하고 전체 DB의 구조를 개선하였다. 데이터베이스 보안시스템을 도입하고 데이터 품질관리 성숙도 유효성 3레벨을 획득하였다.

<표 13> NTIS의 데이터 품질확보를 위한 노력

연도	'07 ~ '08년	'09 ~ '10년	'11 ~ '14년
목표	데이터품질관리방안 마련	데이터품질관리체계 정립	데이터품질관리체계 확산
주요활동	<ul style="list-style-type: none"> <li>· 데이터품질관리체계 구축방안 마련</li> <li>· NTIS 표준운영절차(SOP)수립</li> <li>· ISO20000(IT서비스 관리) 인증</li> <li>· 데이터품질관리지침 수립</li> <li>· NTIS 데이터품질평가 실시 (15개 대표전문기관)</li> </ul>	<ul style="list-style-type: none"> <li>· 데이터품질관리체계 정립 및 내재화</li> <li>· ISO27001(정보보호 관리) 인증</li> <li>· 데이터품질관리 교육실시</li> <li>· NTIS 데이터품질평가 실시 (16개 대표전문기관)</li> </ul>	<ul style="list-style-type: none"> <li>· 메타데이터관리시스템을 통한 데이터 표준, 구조, 흐름관리</li> <li>· 전체 DB 구조개선</li> <li>· 정보연계 모니터링 시스템을 통한 오류 사항 일일 피드백</li> </ul>
달성수준	· 품질관리 성숙도 1레벨(도입)	· 품질관리 성숙도 2레벨(정형화)	· 품질관리 성숙도 유효성 3레벨(통합화)
관련 인프라	<ul style="list-style-type: none"> <li>· 데이터품질관리시스템(DQMS)</li> <li>· NTIS 콜센터</li> <li>· 데이터정제시스템</li> </ul>	<ul style="list-style-type: none"> <li>· 메타데이터관리시스템(MDMS)</li> </ul>	<ul style="list-style-type: none"> <li>· 데이터베이스 보안 시스템</li> </ul>

NTIS 데이터품질관리의 관점에서 접근성, 신뢰성, 통합성, 활용성, 보안성의 5가지 핵심가치를 도출하고, 이에 따른 5가지 원칙을 정의하여 데이터관리를 수행하고 있다. 5가지 데이터품질관리 원칙으로는, ① 데이터 성능(국가R&D 데이터의 서비스성능 극대화), ② 데이터 품질(데이터의 정확성과 신뢰도 향상), ③ 데이터 통합(데이터 모델과 데이터의 통합관리), ④ 데이터 활용(사용자가 원하는 데이터를 활용할 수 있도록 관리), ⑤ 데이터 보안(데이터의 활용을 최대로 보장하되 보안성을 강화)이다.

NTIS 데이터품질관리 정책은 NTIS의 안정적 운영과 범부처와 효율적인 정보연계, 협력체계 마련을 위해 수립한 NTIS 표준운영절차(SOP), ISO20000/27001의 관련 프로세스를 기반으로 하고 있다. NTIS 표준운영절차는 계획과 조직, 운영

등으로 구분된다. 데이터품질관리 체계 프레임워크는 ① 요구사항 관리, ② 데이터 구조 관리, ③ 데이터베이스 관리, ④ 데이터 흐름 관리, ⑤ 데이터 표준 관리, ⑥ 데이터 활용 관리, ⑦ 데이터 오너쉽 관리, ⑧ 사용자 뷰 관리 등으로 구성되어 있다.

<표 14> NTIS의 데이터품질관리 체계 프레임워크

프레임워크 요소	개요	특징 및 장점
요구사항 관리	<p>사용자의 요구사항 관리는 ISO20000의 인시던트관리 및 변경 관리 프로세스를 준수 하고 있으며, NTIS 콜센터 시스템을 통해 종합 관리됨</p>	<ul style="list-style-type: none"> <li>· 일반사용자 요구는 ISO20000 프로세스 준수</li> <li>- 사용자의 콜, 이메일, 홈페이지를 통한 요구사항 접수는 콜센터 시스템을 통해 접수 및 기록되며, 우선순위에 따라 담당자 처리</li> <li>- 처리 후 결과를 문자를 통해 사용자에게 피드백</li> <li>· 연계기관의 단순한 요구는 업무담당자가 대응</li> <li>- 업무 담당자 처리 후 콜, 이메일을 통한 피드백 수행</li> <li>- 처리 결과를 콜 센터에 기록 유지</li> <li>· 복잡한 요구는 선별 후 개발 프로젝트 추진</li> <li>- 단순 대응이 불가능한 복잡한 요건의 경우, 차년도 사업계획에 반영</li> <li>- 개발프로젝트 추진 시 요구사항 정의서 및 추적표 등 개발산출물로 관리되고, 개발완료 후 검수를 통해 반영 확인</li> <li>· (인프라) 콜센터 시스템</li> </ul>
데이터 구조 관리	<p>데이터 구조관리는 메타데이터관리시스템 기반 단계별 승인절차를 통해 데이터모델과 물리적 데이터베이스를 변경함</p>	<ul style="list-style-type: none"> <li>· 구조관리 프로세스의 시스템화</li> <li>- 메타데이터관리시스템(MDMS)을 통해 모든 구조변경 절차를 시스템화하고, 모델구조 변경영향 및 중복확인</li> <li>· 논리모델과 물리모델 일치성 확보</li> <li>- 구조관리자가 모델의 추가, 변경, 삭제 등을 검증 &amp; 승인</li> <li>- DBA는 메타데이터관리시스템을 통해 승인된 것만 DB 생성</li> <li>· (인프라) 메타데이터관리시스템 (MDMS)</li> </ul>

프레임워크 요소		개요	특징 및 장점
데이터베이스 관리	보안	데이터베이스 보안관리는 ISO20000(DB접근권한 관리)와 ISO27001(정보 시스템 보안관리) 프로세스를 준수함	<ul style="list-style-type: none"> <li>· DBA에 의해 통제, 계정 및 권한 기록관리</li> <li>- DB 접근 로그(아카이브 로그) 기록관리, 사고 발생시 원인추적 및 복구가능</li> <li>· 보안관리는 ISO20000/27001 절차준수</li> <li>- 정기적인 보안점검과 보안감사 실시, 정기적인 보안교육 실시</li> <li>- 침해사고 대응 훈련을 통해 취약점 도출 및 취약점 개선</li> <li>· (인프라) 암호화시스템(D'Amo), 접근제어시스템(DBSAFER)</li> </ul>
	백업 / 복구	데이터베이스 백업 및 복구관리는 ISO20000 (서비스연속성 관리) 프로세스 준수	<ul style="list-style-type: none"> <li>· 백업계획을 수립 및 DB복구 모의 훈련 수행</li> <li>- 서비스 등급에 따라 보안등급을 적용하여 재해복구 우선순위 선정</li> <li>- 정해진 스케줄대로 백업시스템에 의해 정기적으로 백업 수행</li> <li>· Daily: 1일 4회, 2주분 보관</li> <li>· 풀백업: Full(주 3회) / Incremental (주 4회) 포함 3일 분량 보관</li> <li>· TAPE : 월 1회 , 분산보관(서울분원)</li> <li>· (인프라) 백업시스템</li> </ul>
	성능	데이터베이스 성능 관리는 ISO20000 용량 관리 프로세스에 따라 이루어지며, 성능관리 시스템을 통해 상시 모니터링하고, 성능 이슈에 대한 개선활동을 수행하고 있음	<ul style="list-style-type: none"> <li>· 서비스 성능의 ISO20000 절차 준수</li> <li>- ISO20000의 프로세스에 따라 서비스에 대한 성능관리의 절차 및 역할 정의</li> <li>- 서비스에 대한 성능 목표를 3초 이내로 정의하고, 각 서비스에 대해 제니우스 툴을 통해 실시간으로 성능을 측정</li> <li>- 월 단위로 측정결과를 보고서로 작성</li> <li>· 외부 전문가를 통한 성능 튜닝 수행</li> <li>- 부분적 성능개선을 위한 전문가 활용</li> <li>- 위탁연구를 통해 NTIS 시스템 최적화 방안 연구 수행</li> <li>· (인프라) 제니우스</li> </ul>

프레임워크 요소	개요	특징 및 장점
데이터 흐름관리	데이터 흐름관리는 메타데이터관리시스템을 통해 변경관리가 이루어지며, 데이터 수집에서 서비스까지 정보연계모니터링시스템으로 데이터의 흐름을 파악함	<ul style="list-style-type: none"> <li>· 데이터 수집부터 서비스까지 모니터링 수행</li> <li>- 정보연계모니터링시스템을 구축하여 수집, 정제, 관리, 서비스DB로 이어지는 주요 데이터흐름 확인</li> <li>- 상황판으로 수집건수, 오류건수, 처리건수 등 전체적인 데이터 흐름 파악</li> <li>· 소스-타겟 매핑 시스템화</li> <li>- 소스-타겟 매핑 정의서를 메타데이터관리시스템으로 등록, 변경 관리 수행</li> <li>- 데이터모델 변경 시 흐름 관리에 의한 변경영향분석 수행</li> <li>· (인프라) 정보연계모니터링시스템, 메타데이터관리시스템(MDMS)</li> </ul>
데이터 표준 관리	국가R&D표준정보는 매년 국가과학기술심의회(국과심)에서 확정·배포함	<ul style="list-style-type: none"> <li>· 국가R&amp;D표준정보 개정 및 배포</li> <li>- 매년 국가R&amp;D표준정보(389개)를 연계기관과 협의하여 여러 협의회(실무자 협의회, 전문가 협의회, 국가연구개발정보관리위원회)를 거쳐 국과심에서 확정 및 배포</li> </ul>
	데이터 표준관리는 데이터에 대한 공통된 시각을 유지할 수 있도록 데이터 표준화 원칙에 따라 관리함	<ul style="list-style-type: none"> <li>· 데이터 표준관리 프로세스의 시스템화</li> <li>- 메타데이터관리시스템을 활용하여 표준 추가, 변경, 삭제 수행</li> <li>- 표준의 변경영향, 표준준수 여부, 비표준매핑 등을 시스템으로 관리</li> <li>· 표준관리자가 데이터 표준관리</li> <li>- 표준관리자가 데이터 표준(단어, 용어, 코드, 도메인) 검토 및 승인</li> <li>· (인프라) 메타데이터관리시스템 (MDMS)</li> </ul>

프레임워크 요소		개요	특징 및 장점
데이터 활용 관리	품질 관리	데이터 활용 극대화를 위해 연 2회 데이터 품질평가 실시	<ul style="list-style-type: none"> <li>· 데이터 품질평가</li> <li>- 연계 기관별 수집데이터 점검 및 오류 결과 피드백(일일 피드백)</li> <li>- 수집된 데이터의 품질수준을 평가하여 시상</li> <li>※ NTIS 데이터 품질평가(연 2회) 실시('08년~ )</li> <li>· 점검규칙(BR) 작성 및 통합관리</li> <li>- 데이터를 완전성, 유일성, 유효성, 일관성 등 다각적인 측면에서 점검하기 위한 점검규칙(BR) 작성 및 통합관리</li> <li>- 정보연계모니터링시스템, 데이터정제 시스템, 데이터품질관리 시스템 등에 점검규칙(BR) 활용</li> <li>· (인프라) 데이터품질관리시스템 (DQMS) 데이터정제시스템</li> </ul>
	모니터링	데이터 활용 모니터링은 ISO20000 고객관계관리 프로세스에 따라 수행 정기적으로 고객불만 사항(VOC)과 고객 만족도를 조사하여 서비스 개선에 활용	<ul style="list-style-type: none"> <li>· 고객불만사항(VOC) 조사</li> <li>- 콜센터를 통해 등록된 사항의 처리결과에 대한 사용자 만족도 입력</li> <li>- 서비스 오픈 전 시범운영을 통해 사용시 불편사항 및 개선내용 반영</li> <li>· 고객만족도 조사</li> <li>- NTIS서비스 전반에 대한 사용자 만족도 조사(연 1회)</li> <li>- 사용자 불편사항 및 원인분석 수행</li> <li>· (인프라) 콜센터 시스템</li> </ul>
데이터 오너쉽 관리		데이터 오너쉽 관리는 서비스 영역별로 담당자를 지정하고, 담당자별 데이터 오너쉽 부여 각 서비스별 데이터 관리지침을 정의하여 데이터관리 일관성 유지	<ul style="list-style-type: none"> <li>· 세부 프로세스명</li> <li>- 서비스 영역별로 담당자가 정해져 있고, 담당자별 데이터에 대한 오너쉽이 정해져 있음</li> <li>· 원천 데이터 추적 가능</li> <li>- 연계기관 별로 수집된 데이터에 대한 문제 발생 시 원천 데이터 추적가능</li> <li>· (인프라) NTIS SOP</li> </ul>

프레임워크 요소		개요	특징 및 장점
사용자 뷰 관리	사용자 교육	사용자교육 관리는 ISO20000/27001 교육 훈련 프로세스에 따라 사용자 교육실시 데이터에 대한 접근 장벽을 최소화하기 위해 사용자에게 데이터를 이용하는 환경에 대한 정보관리 교육, 품질 관리 교육 등을 실시함	<ul style="list-style-type: none"> <li>· 내부 사용자교육 수행</li> <li>- ISO20000/27001 공통 교육훈련 프로세스에 따라 수행</li> <li>· 외부 사용자교육 수행</li> <li>- 각 부처 및 과제관리기관(비연계) 등에 NTIS서비스 시연</li> <li>- NTIS 정보관리 교육(과제관리기관 실무자 및 개발자 대상)</li> <li>- 연계기관 데이터품질향상을 위한 맞춤 교육</li> <li>- 홈페이지 교육 동영상, 사용자매뉴얼 및 기타 팜플렛 등 제공</li> </ul>
	문서 관리	문서관리는 ISO20000/27001 문서관리 및 기록관리 프로세스에 따라 수행 정의된 표준에 따라 사용자 뷰가 설계 및 개발되고, 개발표준 지침서를 제공함	<ul style="list-style-type: none"> <li>· 문서관리 및 기록관리</li> <li>- ISO20000/27001 공통 문서관리 및 기록 관리 프로세스에 따라 수행</li> <li>· 정의된 표준에 따른 화면 및 시스템 개발</li> <li>- NTIS 화면 표준정의서, 웹 스타일가이드</li> <li>- 개발표준지침서에 따라 개발/운영 및 검증</li> <li>- 통합TF에서 화면표준에 대한 협의 과정을 거쳐 갱신 및 배포하고, 개발 완료 시 검수를 통해 화면표준 준수여부 확인</li> </ul>

## 라. 국가 R&D 데이터 품질평가의 개관

NTIS는 2009년 15개 대표전문기관을 대상으로 연계 데이터의 품질을 평가한 이래로 현재는 17개 대표전문기관을 대상으로 품질관리를 수행하고 있다. 데이터 품질평가를 실시한 목적은 신뢰성 높은 고품질의 정보를 대국민에게 서비스하기 위해 정보가 최초로 생산되는 각 대표전문기관의 데이터 품질 수준을 제고하고자 함이다. 데이터품질관리의 성숙도 향상을 위해 품질평가를 정량적 평가와 정성적 평가로 나누어 실시하며, 정량적 평가는 연 2회, 정성적 평가는 연 1회 실시한다.

---

정보공유와 시스템 연계·통합의 경우, 정보보유기관의 입장에서 정보제공을 통해 이익이 얻을 수 없다는 점에서 수동적인 자세를 보이게 되며, 일반 정책 업무 담당자들은 해당 업무가 본인의 직접 업무와는 관련이 없는 부가적 업무로 판단하고 업무부담만 가중한다는 점에서 부정적 태도를 취할 우려가 있다. 정보공유와 시스템 연계·통합 업무와 관련된 다양한 업무 주체들의 적극적 참여를 이끌어 내기 위해 인센티브 제공을 통한 동기유발이 필요하다(이민호·엄석진·김명진, 2014). 이와 같은 이유로 범 부처 차원의 품질점검과 평가 후 각 기관별 품질제고 노력에 대한 인센티브로써, 우수기관에게 미래창조과학부 장관 표창을 수여하고 있다.

이러한 노력으로 2014년 현재 NTIS의 데이터 품질관리 성숙도는 활용성 측면에서 2레벨, 유효성 측면에서 3레벨을 유지하고 있다. 여기서 데이터 품질관리의 성숙수준이라 함은 조직이 수행 중인 데이터 품질관리의 체계화 정도를 나타낸다. 성숙수준은 도입, 정형화, 통합화, 정량화, 최적화의 5단계로 구분되며, 성숙수준이 높을수록 체계적이고 정교한 데이터 품질관리가 수행되고 있음을 의미한다(데이터베이스품질인증, 2015).

#### 마. 국가 R&D 데이터 품질평가의 지표

국가 R&D 데이터 품질평가는 정량적 평가와 정성적 평가로 나뉜다. 정량적 평가는 다시 품질측정 부문과 정보연계 부문으로 나누어진다. 2015년을 기준으로 각 지표들에 대해 알아본다. 먼저, 품질측정의 평가 대상이 되는 기준항목은 품질제고 우선순위와 부처별 보유항목 등을 고려하여 정해진다. 정해진 기준항목을 대상으로 데이터의 완전성, 유효성, 일관성을 지표로 평가한다. 완전성 지표는 업무적으로 필요한 데이터가 누락없이 제공되는 여부를 의미하며, 물리적 Null 값의 완전성과 업무적 Null 값의 완전성을 세부 기준으로 정하고 있다. 유효성 지표는 업무적으로 의미있는 데이터를 제공하는 여부를 의미하며, 도메인과 데이터 값의 일치성과 유효범위와 데이터 값의 일치성을 세부 기준으로 한다. 일관성 지표는 동일한 데이터가 서로 일치하는지를 평가하며, 중복 컬럼 데이터 값의 일관성과 기준코드와 데이터 값의 일관성을 세부 기준으로 한다. 정보연계 부문은 기관이 실제로 송신한 연계항목 수, 기관이 목표로 하는 정보연계 칼럼



수와 실제 연계한 칼럼 수의 비율로 산출되는 데이터 충실도와 과제정보 제공을 지표를 사용하여 평가된다. 과제정보 제공율은 과제 협약현황 자료를 매월 제출하였는가와 과제정보를 과제 협약 이후 며칠 이내에 연계하였는지를 평가한다. 정성적 평가는 부처별 품질향상 노력과 정보연계 및 과제신청창구 일원화 협조도, 기관 특성화 자율지표로 나누어 평가한다. 기관 특성화 자율지표는 기관의 데이터 제공 상황 특성, 데이터품질·데이터보안·개인정보보호 인증 등을 반영한다. <표 15>은 2009년부터 2016년까지 평가에 적용된 지표들을 정리하였다.

<표 15> 데이터 품질평가 지표

지 표		'09	'10	'11	'12	'13	'14	'15	'16	
정 량 적 평 가	품질 측정	품질평가 기준항목 수	91	94	94	121	121	121	121	145
		완전성	○	○	○	○	○	○	○	○
		유효성	○	○	○	○	○	○	○	○
		일관성	○	○	○	○	○	○	○	○
	정보 연계	연계항목 목표 달성도	연계 항목 수	-	-	○	-	○	○	○
			데이터 충실도	○	○	○	○	○	○	○
		과제정보 제공율	-	○	○	○	○	○	○	○
정 성 적 평 가	부처 협조 및 노력도		○	-	-	-	-	-	-	
	정보연계 및 국가연구개발 조사·분석 입력협조도		-	○	○	○	○	○	-	-
	정보연계 및 과제신청창구 일원화 협조도		-	-	-	-	-	-	○	-
	부처별 품질향상 노력		-	○	○	○	○	○	○	○
	기관 특성화 자율지표		-	-	-	-	-	-	○	-

정량적 평가 지표는 일부 지표를 제외한 나머지 지표들이 매년 동일하게 적용되었다. 정성적 평가 지표는 2009년 부처 협조 및 노력도 지표가 2010년 정보연계 및 국가연구개발 조사·분석 입력협조도 지표와 부처별 품질향상 노력 지표로 나뉘었으며, 2015년에는 정보연계 및 국가연구개발 조사·분석 입력협조도 지표가 정보연계 및 과제신청창구 일원화 협조도 지표로 바뀌었다. 기관 특성화 자율지표는 2015년에 신설되었다. 2016년에는 기타로, 참여제한, 공고, 과제 참여연구원 정보의 연계여부를 점검하는 지표가 신설되었다.

<표 16>은 2009년부터 2016년까지의 데이터 품질관리 지표별 배점을 나타내었다.

<표 16> NTIS의 데이터 품질평가 지표별 배점

지 표		'09	'10	'11	'12	'13	'14	'15	'16	
정 량 적 평 가	품질측정	45	45	35	35	35	30	30	20	
	연 계 항 목 목표달성도	연계항목 수	-	-	5	-	5	5	5	5
		데이터 충실도	25	15	10	10	5	5	5	5
	과제정보 제공율	-	20	30	35	35	35	35	20	
	기타 (참여제한, 공고, 과제참여연구원 정보 제공)	-	-	-	-	-	-	-	10	
	계	70	80	80	80	80	75	75	75	
정 성 적 평 가	부처 협조 및 노력도	30	-	-	-	-	-	-	-	
	정보연계 및 국가연구개발 조사·분석 입력협조도	-	10	10	10	10	10	-	-	
	정보연계 및 과제신청창구 일원화 협조도	-	-	-	-	-	-	10	-	
	부처별 품질향상 노력	-	10	10	10	10	10	10	25	
	기관 특성화 자율지표	-	-	-	-	-	5	5	-	
	계	30	20	20	20	20	25	25	25	
총 계		100	100	100	100	100	100	100	100	

---

품질측정 지표의 배점은 2010년 45점에서 2015년 30점으로 15점 낮아졌으나, 과제정보 제공율 지표의 배점은 2010년 20점에서 2015년 35점으로 15점 높아졌다. 품질측정 지표의 배점의 하향으로 보아 데이터의 품질이 점차 안정화되고 있음을 반영하고 있다고 볼 수 있다. 반면 과제정보 제공율 지표의 배점의 상향으로 보아 각 기관에서의 과제정보의 전송 시기를 중요한 평가 요소로 고려하고 있음을 알 수 있다. 정성적 평가의 정보연계 관련 지표를 보면, 국가연구개발 조사·분석 입력협조도 지표가 과제신청창구 일원화 협조도 지표로 전환되었음을 알 수 있다. 이는 국가연구개발 조사·분석자료의 입력 및 자료연계가 안정화에 도달하였음을 반영하는 것으로 판단된다.

## 2. 국가 R&D 데이터 품질평가의 효과

국가 R&D 데이터 품질관리·평가의 효과는 크게 정량적 효과와 정성적 효과로 나누어 볼 수 있다. 정량적 효과는 데이터 오류율과 과제정보 제공률 등 가시적 성과로 생각할 수 있고, 정성적 효과는 NTIS의 사용자 서비스 만족도와 고객가치 조사 결과 등 NTIS 이용자 관점의 평가결과로 볼 수 있다.

먼저 정량적 효과에 대해 알아본다. <표 17>는 2009년부터 2015년까지의 전체 수집 데이터의 오류율과 적시 과제정보 제공율의 추이를 나타낸다. 전체 수집 데이터의 오류율은 2009년 2.30%에서 지속적으로 감소하여 2015년 1.00%까지 감소하였다. 반면 적시 과제정보 제공율은 2009년 65.0%에서 2015년 92.0%까지 지속적으로 상승한 것을 볼 수 있다. 또한 연도별로 하반기 품질평가의 결과를 분석한 결과, 과제 데이터의 연평균 오류율은 2012년 0.84%, 2013년 0.06%, 2014년 0.04%로 감소하였다. 이는 전체 수집 데이터의 오류율보다 적은 수치를 가리키고 있다.

<표 17> NTIS의 전체 수집 데이터 오류율과 적시 과제정보 제공율 추이

구 분	'09	'10	'11	'12	'13	'14	'15
전체 수집 데이터 오류율 (단위: %)	2.30	1.75	1.70	1.54	1.37	1.10	1.00
적시 과제정보 제공율 (단위: %)	65.0	70.0	80.0	85.0	88.0	90.0	92.0

다음으로 정성적 효과에 대해 알아본다. <표 18>는 NTIS의 사용자를 대상으로 조사한 서비스 만족도 점수의 추이를 나타낸다. 2009년 70점대에 머물던 서비스 만족도 점수가 2013년 최초로 80점대를 넘어섰고, 2015년에는 82점을 획득하였다.

<표 18> NTIS 서비스 만족도 추이 (단위: 점)

'09	'10	'11	'12	'13	'14	'15
75.7	76.5	77.1	78.9	80.1	81.2	82.0

한편, 2014년 NTIS 정보이용자를 대상으로 고객가치를 조사한바 있다. 연구자와 연구관리자를 대상으로 사전에 정의된 지표를 이용하여 고객가치를 측정하였다. <표 19>는 연구자 대상 측정결과이고, <표 20>는 연구관리자 대상 측정결과이다. 연구자의 경우 '보고서 작성시간 절감', '연구아이디어 탐색시간 절감', '연구개발소요비용 절감', '연구 질적 수준 제고'측면에서 상대적으로 높은 점수를 얻었다. 아울러 2013년도와 비교해보면 대체로 점수가 높아졌으며, 특히 '연구개발 소요시간절감', '연구장비 활용비용 절감'측면에서 거의 두자리 수의 비율로 향상되었다. 반면에 '연구 질적 수준 제고'와 '과제성공률 제고'측면은 미세하지만 점수가 다소 더 낮아짐을 보였다. 연구관리자의 경우 '유사과제 탐색시간 절감'측면에서 가장 높은 평가를 받았으며, 아울러 전년도에 비해 가장 긍정적으로 변화하였다. 반면, '논문, 특허출원 질적 향상'측면은 점수가 다소 더 낮아짐을 보였다.

<표 19> 연구자 대상 NTIS의 고객가치 지표 항목별 변화('13~'14)

지 표		'13	'14	변화(%)
경제적 효과(8)	연구아이디어 탐색시간 절감	3.64	3.63	-0.27
	<b>보고서 작성시간 절감</b>	<b>3.60</b>	<b>3.87</b>	<b>+7.50</b>
	시간 노력대비 높은 가치 획득	3.52	3.58	+1.70
	전문가 탐색시간 절감	3.47	3.58	+3.17
	외부정보 사용료 절감	3.31	3.39	+2.42
	<b>연구개발 소요시간 절감</b>	<b>3.20</b>	<b>3.51</b>	<b>+9.69</b>
	<b>연구장비 활용 비용 절감</b>	<b>3.21</b>	<b>3.50</b>	<b>+9.03</b>
	연구개발 소요비용 절감	3.54	3.61	+1.98
성과제고 효익(6)	<b>연구 질적 수준 제고</b>	<b>3.64</b>	<b>3.58</b>	<b>-1.65</b>
	개발애로사항 해결	3.44	3.44	+0.00
	과제 성공률 제고	3.42	3.39	-0.88
	논문특허의 양, 질적 향상	3.37	3.38	+0.30
	기술개발수준 향상	3.36	3.35	-0.30
	기술사업화 제고	3.29	3.31	+0.61
평 균		3.43	3.51	+2.3

<표 20> 연구관리자 대상 NTIS의 고객가치지표 항목별 변화('13~'14)

지 표		'13	'14	변화(%)
경제적 효과(5)	유사과제 탐색시간 절감	3.73	3.87	+3.75
	시간 노력대비 높은 가치 획득	3.54	3.61	+1.98
	과제관리 의사결정	3.49	3.56	+2.01
	장비탐색 및 관리시간 절감	3.45	3.42	-0.87
	연구관리 소요비용 절감	3.41	3.48	+2.05
성과제고 효익(2)	논문, 특허출원 질적 향상	3.31	3.22	-2.72
	논문, 특허출원 양적 향상	3.21	3.27	+1.87
평 균		3.45	3.49	+1.2

연구자와 연구관리자의 고객가치를 수치로 계산하기 위해, 다음과 같이 총 점수의 평균치를 산출하였다.

연구자: 총점 / 14(항목수)

연구관리자: 총점 / 7(항목수)

산출식에 의거하여 산출한 평균치를 원래의 5점 척도로 나눈 후 100을 곱하여 지수 설정한 결과, 연구자의 고객가치는 70.2점, 연구관리자는 69.8점으로 나타났다.

---

# 제4장

## 국가 R&D 데이터 품질평가 개선 방향

---

이번 장에서는 국가 R&D 데이터 품질관리와 평가에 대한 개선 사항에 대해 알아보려고 한다. NTIS의 DB는 17개 부처·청의 연계 정보로 구축되므로, 각 기관들의 능동적이고 높은 정확도의 정보연계가 필수적이다. NTIS는 데이터 품질관리를 위해 데이터 품질평가를 연 2회 실시하는 등 데이터 품질개선 활동을 하고 있다. 또한 연계기관들이 참여하는 국가연구개발정보관리위원회, 전문가 협의회, 실무자 협의회 등 여러 협의회를 개최하고, 연계기관의 실무자들이 참여하는 워크숍 등을 개최하여, 데이터 품질의 중요성을 강조하고 공유하고자 노력하고 있다.

하지만, 적은 예산과 인력으로 R&D관리시스템을 운영하고 있는 일부 일선 기관에서는 데이터 품질관리에 부담을 느끼고, 관련 업무가 부수적 업무라 생각하여 후순위로 밀리기 마련이다. 데이터 품질평가의 정량적 평가결과를 종합하여 분석하여 보면, <표 21>과 같다. <표 21>은 데이터 품질관리 지표별 도달률을 나타내었다. 품질측정 지표는 2012년 75.13%에서 2014년 98.82%로 향상되었다. 연계항목 목표달성도 중 연계항목 수 지표는 2012년 77.65%에서 2014년 97.65%로 향상되었고, 데이터 충실도 지표는 2012년 80.59%에서 2014년 97.06%로 향상되었다. 그러나 과제정보 제공을 지표는 타 지표와 달리 2012년 83.19%에서 2014년 91.67%로 도달률의 상승폭이 높지 않다. 이는 각 연계기관에서 과제 협약 후 신속히 NTIS로 연계하지 않은 문제점을 나타낸다. 또한, 연계기관에서 데이터 실시간 연계 부분의 필요성과 중요성을 크게 인식하지 않고 있음을 반증한다.

<표 21> 데이터 품질관리 지표별 도달률(정량적 평가)

(단위: %)

지 표		'12 도달률	'13 도달률	'14 도달률	평균 도달률
품질측정		75.13	99.41	98.82	91.12
연계항목 목표달성도	연계항목 수	77.65	91.76	97.65	89.02
	데이터 충실도	80.59	91.18	97.06	89.61
과제정보 제공율		83.19	88.99	91.67	87.95



---

이러한 점들을 근거로 현 국가 R&D 데이터 품질관리와 평가에 대한 개선 방안을 제시하고자 한다.

첫째, 각 정보연계 부처·청과 데이터 품질 이슈와 중요성의 공유가 필요하다. NTIS의 데이터품질관리 방식은 각 기관에서 발생한 오류를 NTIS에서 발견하고 이를 해당 기관에서 수정하여 재연계하는 방식이다. 이러한 방식을 통해 NTIS의 데이터품질이 향상되었다면, 각 기관의 데이터품질 역시 향상되었을 것이다. 하지만, 지금까지의 데이터품질향상 홍보는 NTIS 중심으로 이루어졌다. 각 정보연계 부처·청에서 데이터 품질관련 업무가 중요한 업무라 인식되기 위해서는, 각 기관의 입장에서 데이터 품질관리를 통해 얻을 수 있는 긍정적 효과를 강조하여야 한다. 예를 들어, 데이터 품질관리를 통해 예상되는 경제적 효과를 계량적 수치로 언급하거나, 업무의 효율성 향상, 통계자료의 정확성 향상 등을 언급해야 한다.

둘째, 일선 부처·청에 데이터 품질관리 노하우를 공유하여야 한다. 그간, NTIS가 데이터 품질확보를 위해 어떠한 노력을 했고, 어떠한 방식으로 품질관리를 하면, NTIS의 수준으로 도달할 수 있는지 전달해야 한다. 또한 각 기관의 데이터 오류 등 NTIS가 보유한 품질관리지표별로 분석하여 어떠한 부분이 잘 이루어져있고, 어떠한 부분에서 결점이 발생하였는지 등의 자료를 전달하여야 한다.

셋째, 실시간 수집률을 제고하기 위한 방안을 수립하여야 한다. 매년 지표별 구간을 상향하여, 2014년 기준, 협약 후 1일 이내에 과제정보를 연계하도록 하고 있다. 하지만, 실시간 수집률이 크게 향상되지 않고 있어, 평가 방법 개선의 필요성이 존재한다. 현장에서 과제정보를 실시간으로 연계할 수 없는 이유나 여건 등을 조사하여, 개선방법을 모색하여야 할 것이다. 또한, 관련 법령을 개정하여 과제 정보의 실시간 수집에 대한 법적근거를 마련해야 할 것이다. 현재 ‘국가연구개발사업의 관리 등에 관한 규정(2015.01.01.)’을 보면 제25조 제18항에는 ‘중앙행정기관의 장은 연구개발정보를 국가과학기술종합정보시스템에 즉시 제출할 수 있는 수집·연계 체계를 구축하여야 한다.’고 명시되어 있다. 허나, 반드시 연구개발정보를 즉시 연계하여야 한다는 강제조항은 아니다. 또한, 동 규정의 제25조 제5항에는 국가연구개발사업을 통해 취득한 연구시설·장비를 취득 후 30일 이내 국가과학기술종합정보시스템에 등록·관리하도록 하고 있으나, 동

---

---

규정은 연구시설·장비 이외의 정보들에 대한 연계기한을 명시하고 있지 않다. 따라서 연구개발정보의 연계 기한을 명시하여 강제할 수 있는 조항을 신설해야 한다.

넷째, 데이터 품질평가의 지표를 신설 및 변경의 필요가 있다. <표 21>에서 이미 언급한 바와 같이, 운용 중인 지표들의 도달률이 90%를 상회하고 있어 지표의 변별력이 하락하였으므로, 이에 대한 개선이 필요하다. 이를 위해, 2015년 6월 현재, NTIS에서는 데이터 품질평가 지표의 개선을 아래의 <표 22>와 같이 논의 중에 있다. 일부 항목을 개선하거나 삭제하였다. 또한 새로운 지표를 선정하였다. 먼저, 정량적 평가와 정성적 평가의 배점이 조정되었다. 정량적 평가는 75점에서 80점으로, 정성적 평가는 25점에서 20점으로 조정되었다.

두 평가 중 먼저 정량적 평가에 대해 알아본다. 데이터오류율은 과제 오류율과 인력 오류율을 구분하여 평가한 기존 방식과 달리, 과제와 인력을 통합하여 평가하며 배점을 하향조정하였다. 이는 데이터오류율의 관리수준이 일정수준에 도달하여 변별력을 상실하였으나, 본 지표의 중요성을 배제할 수 없어, 지표를 유지한 채 배점만 하향조정한 것으로 보인다. 다음으로 데이터 서비스 달성을 지표가 신설되었다. 최초 수집에서 최종 서비스까지 데이터가 연계됨을 평가한다. NTIS홈페이지에서 표출되는 정보는 최종 서비스 단계까지 도달한 정보로 오류가 없는 정보를 의미한다. 오류가 발생한 정보는 NTIS에 수집되었더라도, 서비스로 이어지지 않으므로 그 중요성을 강조하기 위한 지표로 풀이된다. 연계항목 목표달성도는 데이터 입력 항목 수 및 데이터 충실도로 지표가 변경되었으나, 배점은 동일하다. NTIS에서 정한 121개 연계목표 항목 수 대비 실제 연계된 항목수를 평가한 기존의 지표와는 달리, 데이터 입력 항목을 155개로 확대하였다. 기존 121개 연계목표 항목 수는 몇 해 동안 변경되지 않은 항목 수로, 기존 항목을 신규 항목으로 대체하거나 신규 항목을 추가한다는 측면에서 긍정적으로 보인다. 과제정보 제공률은 전과 동일하나 배점이 30점에서 20점으로 하향 조정되었다. 또한 참여제한 정보 제공, 공고정보 제공율, 과제 참여연구원 정보 제공 지표가 신설되었다. 참여제한 정보 제공 지표는 국가연구개발사업·과제 수행에 있어 제재를 받아 일정기간동안 국가연구개발사업·과제를 중단하거나 신규 수행이 불가능한 자 혹은 기관에 대한 정보를 신속히 NTIS로 연계하였는가를 평가한다. 동 지표를 통해, 현재보다 빠른 시일 내에 최신 제재자 정보를

---

---

기관에서 확인할 수 있을 것으로 보인다. 공고정보 제공을 지표는 국가연구개발 사업의 과제 공고를 NTIS와 기관 홈페이지에 게시를 하였는가를 평가한다. 기관 홈페이지에는 과제 공고가 게시되어 있으나, NTIS에는 해당 과제 공고가 게시되어있지 않은 경우가 종종 발생한 것으로 파악되어 동 평가지표를 신설하였다. 과제 참여연구원 정보제공 지표는 일부 과제에서 참여연구원 정보가 누락·입력되어, 과제협약서 상 참여연구원의 수와 NTIS에 입력된 참여연구원의 수가 불일치하는 문제가 발생하여 신설되었다. 또한 NTIS에 입력된 과제 참여연구원 총인원과 개별 참여연구원의 수의 합이 불일치하는 문제도 제기되고 있다. 이에 동 지표는 앞서 거론된 문제점들을 해결하기 위해 신설되었다. 동 지표를 통해 매년 실시되는 조사분석 통계의 정확도와 신뢰도가 증진될 것으로 보인다.

다음으로 정성적 평가에 대해 알아본다. 정보연계 및 과제신청 창구 일원화 협조도 지표는 국정과제 및 NTIS 협조도 지표로 지표명이 변경되었으나, 지표 내용의 차이는 크지 않고, 배점도 동일하게 유지되었다. 부처별 품질향상 노력 지표는 전과 동일하다. 기관 특성화 자율지표는 삭제되었다. 동 지표는 구체적인 평가 지표가 제시되지 않은 채, 기관의 평가서를 접수하다보니, 평가상 객관성이 떨어지는 문제가 발생하여 삭제된 것으로 판단된다.

<표 22> 데이터 품질평가 지표 개선 전후

	개선 전	개선 후	비고
정량적 평가	<ul style="list-style-type: none"> <li>· 데이터 오류율(30점)</li> <li>- 과제 오류율(20점)</li> <li>- 인력 오류율(10점)</li> </ul>	<ul style="list-style-type: none"> <li>· 데이터 오류율(15점)</li> </ul>	<ul style="list-style-type: none"> <li>· 과제정보, 인력정보 통합 평가</li> </ul>
	-	<ul style="list-style-type: none"> <li>· (신규)데이터 서비스 달성율(5점)</li> </ul>	<ul style="list-style-type: none"> <li>· 최초 수집에서 최종 서비스까지 데이터 서비스 달성율 평가</li> </ul>
	<ul style="list-style-type: none"> <li>· 연계항목 목표달성도 (10점)</li> <li>- 연계목표 항목 수(5점)</li> <li>- 데이터 충실도(5점)</li> </ul>	<ul style="list-style-type: none"> <li>· 데이터 입력 항목수 및 데이터 충실도(10점)</li> <li>- 데이터 입력 항목수(5점)</li> <li>- 데이터 충실도(5점)</li> </ul>	<ul style="list-style-type: none"> <li>· 지표 변경</li> <li>· 기존 121개 연계목표 항목 수 → 155개 입력 항목 수로 변경</li> </ul>
	<ul style="list-style-type: none"> <li>· 과제정보 제공률(30점)</li> <li>- 협약자료제출(5점)</li> <li>- 정보 제공률(30점)</li> </ul>	<ul style="list-style-type: none"> <li>· 과제정보 제공률(20점)</li> <li>- 협약자료제출(5점)</li> <li>- 정보 제공률(15점)</li> </ul>	<ul style="list-style-type: none"> <li>· 전과 동일</li> </ul>
	-	<ul style="list-style-type: none"> <li>· (신규)참여제한 정보 제공(10점)</li> <li>- 참여제한 정보 적시 제공률(5점)</li> <li>- 데이터 충실도(5점)</li> </ul>	<ul style="list-style-type: none"> <li>· 참여제한 정보 적시제공률: (참여제한 시작일 이전 등록건수) / (참여제한 등록건수)</li> <li>· 데이터 충실도: (NTIS승인건수) / (참여제한등록건수)</li> </ul>
	-	<ul style="list-style-type: none"> <li>· (신규)공고정보 제공율 (5점)</li> </ul>	<ul style="list-style-type: none"> <li>· (NTIS공고등록건수)/ (기관홈페이지 내 공고등록건수)</li> </ul>
	-	<ul style="list-style-type: none"> <li>· (신규)과제 참여연구원 정보제공(15점)</li> <li>- 입력된 총인원 대비 개별 참여연구원 일치율(8점)</li> <li>- 입력된 총인원과 과제 협약서의 총인원 오차율(7점)</li> </ul>	-
정성적 평가	<ul style="list-style-type: none"> <li>· 정보연계 및 과제신청 창구 일원화 협조도(10점)</li> </ul>	<ul style="list-style-type: none"> <li>· 국정과제 및 NTIS 협조도(10점)</li> </ul>	<ul style="list-style-type: none"> <li>· 지표명 변경</li> </ul>
	<ul style="list-style-type: none"> <li>· 부처별 품질향상 노력 (10점)</li> </ul>	<ul style="list-style-type: none"> <li>· 부처별 품질향상 노력 (10점)</li> </ul>	<ul style="list-style-type: none"> <li>· 전과 동일</li> </ul>
	<ul style="list-style-type: none"> <li>· 기관 특성화 자율지표 (5점)</li> </ul>	-	<ul style="list-style-type: none"> <li>· 삭제</li> </ul>

---

# 제5장

## 결론

---

---

정보화 시대의 도래와 정보통신기술의 발달은 사회전반에 걸친 변화를 야기하고 있다. 정보의 유통은 우리 사회를 폐쇄형·수직적 체계에서, 모든 시민이 쉽게 접근 가능한 개방형·수평적 체계로 바꾸어 놓았다.

시대의 흐름에 맞게 기관 혹은 기업 내 업무의 모습도 변화하고 있다. 전자결제 시스템, 전자우편 시스템 등 전사에 데이터베이스기반 IT 솔루션이 보급되면서, 업무의 효율성이 증대되고, 업무처리 속도가 빨라지고 있다. IT가 기업 내에 필수 요소로 자리 잡으며, 데이터 품질관리의 역할이 부각되고 있다. 데이터의 품질 저하는 업무의 정확성과 효율성을 감소시키고, 데이터 정제 등의 추가적인 작업과 이에 수반한 비용이 발생한다. 데이터 품질 저하로 발생한 피해는 사기업인 경우, 사기업의 책임으로 마무리될 수 있지만, 대국민 서비스를 담당하는 공공기관의 경우, 일반 국민이 피해를 받는다는 점에서 공공분야의 데이터 품질 관리가 더욱 중요시 되고 있다.

R&D정보 서비스 및 정부R&D 투자 효율성 제고를 위해 구축한 국가과학기술지식정보서비스(NTIS)는 5가지 데이터품질관리 원칙을 수립하여 데이터품질 관리 활동을 수행하고 있다. 5가지 원칙은 ① 데이터 성능(국가R&D데이터의 서비스성능 극대화), ② 데이터 품질(데이터의 정확성과 신뢰도 향상), ③ 데이터 통합(데이터 모델과 데이터의 통합관리), ④ 데이터 활용(사용자가 원하는 데이터를 활용할 수 있도록 관리), ⑤ 데이터 보안(데이터의 활용을 최대로 보장하되 보안성을 강화)이다.

NTIS의 데이터 품질관리 체계 프레임워크는 ① 요구사항 관리, ② 데이터 구조 관리, ③ 데이터베이스 관리, ④ 데이터 흐름 관리, ⑤ 데이터 표준 관리, ⑥ 데이터 활용 관리, ⑦ 데이터 오너쉽 관리, ⑧ 사용자 뷰 관리 등으로 구성되어 있다.

NTIS의 데이터 품질평가는 데이터품질관리 체계 8가지 프레임워크 중 6번째 데이터 활용관리 중 1개 요소로 포함되어 있다. 데이터 품질평가는 신뢰성 높은 고품질의 정보를 대국민에게 서비스하기 위해 정보가 최초로 생산되는 각 대표 전문기관의 데이터 품질 수준을 제고하고자 함에 그 목적이 있다.

이러한 NTIS의 데이터 품질관리 노력의 결과로, 2014년 현재 NTIS의 데이터 품질관리 성숙도는 활용성 측면에서 2레벨, 유효성 측면에서 3레벨을 유지하고

---

있다. 또한, 2014년 기준, 전제 수집 데이터 오류율은 1.10%, 전시 과제정보 제공율은 90%로 NTIS의 높은 관리 수준을 대변하고 있다. NTIS의 이용자를 대상으로 한 서비스 만족도 조사에서는 2013년 80.1점으로 2009년 이래 최초로 80점을 상회하였으며, 2014년에는 81.2점을 기록하였다. 한편, 2014년 NTIS 정보이용자를 대상으로 고객가치를 조사한 결과, 연구자는 NTIS를 보고서 작성 시간 절감, 연구개발 소요시간 절감, 연구의 질적 수준 제고 등의 측면에서 가장 높은 점수를 부여하였다. 연구관리자는 유사과제 탐색시간 절감, 논문, 특허출원 양적 향상 등의 측면에서 가장 높은 점수를 부여하였다.

NTIS는 2006년 ~ 2009년 동안 범부처 협력체계 마련을 위해 이용자를 과제 담당자로 한정하고 기반조성을 위한 정보수집 측면에서 시스템을 구축하였다. 2010년 ~ 2012년 동안 정보연계의 확대를 위해 이용자를 정책 및 기획 입안자로 확대하고 대상정보를 확산·적용하는 방향으로 계획을 수립하고 추진하였다. 2013년 ~ 2015년 동안 정보공동활용 극대화를 목표로 최종 연구자 중심의 시스템을 구축하고 대상정보를 지능형으로 고도화하였다. 2016년 ~ 2020년 동안 국가 R&D CIO를 목표로 고객중심의 개방과 융합, 활용의 시스템을 구축할 계획을 수립하였다. NTIS를 통해 R&D 정보를 상호 개방 및 공유하고, 과학기술의 고부가가치를 창출하는데 기여하여야 할 것이다.

정보의 상호 개방과 공유는 정보의 신뢰성이 담보된다. 지속적인 국가R&D 데이터 품질관리와 평가로 정보의 신뢰성을 높은 수준으로 유지하여야 한다. 데이터 값 이외에 실시간 정보연계와 같은 NTIS의 서비스 운영에 필요한 요소들로 평가범위를 확대하여 넓은 범위의 데이터 품질관리를 수행해야 할 것이다. 앞으로 국가 과학기술의 발전을 위해 NTIS가 선도적인 역할을 하고, 데이터 품질관리 측면에서 롤모델이 될 수 있도록 더욱 발전하기 바란다.

[부록]

## 중 · 장기적 접근이 요구되는 국가 R&D 데이터 품질평가 지표

구분		현황 및 문제점	검토내용
종합	체급별 (또는 그룹별) 평가	<ul style="list-style-type: none"> <li>○ 17개 대표전문기관을 동일하게 평가하여 우수 기관은 계속 수상받고, 하위 기관은 계속 수상받지 못하는 불균형 현상 발생</li> <li>※ 명예 졸업제(상위 3위 3회 수상 기관은 이후 2년간 평가 순위 발표 대상에서 제외)를 시행함에도 불구하고 하위 기관은 제외</li> <li>※ 품질 개선도가 높은 약진하는 기관에 대해 특별상 시상(15년~)</li> </ul>	<ul style="list-style-type: none"> <li>○ 부처·청과 대표전문기관을 분리* 하거나, 데이터양이나 인원을 고려하여 체급별로 분리**하는 방식 검토</li> <li>* 부처·청 중 일부 부처·청의 경쟁력은 상위 수준</li> <li>** 기관별 그룹화 및 기관 보정 계수 등을 검토(기관별 그룹화 기준설정 시 불공정 문제가 발생가능하며, 객관적이고 공정한 체급 구분의 한계)</li> </ul>
	인증기관의 NTIS 데이터품질 평가	<ul style="list-style-type: none"> <li>○ 외부의 공신력 있는 기관에 의뢰하여 현장 품질평가</li> </ul>	<ul style="list-style-type: none"> <li>○ 고가의 심사비용 발생 및 대표 전문기관에서 현장 품질평가에 대한 부담</li> </ul>
	정량적 평가 횟수 확대	<ul style="list-style-type: none"> <li>○ 정량적 평가를 연 2회(7월, 10월) 실시</li> </ul>	<ul style="list-style-type: none"> <li>○ 연 4회(4월, 6월, 8월, 10월)이상으로 평가 횟수 확대. 평가기관과 피평가기관에 평가 부담.</li> </ul>
	미연계 기관 평가 및 시상	<ul style="list-style-type: none"> <li>○ 데이터 품질평가 대상기관에서 제외</li> </ul>	<ul style="list-style-type: none"> <li>○ 미연계 기관을 포함하여 평가하는 방안을 고려하였으나, 데이터양이 소량임</li> </ul>
	외부 평가위원 선정	<ul style="list-style-type: none"> <li>○ 데이터 품질평가에서 미래부와 NTIS센터 심사위원이 참여하나 외부 심사위원은 미참여</li> </ul>	<ul style="list-style-type: none"> <li>○ 향후 외부 평가위원 참여 - 한국데이터베이스진흥원, 한국정보화진흥원 등 검토</li> </ul>



구분		현황 및 문제점	검토내용										
정량적 평가	데이터 품질평가 범위 확대	○ 기존 과제, 인력 정보에 대상으로 평가하던 분야를 성과, 연구시설·장비까지 점진적 확대	○ 성과 데이터는 주로 연말에 수집되어 제외되었으나 향후 중장기적으로 성과 정보뿐만 아니라 연구시설·장비까지 확대 ※ 개선 예 <table border="1" style="margin-left: 20px;"> <tr> <td>'16</td> <td>'17</td> <td>'18</td> <td>'19</td> <td>'20</td> </tr> <tr> <td>과제 인력 논문</td> <td>과제 인력 논문 특허</td> <td>과제 인력 논문 특허 사업화 기술료</td> <td>과제 인력 논문 특허 사업화 기술료 기 탄 과</td> <td>과제 인력 논문 특허 사업화 기술료 기 탄 과 신 장 비</td> </tr> </table>	'16	'17	'18	'19	'20	과제 인력 논문	과제 인력 논문 특허	과제 인력 논문 특허 사업화 기술료	과제 인력 논문 특허 사업화 기술료 기 탄 과	과제 인력 논문 특허 사업화 기술료 기 탄 과 신 장 비
	'16	'17	'18	'19	'20								
과제 인력 논문	과제 인력 논문 특허	과제 인력 논문 특허 사업화 기술료	과제 인력 논문 특허 사업화 기술료 기 탄 과	과제 인력 논문 특허 사업화 기술료 기 탄 과 신 장 비									
오류 개선율이 높은 기관 대상으로 정량적 평가	○ 품질 향상 및 하위 기관에 대한 고려 필요	○ 하위 기관에게만 유리한 평가 ○ 정량적 평가를 17개 대표전문 기관에 전체 적용 시, 객관성 및 공정성 문제 발생 가능											
정성적 평가 점수	정보연계 모니터링 서비스 (ReMon) 접속 및 피드백 정도	○ 기관별 ReMon 담당자의 월 평균 접속 횟수와 ReMon을 통해 오류 조회 후 피드백(수정)을 평가	○ ReMon 접속 활성화가 기대되나, 현재 로그(Log) 기록 미보유										
	기관특성화 자율지표	○ '14년 기관특성화자율지표 (5점)을 시행하였으나 객관성 및 공정성 떨어짐	○ 대표전문기관에서 기관특성화 자율지표에 따른 자료 작성의 부담과, 평가시 객관성 및 공정성이 낮아 무의미한 지표임										

---

## 참 고 문 헌

- 김문영 (2005), “체계적인 데이터 품질 관리를 위한 대안을 찾아서”, 디지털 콘텐츠.
- 김재수 (2009), “국가과학기술종합정보서비스(NTIS)”, 「한국블리비아」, 20: 187-206.
- 김진만 · 신성호 · 운영준 · 양명석 · 손강렬 (2010), “서비스 중요도의 가중치를 적용한 R&D과제 데이터에 대한 품질평가 방법”, 「한국인터넷정보학회 추계학술발표대회」, 289-290.
- 미래창조과학부 (2016가), “국가연구개발사업의 관리 등에 관한 규정”.
- 미래창조과학부 (2016나), “국가연구개발정보표준”.
- 서상혁 · 이선영 · 이병희 (2012), “국가 R&D정보 이용자의 고객가치 및 고객 만족도 영향요인 분석”, 「기술혁신학회지」, 15(4): 837-861.
- 서상혁 · 이선영 · 이병희 (2014), “연구관리자의 국가R&D정보서비스 고객가치 및 업무성과 영향요인”, 「한국콘텐츠학회논문지」, 14(7): 480-494.
- 손강렬 · 임종태 (2009), “품질문제 요인분석을 통한 데이터 품질 개선방안 연구: 국가R&D정보연계 방식 중심”, 「한국콘텐츠학회논문지」, 9(1): 1-14.
- 신성호 · 운영준 · 양명석 · 김진만 · 손강렬 (2011), “데이터 품질을 고려한 국가 R&D정보 데이터베이스의 통합 사례 연구 - NTIS 데이터베이스 통합 사례”, 「한국컴퓨터정보학회논문지」, 16(6): 119-130.

---

양명석 · 최광남 · 정옥남 · 김재수 (2013), “국가과학기술지식정보서비스(NTIS)에 관한 고찰”, 「한국기술혁신학회 학술대회」, 294-304.

이민호 · 엄석진 · 김명진 (2014), “정보 공유 및 시스템 연계 · 통합 활성화 방안”, 한국행정연구원.

이병희 · 임철수 · 신동구 · 정옥남 · 김재수 (2009), “NTIS의 ISO 27001 체계 수립과 ISO 20000 인증 사후관리”, 「한국기술혁신학회 추계학술대회」, 373-380.

정영기 (2015), “공공개방데이터 품질진단모델에 관한 연구”, 숭실대학교 석사 학위논문.

정혜정 (2007), “데이터 품질 평가에 관한 연구”, 「한국인터넷정보학회논문지」, 8(4): 119-128.

최승준 · 박제원 · 김종배 · 최재현 (2014), “빅데이터 분산처리시스템의 품질평가 모델”, 「디지털콘텐츠학회지」, 15(4): 533-545.

최희석 · 정옥남 · 김윤정 · 김태현 · 임철수 · 이병희 · 김재수 (2009), “효율적인 NTIS 서비스 운영을 위한 표준운영절차 : NTIS SOP”, 「한국정보처리학회 춘계학술발표대회」.

최희석 · 김재수 (2012), “표준연계플랫폼 기반의 NTIS 정보연계모델 설계”, 「정보과학회논문지: 컴퓨팅의 실제 및 레터」, 18(6): 484-488.

특허청 (2012), “특허정보DB 관리체계 선진화 4개년 종합계획(안)”.

한국데이터베이스진흥센터 (2006), “데이터 품질관리 지침(Ver 2.1)”.

---

한국데이터베이스진흥센터 (2009), “데이터 품질진단 절차 및 기법(Ver 1.0)”.

한국데이터베이스진흥원 (2014), “미래부 R&D 등 공공데이터 품질관리 체계 마련 연구 보고서”.

한국정보화진흥원 (2012), “관세청 데이터 품질관리 사례”.

한국정보화진흥원 (2013), “정부3.0 및 창조경제를 위한 데이터 품질관리 사례집”.

행정안전부 (2011), “공공기관의 데이터베이스 품질관리 지침”.

호진원 · 이미영 (2013), “IT활용 감사의 효과성 향상을 위한 데이터 품질관리 방안 연구”, 「한국사회와 행정연구」, 23(4): 31-53.

Larry P. English (1999), “Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits”.

국가과학기술지식정보서비스 (NTIS) <http://www.ntis.go.kr>

데이터베이스품질인증 (DQC) <http://www.dqc.or.kr/>

국가R&D 데이터 품질평가와 개선방안  
Evaluation of National R&D Data Quality and its Direction for Improvement

---

2016년 10월 일 인쇄  
2016년 10월 일 발행  
I S B N 978-89-294-xxx-x-xxxxx  
저 자 이병희, 임철수, 최기석  
발 행 처 한국과학기술정보연구원  
대전광역시 유성구 과학로 335  
전화 : 042-869-1234  
www.kisti.re.kr  
인 쇄 미래미디어 042-487-4012

---