KOLIS-NET 종합목록 DB의 품질평가*

Quality Evaluation of a Shared Cataloging DB: the Case of KOLIS-NET

김 선 애(Sun-Ae Kim)**
이 수 상(Soo-Sang Lee)***

목 차

1. 서 론
2. KOLIS-NET 종합목록 DB의 품질평가 방법 및 절차
 2.1 품질평가를 위한 방법
 2.2 품질평가를 위한 표본레코드
3. KOLIS-NET 종합목록 DB의 품질평가 결과
 3.1 포괄성 평가
 3.2 중복성 평가
 3.3 최신성 평가
 3.4 정확성 평가
 3.5 일관성 평가
 3.6 완전성 평가
4. 결 론

초 록

이 연구는 국립중앙도서관을 중심으로 전국의 공공도서관들이 소장하고 있는 자료를 구축해 놓은 국가자료종합목록시스템(KOLIS-NET) DB의 품질을 평가하는데 그 목적이 있다. 국가자료 종합목록 DB의 품질평가는 선행연구에서 제시된 품질평가 모델을 참조하여 포괄성, 중복성, 최신성, 정확성, 일관성 그리고 완전성의 6개의 항목에 대해 사례에 의한 발견적 평가를 시도하였다. 실행모델에 근거한 평가를 통해 국가자료 종합목록 DB의 품질을 종합적으로 검증하고, 품질을 열악하게 만드는 요인을 진단하였으며 KOLIS-NET종합목록 DB의 품질수준 제고를 위한 개선방안을 평가에서 나타난 구체적 문제점을 중심으로 제시하였다.

ABSTRACT

The research purpose for this study is to evaluate the quality of the KOLIS-NET DB which builded bibliographic data of the holding collections of nationwide public libraries. The quality evaluation of the KOLIS-NET DB was inspired by the successful experience from researches precedent and tried to approach the case study focuses on six quality dimensions: coverage, duplication, currentness, accuracy, consistency completeness. The study verified comprehensively the quality of the KOLIS-NET DB through a quality evaluation model and analyzed the factors causing such inferior and substandard bibliographic records in the KOLIS-NET DB. Based on the results of the quality evaluation, the quality improvements of the KOLIS-NET DB was suggested.

키워드: 국가자료공동목록시스템, 품질평가, 품질관리
KOLIS-NET, Quality Evaluation, Quality Management

---

* 이 논문은 국립중앙도서관의 "국가자료종합목록 발전 및 운영모델 개발에 관한 연구"의 일환으로 연구되었음
** 경성대학교 문헌정보학과 전임강사(kimsa@ks.ac.kr) (제1저자)
*** 부산대학교 문헌정보학과 조교수(sslee@pusan.ac.kr) (공동저자)
논문접수일자 2006년 2월 8일
게재확정일자 2006년 3월 17일

1. 서 론

국가자료공동목록시스템(이하 KOLIS-NET이라 함)은 국립중앙도서관을 중심으로 전국의 공공도관들이 소장하고 있는 자료에 대한 목록을 통합 구축할 수 있도록 설계된 데이터베이스이다. 국립중앙도서관은 이를 통해 도서관간 정보공유 및 상호협력을 가능케 하는 종합정보 유통망을 구축하고자 하였다. 다시 말해, KOLIS-NET은 통합된 전국 공공도서관의 소장자료 정보를 다시 전국의 공공도서관들이 공유함으로써 정보자원의 효율적 이용을 기대하였다. 또한 타 도서관에서 이미 작성한 목록, 목차, 초록 등 자료에 대한 다양한 정보를 공동 활용할 수 있게 함으로써 도서관의 업무효율을 향상시키는데 기여하고자 하였다. 이처럼 KOLIS-NET은 국가자료에 대한 종합적인 정보유통망의 구축을 통해 정보서비스의 질적 향상을 기하고, 동시에 전국공공도서관의 자료관리 및 목록정보 구축의 표준화 체제를 확립할 목적으로 지난 2000년에 개발되기 시작하였다.

이러한 취지로 출발한 KOLIS-NET은 2001년 3월부터 2001년 4월까지 공공도서관 지역센터 9개관 업무관계자를 중심으로 전담반을 운영하고, 2001년 6월 설명회를 거친 후 2001년 9월부터 전국의 공공도서관에 보급되기 시작하였다. 이듬해인 2002년 3월에는 30개관을 대상으로 시범운영을 실시하였으며, 이를 바탕으로 2002년 4월 22일부터 정상운영에 들어갔다. 2005년 현재 전국의 공공도서관에 설치된 KOLIS-NET은 통합형[1]이 370개관, 독립형이 9개관, 그리고 약 310개의 공공도서관들이 KOLIS-NET을 업무에 활용하고 있는 것으로 나타난다.

프로그램의 보급과 더불어 종합목록 DB의 구축작업도 시작되었다. DB 구축초기인 2001년도에는 국립중앙도서관과 전국의 공공도서관 및 행정부처 자료실로부터 소장자료에 대한 목록데이터를 일괄 제출받아 150만 건의 종합목록 DB를 구축하였다. 이어 2002년에는 200만 건, 2003년에는 100만 건, 2004년에는 162만 건의 서지레드를 구축하였다. DB 구축과 더불어 구축된 DB에 대한 품질관리도 이루어졌다. 2003년에는 KOLIS-NET을 업로드하는 과정에서 발생한 오류데이터와 중복데이터 100만 건을 처리하기 위해 구축된 DB 전량을 점검하는 '국가자료종합목록 DB 점검팀'과 DB 구축 및 점검 상태를 최종적으로 확인하는 'DB 품질관리팀'을 운영하였으며, 이와 별도로 단위도서관에서 오류레코드의 수정요청이나 중복레코드의 통합요청이 있을 때마다 목록의 품질을 수시로 관리할 수 있는 체제를 두었다. 이렇게 하여 KOLIS-NET은 3년 남짓한 짧은 기간 동안에 443만 건의 서지레코드와 1,735만 건의 소장레코드(2005년 10월 7일 기준)을 가진 국내의 대표적인 종합목록으로 발전하였다.

그러나 이러한 양적 성장에도 불구하고 KOLIS-NET은 현재 그 유용성에 상당한 의문이 제기되고 있다. 특히 서지공유형 방식을 통해 국내 공공도서관의 자료관리 및 목록정보의 표준체계를 확립하여 개별도서관들이 편목업무에 투자하는 시간과 비용을 절감해 주고자 했던 목적에 도달하지 못하고 있다. 본 연구팀의 설문

1) KOLIS-NET은 KOLAS 및 KOLASII를 사용하고 있는 도서관을 위한 'KOLIS-NET 클라이언트 통합형'과 기타 상용프로그램을 사용하고 있는 도서관을 위한 'KOLIS-NET 클라이언트 독립형'으로 구분된다.

조사[2]에 의하면, 공공도서관의 편목담당사서들 사이에서 KOLIS-NET에 대한 평가는 매우 부정적이었다. 그들은 KOLIS-NET을 개별 도서관들이 편목업무에 투자하는 시간과 비용을 최소화할 수 있는 최적의 서지 도구로 인식하지 않고 있었으며, 따라서 편목업무에 있어서 KOLIS-NET의 활용도는 매우 낮았다. KOLIS-NET을 이용하지 않은 주된 이유로는 DB 품질의 열악함과 최신 데이터의 부재를 들었다.

이러한 상황에서 KOLIS-NET 종합목록에 대한 종합적인 품질평가를 실시하여 현단계의 품질을 실증적으로 검증하고 특히, KOLIS-NET 종합목록 DB의 품질을 열악하게 만드는 주요 원인을 밝혀보고자 하는 의도에서 이 연구는 시도되었다. KOLIS-NET 종합목록이 지금까지 양적 성장을 추구했다면 앞으로 나아갈 방향은 질적 성장이다. 질적 성장 없이는 도서관간 정보공유 및 상호협력을 위한 국가자료 종합 정보유통망의 중심기관으로서 역할을 수행할 수 없기 때문이다. 따라서 이 연구는 국립중앙도서관이 KOLIS-NET에 구축해 놓은 종합목록 DB의 품질을 체계적인 검증을 통한 평가를 실시하여 종합목록 DB의 오류레코드 유형, 이용률 저하원인 및 중복레코드 유형 등을 파악하는데 그 목적이 있다. 이를 위해 이 연구에서는 종합목록 DB의 품질을 평가하기 위한 실행모델을 개발하였으며, 개발된 모델에 근거하여 종합목록 DB의 품질을 사례를 통한 실증적 발견적 평가를 실시하였다.

2. KOLIS-NET 종합목록 DB의 품질평가 방법 및 절차

2.1 품질평가를 위한 방법

DB의 품질평가와 관련한 국내의 선행 연구를 분석해 보면 품질평가의 목적과 평가대상 그리고 평가주체에 따라 다양한 평가기준과 절차 그리고 방법이 개발되어 적용되어 왔음을 알 수 있다. 그리고 DB 품질에 대한 인식의 차이 또는 DB 품질에 대한 정의의 차이 또한 품질기준의 차이를 형성하고 있음을 알 수 있다. 선행연구에서 조사된 바에 의하면 DB 품질은 DB 데이터의 품질과 DB 서비스의 품질로 정의되며, 전자가 데이터의 정확성, 완전성, 최신성, 일관성, 포괄성 등을 평가항목으로 제시하는 반면, 후자는 검색성, 사용용이성, 사용자지원성, 비용 등을 평가항목으로 제시하고 있다.

---

2) 『국가자료종합목록 발전 및 운영모델 개발에 관한 연구』를 수행하는 과정에서 연구팀은 KOLIS-NET의 활용도와 참여도에 대한 공공도서관 실무자들의 솔직한 의견을 수렴하고자 방문조사, 면담조사 및 설문조사를 실시하였다. 설문조사는 『한국도서관연감 2004』를 토대로 461개의 공공도서관에 근무하는 편목담당사서들 중 389개 도서관을 대상으로 이루어졌다. 이메일을 통한 설문조사는 2005년 8-9월에 실시되었으며 회수률은 55.5%(216개 도서관)였다. 설문조사의 결과는 필요한 경우 전화인터뷰를 통해 그 내용을 보충하였다. 설문내용은 참여도서관의 편목환경(편목방법, 담당인원, 데이터양 등), KOLIS-NET에 대한 참여도서관의 인식 및 참여도서관의 업무 활용도로 구성되었다. 결과를 간략히 살펴보면 편목담당사서들은 KOLIS-NET의 주된 기능으로 48.6%가 편목업무의 효율성 향상을 꼽았으며, KOLIS-NET에서 가장 중요하게 고려되어야 할 요소로는 35.2%가 자료의 최신성이라고 응답하였다. 이러한 KOLIS-NET에 대한 인식과 달리 KOLIS-NET의 실질적인 활용부문에서는 참여도서관의 대부분(50.5%)이 분류기호 등 특정데이터를 단순히 참조하기 위해서 KOLIS-NET을 이용한다고 대답하였다.

국내의 품질평가에 관한 주요 연구 중 안계성과 함정훈의 한국데이터베이스진흥센터(2000), 이응봉 등(2001)의 19개 평가기준, 11개 선행연구를 분석한 이제환 등(2002), 4개 선행연구를 분석한 이만수 등(2003), 17개의 선행연구를 분석한 KERIS(2005)의 연구들에 나타난 데이터베이스의 품질평가 기준을 정리하면 〈표 1〉과 같다. 〈표 1〉에서 알 수 있듯이 국내 선행연구들은 데이터베이스의 품질평가를 위해 한국데이터베이스진흥센터에서 제시하는 동일한 평가항목을 사용하고 있다. 평가항목 수에서 한 두 개의 차이가 있지만, DB의 품질을 데이터의 품질, 서비스의 품질 그리고 기타로 구분한 것은 4개의 연구에서 공통적으로 나타난다. 그렇다면 선행연구들간에 나타나는 차이는 평가항목 수보다 그들 항목들이 측정하고자 하는 내용의 차이이다. 즉 평가의 대상이 되는 DB의 특성에 따라 측정하고자 하는 내용에 있어 차이점을 나타내고 있다.

KOLIS-NET 종합목록 DB의 품질평가는 DB 품질 중 데이터의 품질만을 평가한다. 측정항목의 선정에 있어서 연구팀은 선행연구들에서 공통적으로 사용되었던 5가지 항목 -최신성, 정확성, 완전성, 포괄성, 일관성- 을 그대로 사용하고자 한다. 그리고 KOLIS-NET이 전국의 공공도서관들의 참여로 구축되는 종합목록임을 감안하여 전체 레코드 중 동일한 레코드의 중복정도를 측정하는 중복성을 추가하였다. 이들 평가항목은 공공도서관의 편목담당사서들을 대상으로 한 종합목록에서 중요하게 고려되어야 할 요소가 무엇인가에서 나타난 결과와 유사하다. 이들은 종합목록의 유용성에서 가장 중요한 요소로 최신자료의 신속한 반영을 그 다음으로는 데이터 품질의 우수성 그리고 표준화된 형식과 기술규칙의 준수 등을 들었다. 이는 종합목록 DB를 유용성의 관점에서 봤을 때 현장의 편목담당사서들은 데이터의 최신성, 정확성, 완전성 그리고 일관성을 주요한 요소로 생각하고 있는 것으로 나타났다. 선행연구들에서 사용한 평가항목과 현장의 편목담당사서들의 의견을 종합하여 그리고 종합목록이 갖는 특성을 고려하여 KOLIS-NET 종합목록 DB의 품질 평가 기준을 다음과 같이 한다.

- 포괄성: KOLIS-NET 종합목록이 단위 공공도서의 소장 DB을 모두 포괄하고 있는가?
- 중복성: 하나의 동일한 레코드에 대해 둘 이상의 상이한 레코드가 존재하는가?
- 최신성: 최신 발간자료의 서지레코드가 사용가능한가?
- 정확성: DB를 구성하는 레코드 내용이 실제 값과 동일한가?
- 일관성: 제공하는 정보의 내용 표기에 있어 레코드의 구조가 일정한가?
- 완전성: 제공하는 정보의 내용표기에 있어 레코드 필드 또는 데이터 값이 누락 없이 완전한가?

선정한 6개의 평가기준을 측정하기 위한 평가지표의 개발은 이제환 등(2002)이 KERIS 종합목록 DB의 평가를 위해 사용한 지표를 참조하였다. 왜냐하면 그들의 연구에서 사용한 측정방법이 객관적인 비교평가가 수월하고 측정결과를 계량화 할 수 있는 방법을 제시하고 있기 때문이다. 항목별 평가지표와 측정방법은 〈표 2〉를 참조한다.

〈표 1〉 국내 주요 선행연구의 데이터베이스 품질평가 기준 정리표

구분	DB진흥센터의 품질기준	주요선행연구[3]의 기준별 측정기준			
		이응봉의 연구	이제환 외(2002)와 김성희(2003)의 연구	이만수 외 연구(2003)	KERIS(2005)
데이터 품질	최신성 (Currentness)	- 데이터 갱신 빈도	- 현행성 - 신속성 - 갱신주기 - 타임래그	- 최근 데이터의 갱신 - 신속한 갱신	- 갱신주기 - 출판 전 논문서비스 - 최근 원문제공 저널 수
	정확성 (Accuracy)	- 레코드 내 오류, 데이터의 양 - 레코드 작성기준 유무 (완전성에도 해당됨)	- 데이터의 신뢰성 - 데이터 표현 정확성 - 데이터 기술 정확성(철자 오류, 누락 등)	- 수록데이터의 정확성과 일치성 - 데이터 표현의 정확성 - 철자 오류와 누락	- 데이터의 오류 - 오류데이터의 양
	완전성 (Completeness)	- 레코드 필드 구조의 완전성과 일관성 - 공표된 레코드와 실제 레코드 수의 차이	- 레코드 구조 적합성 - 데이터 기술 완전성 - 레코드의 고유성 비중복성	- 레코드 구조의 적합성 - 데이터 기술의 완전성자료 속성의 완전 표현 고유레코드(비중복성)	- 공표된 레코드 수와 실제 레코드 현황 - 레코드의 고유성, 중복율
	포괄성 (Coverage)/수록 범위(Scope)	기타(수록범위)	- 수록 자료의 범위 - 수록 레코드의 수 - 유형과 시기 망라성 - 질의에 대한 응답률	- 자료범위/종류 다양 - 수록레코드의 수 - 데이터의 유형과 시기 - 가치 있는 자료 수록량	- 전체 레코드 수 - 논문/초록/원문제공별 구성 비율
	일관성 (Consistency)	* 일관성 대신 전문성항목만 별도로 품질 기준에 표시	- 레코드 구조 일관성 - 데이터 표현 일관성 - 데이터 기술 통일성	- 레코드 구조의 일관성 - 데이터 표현과 기술의 통일성	- 저널 명 표기방법의 통일성 - 레코드 구조의 일관성
서비스 품질	편의성 (Easy of Use)	- 메뉴검색 체계성/다양성 - 키워드 검색 용이성/다양성 - 디자인 이용자 지향성 - 검색결과 표시의 내용 주제별 다양성	- 인터페이스의 편의성 - 작업처리의 편의성 - 접속의 편의성	- 인터페이스의 편리성 - 접속의 편리성 - 결과의 손쉬운 활용	- 인터페이스의 용이성 - 검색결과 활용의 편리성 - Alert 서비스
	지원성 (Customer support)	- 오류데이터 수정의 체계성 - DB이용 안내 제공 - Help 메뉴 제공 - Q & A 제공	- 이용자 지원 및 교육 - 도큐멘테이션 유무 - Help 기능의 적절성 - 원문제공의 편의성	- 사용자 지원 및 교육 도큐멘테이션, 도움말 제공 - 원문 접근의 용이성	- 이용안내 - 도움말 메뉴 제공
	검색성 (Searching)	- 시소러스 관리 우수성 - 다양한 탐색기법제공 - 복수 DB검색 가능성 - 검색 시 1건의 평균 소요시간 - 검색결과 출력 다양성	- 검색속도 - 검색방법의 다양성 - 출력 기능의 다양성 - 시소러스의 유무	- 검색속도 - 검색방법의 다양성 - 출력기능의 다양성 - 시소러스 유무	- 검색방법의 다양성 - 원문링크의 다양성
	비용 (Cost)	* 품질기준으로 표시되었으나 평가 항목서 제외	- 구입비용 - 이용비용		- 구입비용과 상대적 타당성
	네트워크/하드웨어	* 품질기준으로 표시되었으나 평가 항목서 제외	- 접근성 - 접근 속도 - 접속방식		- 인터넷 외 추가 소요사항
기타	일반적인 요구사항		- 권위성 - 통합성 - 유일성 - 객관성 - 이용자 만족도 - 데이터 작성 지침준수		

3) 김성희 2003. 데이터베이스 평가모델 개발을 위한 해외사례조사연구. 『국회도서관보』, 40(2): 15-33.
안계성, 함정훈 2000. 『데이터베이스 품질평가 항목』, 서울: 한국데이터베이스진흥센터.
이만수, 이상열, 신인수 2003. 서지 데이터베이스의 품질평가에 관한 연구. 『정보관리연구』, 34(3): 21-39.
이응봉, 조현양, 류범종, 최재황 2001. 과학기술분야 데이터베이스의 품질향상을 위한 품질 평가 연구. 『한국문헌정보학회지』, 35(2): 109-132.
이제환, 노지현 2002. 『KERIS 서지 DB의 품질관리를 위한 평가모델 개발 및 개선방안 수립』, 서울: 한국교육학술정보원.
한국교육학술정보원. 2005 『경영·경제 분야 해외전자정보 비교 분석』, 서울: 한국교육학술정보원.

〈표 2〉 평가지표의 측정방법

평가항목	평가지표	측정방법
포괄성	▪포괄 레코드 수	이용 가능한 통계를 활용하여 절대값(수, %)을 측정
중복성	▪중복레코드의 비율	표집분석 통한 지표의 절대값을 측정 100개의 레코드 당 발생비율(%)을 산출
최신성	▪최근 발간자료의 서지레코드 구축 건수	서지대조법, 교차탐색법을 이용하여 지표의 절대값(%)을 측정
정확성	▪데이터 표기에 있어서 오류가 발생한 레코드의 비율	표집분석을 통한 지표의 절대값을 측정 100개의 레코드 당 발생비율(%)을 산출
일관성	▪구조가 상이한 레코드의 비율	표집분석 통한 지표의 절대값을 측정 100개의 레코드 당 발생비율(%)을 산출
완전성	▪데이터 필드가 누락된 레코드의 비율 ▪데이터 값이 누락된 레코드의 비율	표집분석 통한 지표의 절대값을 측정 100개의 레코드 당 발생비율(%)을 산출

KOLIS-NET 종합목록 DB의 품질평가를 위해 선정된 6개의 평가항목을 통해 연구팀은 KOLIS-NET 종합목록 DB의 오류레코드 유형, 이용률 저하원인 및 중복레코드 유형 등에 대한 원인 파악을 시도한다.

2.2 품질평가를 위한 표본레코드

이 연구는 KOLIS-NET 종합목록 DB의 실질적 유용성을 측정하기 위하여 두 종류의 표본레코드를 활용하였다. 선정된 6개의 품질평가 항목 중 포괄성, 중복성, 정확성, 일관성, 완전성 측정을 위해서는 연구팀이 개발한 탐색질의를 이용하여 KOLIS-NET 종합목록 DB를 검색하였으며, 검색된 결과를 종합하여 리스트를 만들어 사용하였다. DB의 우수성과 표준화된 형식 및 규칙과 관련된 DB의 품질평가를 위한 표본레코드의 추출을 위해서 다음의 탐색질의를 사용하였다: 자료의 유형을 단행본으로 선택한 한 후 자료의 형태를 일반자료로 제한한 후, 표제키워드 '정보사회' 사용. 이러한 검색조건에 의해 검색된 출력레코드는 386건 이었으며 이 중 부적합레코드 60[4]건을 제외한 326건을 종합목록 DB의 품질측정을 위한 표본레코드 집단으로 활용하였다. 그리고 최신성 평가를 위해서는 단위도서관의 1회분 수서자료리스트를 활용하였다. 종합목록 DB에 최신자료가 얼마나 신속히 유입되는지, 다시 말해 단위 공공도서관에서 최신자료의 서지레코드 구축 시 사용가능한 최신자료의 레코드는 어느 정도인지를 가늠할 수 있는 최적의 표본레코드가 개별도서관의 수서자료리스트라고 판단했기 때문이다. 이를 위해 대도시에 위치한 중규모 공공도서관 한 곳[5]의 8월분 수서자료리스트 921건을 확보하여 활용하였다.

4) 386건의 검색레코드 중에서 자료유형 필드에 단행본이라고 기술되어 있으나, 실제로는 학위논문인 레코드 3건, 학술지수록논문(연속간행물) 48건, 그리고 단행본의 일반자료가 아님에도 불구하고 검색된 레코드가 9건 포함되어 있었다. 이처럼 자료유형을 단행본 중에서 일반자료를 대상으로 한 검색결과 386건 중 총 60건이 탐색조건(자료유형)에 적합하지 않은 자료로 확인되었다.

5) 부산광역시에 소재한 시립공공도서관으로 연간 2억3천여원의 도서구입비를 사용하고 있다.

3. KOLIS-NET 종합목록 DB의 품질평가 결과

3.1 포괄성 평가

포괄성을 측정하는 목적은 KOLIS-NET 종합목록 DB가 단위도서관의 학술자료를 모두 포함하고 있는지를 평가하는데 있다. 종합목록 DB는 당연히 단위 공공도서관의 학술자료를 모두 포함하여야 하며, 이는 이용자가 검색한 결과에 대한 신뢰도를 측정하는 기준이 될 것이라는 가정에 기초한다. 또한 이러한 포괄성 측정은 종합목록 DB의 운영, 관리 소홀 및 참여자들의 책임의식부족도 엿볼 수 있는 항목이다.

앞서 언급한 실험용 탐색질의, 단행본의 일반자료를 대상으로 한 표제키워드 '정보사회'를 이용한 검색결과를 표본레코드로 사용하였다. KOLIS-NET 종합목록 DB가 단위 공공도서관의 자료를 모두 포괄하고 있는지를 나타내는 포괄성 측정을 위해서는 비교대상 DB의 선정이 우선 되어야 한다. 연구팀은 비교 조사대상 DB를 3개 대도시의 공공도서관중 소장자료의 규모가 가장 큰 기관을 대상으로 하였다: 서울 정독도서관, 부산 시립시민도서관, 대구 시립중앙도서관. 비교대상으로 선정된 3개 공공도서관의 소장 DB를 대상으로 KOLIS-NET 종합목록에서 사용했던 동일한 검색식을 사용하여 검색된 결과는 〈표 15〉와 같다.

포괄성 측정에 앞서 검색결과와 검색조건과의 비교를 통한 레코드에 대한 적합성 여부를 분석하였다. 우선 이 연구의 분석대상인 KOLIS-NET은 전국의 487개 공공도서관 중 50.7%에 해당하는 247개관에서 분류 및 편목에 활용되고 있으며, 이들 도서관들의 참여로 구축되는 서지레코드를 통합하여 만든 '종합목록'이다. 따라서 다수의 구축주체로 인한 중복 및 오류레코드로 인한 부적합 레코드가 존재할 개연성이 상당히 높다. 따라서 검색결과를 갖고 부적합레코드를 제외하는 작업을 하였다. 아울러 비교 대상으로 선정된 서울 정독도서관, 부산 시립시민도서관, 대구 시립중앙도서관의 소장 DB를 대상으로 한 검색결과에 대해서도 동일하게 부적합한 레코드의 선별작업을 거쳤다.[6)]

이러한 선별작업을 거친 후 4개의 기관에서 동일한 키워드를 이용한 동일한 검색방법에 의해 검색된 결과를 갖고 서지 대조법을 이용하여 비교 평가를 실시하였다. 만약 전국의 단위 공공도서관이 자관에서 구축한 모든 자료를 업로드를 했다면, 원칙적으로 KOLIS-NET 종합목록은 전국의 단위 공공도서관이 구축한 모든 서지레코드를 포괄하고 있어야 한다. 그러나 〈표 3〉의 결과에서 보듯이 3개의 기관을 표본으로 하여 비교 평가한 결과, 3개 기관의 평균 포괄성이 73.4%였으며 포괄성은 전반적으로 매우 낮게 나타났다. 포괄성은 단위 공공도서관에서 검색된 레코드가 종합목록을 대상으로 검색된 레코드와 모두 중복될 경우를 포괄성 100으로 산정하였다. 구체적으로 살펴보면, 서울 정독도서관의 경우 48건의 레코드 중 검색되지 않거나 소장추기 되지 않은 레코드가 19

6) 이 과정에서 부적합레코드로 판명된 사례는 다음과 같다. (ㄱ) 중복레코드: KOLIS-NET 144건, 대구 시립중앙도서관 2건, (ㄴ) 단행본 자료가 아님에도 불구하고 실제로 검색된 레코드: KOLIS-NET 60건, 정독도서관 1건, 부산 시립시민도서관 8건, 대구 시립중앙도서관 162건.

건으로 포괄성이 60.4%로 가장 낮게 나타났다. 한편 부산 시립시민도서관의 경우, 검색된 54건의 레코드 중 9건의 레코드를 정독도서관의 경우와 같은 이유로 종합목록에서 찾을 수 없었으며, 대구 시립중앙도서관의 경우는 34개의 레코드 중 8개의 레코드가 종합목록에 포함되지 않은 걸로 나타나 포괄성이 76.5%로 낮게 나타났다.

KOLIS-NET과 같이 다수의 구축주체에 의해 운영되는 종합목록의 포괄성 문제는 운영주체인 국립중앙도서관의 관리 소홀 측면보다는 운영주체이자 참여자인 단위 공공도서관의 책임의식부족에서 비롯된다고 할 수 있다. 물론 포괄성 측정을 위해 표본으로 사용한 레코드의 수가 충분치 않아서 위의 결과만으로 단위 공공도서관이 책임감을 갖고 자관이 구축한 레코드를 성실히 업로드하고 있는지를 판단하는 것이 결코 쉬운 일은 아니다. 그렇지만 위 결과에서 제시하는 KOLIS-NET이 전국 단위 공공도서관이 구축한 레코드들의 포괄율이 매우 낮다는 사실이 시사하는 바는 무엇인가? 이는 단위 공공도서관이 신규구축자료에 대해 업로드를 하지 않고 있다는 것이며, 구축주체로서의 책임의식이 결여돼 있다는 것이다. 이러한 사실을 통해서 알 수 있는 보다 분명한 것은 단위 공공도서관이 책임감을 상실하여 자관이 신규로 구축한 레코드를 신속히 업로드 하지 않는다거나 자관에서 구축한 모든 자료를 시의 적절하게 업로드 하지 않는다면, KOLIS-NET는 이용자에게 국가자료 종합목록의 기능을 못할 뿐만 아니라 종합목록 구축의 본래 취지인 자원공유를 통한 분담편목, 분담수서 및 상호대차를 통한 적극적 이용자봉사는 요원하다.

3.2 중복성 평가

중복성은 DB의 절대적 유용성을 평가하는데 있어 매우 중요한 기준이다. 사용자가 원하는 자료를 찾고자 특정 DB를 검색할 때, 검색결과로 제공되는 레코드 중에 동일한 레코드가 중복되어 반복적으로 출현한다면 사용자는 해당 DB의 유용성에 대해 근본적인 의문을 갖는다. 그러나 구축 주체가 다수인 종합목록 DB의 경우, 중복레코드 발생 가능성은 상대적으로 높아 질 수밖에 없으며, 이러한 현상은 해당 종합목록 DB의 유용성을 저하시키는 주된 요인으로 지

〈표 3〉 실험용 탐색질의를 이용한 검색결과

DB명	검색 결과의 수정 (최초검색결과)	미 발견 레코드 수[7]	포괄성
KOLIS-NET	182(386건)	-	-
서울 정독도서관	48(49)	19	60.4%
부산 시립시민도서관	54(62)	9	83.3%
대구 시립중앙도서관	34(198)	8	76.5%

7) 미 발견 레코드란 비교대상 단위 도서관에서 소장한 레코드를 KOLIS-NET 종합목록 DB에서 찾을 수 없는 경우와 해당레코드가 검색이 된 경우라도 비교대상의 단위도서관명이 소장추기 되어 있지 않은 경우의 두 가지 유형을 의미한다.

적되고 있다. 분담편목 방식으로 구축되고 있는 KOLIS-NET 종합목록 DB도 이러한 이유로 인하여 단위도서관으로부터 활발한 이용을 이끌어 내지 못하고 있는 실정이다. 관리 주체인 국립중앙도서관은 2003년 KOLIS-NET에 서지레코드를 업로드하는 과정에서 발생한 오류 및 중복 데이터 100만 건을 처리하기 위하여 'KOLIS-NET 종합목록 DB 점검팀'을 구성하여 구축된 DB 전량을 점검하였다. 그리고 점검상태를 최종적으로 확인하는 'DB품질관리팀'을 운영하면서 중복 및 오류레코드에 대한 점검을 실시하였다. 이처럼 국립중앙도서관이 중복레코드 발생방지를 위해 다각도로 노력하고 있지만 중복레코드의 유형과 정확한 원인을 파악하는 것은 여전히 어려운 작업이다. 분담편목 및 자료공유를 통해 이용자 서비스의 최적화와 개별도서관의 편목업무의 비용절감을 기대하고 출발했던 KOLIS-NET은 이 문제의 해결 없이는 종합목록 DB의 활용도를 높일 수 없는 상황에 직면하고 있다.

연구팀은 2005년 9월, KOLIS-NET 종합목록 DB를 대상으로 중복레코드로 인한 심각성의 정도를 파악하기 위하여 표집분석을 통해 실증적으로 레코드의 중복성을 측정하였다. 표집분석을 위해서는 앞서 설명한 실험용 탐색질의 표제키워드 '정보사회'를 이용하여 KOLIS-NET 종합목록 DB로부터 검색한 '표본레코드'를 활용하였다.[8] 총 326건의 검색레코드 중에서 144건[9]이 중복레코드로 밝혀져 44.2%의 중복률을 보였다. 특히 동일 자료에 대한 레코드가 11건이나 중복하여 출현하는 사례가 있을 정도로 중복레코드로 인한 문제는 매우 심각한 상황이었다. 중복레코드가 이처럼 심각하게 발생하는 원인을 파악하고자 중복레코드를 유형별로 정리하여 발생비율을 조사하였다.

중복레코드의 유형을 파악함에 있어 연구팀은 다음과 같은 기준을 적용하였다. 첫째, 각 중복레코드마다 표준레코드[10]를 설정하여 비교해 본 결과 상이한 점 모두를 체크하였다. 둘째, 저자사항 표기 방식의 차이에서 한자와 한글은 동일한 것으로 보았다. 셋째, 서명 기술방식의 차이에서는 부서명의 유무는 있는 것을 기준으로 하였다.

이러한 기준에 의해 중복레코드의 유형을 파악했을 때 결과는 매우 다양하였다. 144건의 중복레코드 중에서 '출판년도의 차이로 인해 생

8) 총 386건의 검색레코드 중에서 검색오류로 분류된 60건을 제외한 326건이 분석대상이 된다.

9) 중복레코드의 발생 건수(144건)는 동일 자료로 판명된 총 수에서 레코드 하나만을 남긴 나머지의 수를 의미한다. 다음과 같은 계산식을 통해 산출된다.
-- 2번 중복된 레코드 30쌍에서 한 쌍에 한 개씩의 레코드를 제한 나머지 수: 30건
-- 3번 중복된 레코드 17쌍에서 한 쌍에 한 개씩의 레코드를 제한 나머지 수: 34건
-- 4번 중복된 레코드 8쌍에서 한 쌍에 한 개씩의 레코드를 제한 나머지 수: 24건
-- 5번 중복된 레코드 5쌍에서 한 쌍에 한 개씩의 레코드를 제한 나머지 수: 20건
-- 6번 중복된 레코드 4쌍에서 한 쌍에 한 개씩의 레코드를 제한 나머지 수: 18건
-- 8번 중복된 레코드 2쌍에서 한 쌍에 한 개씩의 레코드를 제한 나머지 수: 8건
-- 11번 중복된 레코드 1쌍에서 한 쌍에 한 개씩의 레코드를 제한 나머지 수: 10건

10) 중복으로 판명된 레코드 내에서 그중 한 레코드를 기준으로 설정하기 위하여, 연구팀은 KERIS의 종합목록, 대규모 대학 도서관의 온라인 목록, 출판목록 등과의 비교를 통하여 원자료와 동일하게 내용을 기술한 레코드를 선별하여, 그 레코드를 기준으로 중복레코드의 유형을 분석하였다.

성된 중복레코드가 가장 많았으며 저자사항표기방식의 차이, 서명기술방식의 차이, 형태사항 필드의 누락, ISBN 필드의 누락, 그리고 기타 등으로 인해 생성된 중복레코드 순으로 나타났다. 특히 이중 대표적인 유형에 대해 간략하게 살펴보고자 한다.

레코드의 중복이 발생한 원인 중 가장 많은 경우가 '출판년도'의 차이로 인한 것이었다. 표본레코드를 분석한 결과 동일한 자료에 대해 출판년도가 1년 또는 2년의 차이로 인해 중복으로 생성된 레코드들이 많았다. 정확한 출판년도를 확인하기 위해 KOLIS-NET의 MARC 레코드를 찾아 조사하였으나 대부분의 레코드는 정확한 근거를 찾을 수가 없었다. 따라서 연구팀은 '판차사항' 필드에 출판년도가 따로 구분이 되어 있는 경우는 중복레코드가 아닌 것으로 판단하였고, '판차사항' 필드에 아무런 데이터도 수록되어 있지 않고 다른 모든 사항이 동일함에도 불구하고 출판년도가 다르게 기술된 레코드들을 중복레코드로 처리하였다. 이는 편목담당사서들이 레코드 기술에 있어 '쇄'를 '판'으로 오인한 것에서 비롯된 것보다는 단위도서관에서 동일한 판임에도 불구하고 자관에서 소유하고 있는 자료가 최신자료라는 것을 부각시키기 위하여 출판년을 무시하고 쇄 연도를 기술사항으로 채택하는 잘못된 이기심 또는 관행이 그 원인인 것으로 추측된다.

저자사항을 표기하는 방식의 차이로 인해 중복이 발생한 경우, 표본레코드를 분석한 결과 동일한 자료에 대해 '저자'를 표기하는 방식의 차이는 다음의 몇 가지 유형으로 요약될 수 있다. 첫째, 저자 역할어의 표기상의 차이로 중복레코드가 발생한 경우(예를 들면 김병준...[등] 저 ; 김병준...외[등저], 현규섭 저 ; 현규섭 지음), 둘째, 외국인명을 한글로 표기한 경우와 원어 그대로 표기한 레코드(예를 들면, 마쓰오 사브로 著 ; 松尾三郎 著 ; 增田米二 ; コニピュータ 著), 셋째, 동일한 레코드 임에도 불구하고 저자가 완전히 상이한 경우, 넷째, 공저자의 경우, 표기에 있어서 책임사항을 모두 나열한 경우와 주요 책임사항만을 표기하고 생략한 레코드, 다섯째, 연구단체에서 발간된 자료의 경우 연구주관명을 표기한 레코드와 연구책임자명 또는 기관명과 책임자명을 병기 표기한 레코드(예를 들면 정보사회학편 ; 조상호, 정보사회학회 편). 이와 같이 동일한 자료임에도 불구하고 개별 레코드에 표기된 '저자사항'의 차이로 인해, 종합목록 DB에서는 이들 레코드가 각기 다른 것으로 인식되어 반복적으로 수록되어 나타나고 있다. 한편 서명의 기술방식에 있어서의 차이로 인해서도 많은 중복레코드가 발생하였다. 모두 59건, 즉 40.9%의 레코드가 이 유형에 해당되었다. 구체적으로 권차 제목을 서명관련사항으로 입력했거나 '본서명 이외에 부서명이나 잡제 등의 기타 서명 관련 정보의 유무'에 따라 또는 관제 및 관칭을 다르게 처리함으로써 동일 자료에 대한 레코드가 중복 생성되었다.

다른 유형의 대표적 사례로는 ISBN 필드의 누락과 형태사항 필드의 누락이 있다. 이 중 특히 심각한 것은 ISBN 필드의 누락이다. 왜냐하면 이 필드는 KOLIS-NET에서 중복레코드를 처리하기 위한 알고리즘에서 UB 제어번호를 제외할 경우 가장 중요한 변별 요소로 취급하는 항목이다. 다시 말해 ISBN 10자리 일치 검색만으로 충분히 중복레코드로 판별될 수 있는 레코드

〈표 4〉 중복레코드의 유형과 발생비율

중복레코드 발생 유형	발생건수	발생비율
저자사항 표기방식의 차이	78	54.1%
출판년도의 차이	89	61.8%
서명 기술방식의 차이	59	40.9%
ISBN 필드의 누락	39	18.1%
형태사항 필드의 누락	43	13.9%
기타	1	0.7%

* 발생비율은 중복레코드 건수 144건의 레코드를 대상으로 산출한 값임.
* 동일한 레코드에 대해 다수의 중복 유형이 있을 경우 중복 계산.

가 단위 공공도서관에서 레코드를 생산할 때 목록규칙의 미 준수로 인하여 중복레코드 발생의 원인을 제공하고 있다. 사실, 중복레코드로 판명된 사례 중 외국자료의 저자사항 표기방식의 차이로 인한 중복 사례는 단위 도서관에 따라 자의적으로 판단하고 기술할 수 있는 목록규칙의 적용상의 차이로 발생한 중복 유형이다. 그러나 ISBN 필드는 목록규칙의 적용에 있어 해당시 필수 항목으로서 단위 공공도서관의 자의적 판단이 개입될 여지가 없는 사항이다. 이는 공공도서관의 편목담당사서들의 목록규칙에 대한 이해의 부족뿐만 아니라 책임의식의 결여에서 비롯된 중복레코드 생성 유형에 해당된다고 할 수 있다. 이처럼 KOLIS-NET의 목록 DB을 대상으로 한 중복레코드의 발생 원인은 국내 선행연구들[11]이 지적한 데이터의 비표준화 문제, 사서들의 목록시스템 사용 미숙, 전문적인 지식을 요구하는 목록기술규칙과 MARC에 대한 이해부족 등이 복합적으로 작용하고 있다. 〈표 4〉는 분석대상 표본레코드에서 파악된 중복레코드의 유형과 발생건수 그리고 발생비율을 요약해 놓은 것이다.

3.3 최신성 평가

최신성을 측정하는 목적은 KOLIS-NET 종합목록 DB가 분담편목에 대한 단위 공공도서관의 요구를 얼마나 많이 충족시켜줄 수 있는가를 평가하는데 있다. 즉, KOLIS-NET 종합목록 DB에서 사용할 수 있는 최신 자료의 서지레코드 수가 많으면 많을수록 단위도서관의 편목담당자가 느끼는 KOLIS-NET 종합목록 DB의 유용성과 활용성은 더욱 클 것이라는 것을 전제로 한다. 최신성은 KOLIS-NET 종합목록 DB의 신규레코드 구축이 참여도서관 상호간 얼마나 적극적으로 그리고 신속하게 DB 구축에 기여하고 있는지를 판단할 수 있는 기준이 될 수 있다.

앞서 설명했듯이, 최신성 측정을 위해 연구팀

11) 이제환, KERIS 서지 데이터베이스의 품질관리를 위한 평가모델 개발 및 개선방안 수립, 서울: 한국교육학술정보원, 2001.
이지은, 첨단학술정보센터 종합목록데이터베이스 품질관리에 관한 연구, 석사학위논문, 숙명여자대학교 : 문헌정보학과, 1999.
김지훈, 서지데이터베이스의 품질관리-k관의 MARC 레코드 분석을 중심으로, 도서관학논집 제21집(1994) : 401-429.

은 단위 도서관의 최신 수서자료리스트를 이용하여 검색을 통한 서지대조법을 이용하였다. 이때 최신 수서자료리스트의 최신성을 100으로 하여 KOLIS-NET 종합목록 DB의 최신성을 측정하였다. 단위도서관의 2005년 7월말 수서자료리스트를 확보한 후, 연구팀은 최신성의 측정시기를 언제로 할 것인가에 대해 논의하였다. 연구팀은 최종적으로 수서자료리스트의 확보시점에서 한 달이 경과한 후에 최신성을 측정하기로 결정하였다. 수서자료리스트 확보후의 한 달이라는 시간은 책의 출간일을 기준으로 했을 경우 2-3개월의 시간이 경과한 시점이며, 개별 도서관을 기준으로 했을 경우, 주문한 자료가 도서관에 도착하여 정리를 기다리고 있는 시점이라는 판단에서다. 이 기간이면 구축주체 및 참여도서관의 협력적 노력만 있다면 최신자료의 신규레코드가 충분히 생성될 수 있는 기간이라고 판단하였다. 이러한 기준에 의해 연구팀은 단위도서관의 수서자료리스트를 활용하여 8월 25-26일 양일간에 걸쳐 KOLIS-NET 종합목록 DB에서 필요한 서지 레코드의 이용가능성 여부를 검색하였다.

〈표 5〉에서 보듯이 KOLIS-NET 종합목록 DB의 최신성은 36.2%로 나타났다. 즉 KOLIS-NET 종합목록 DB에서 검색된 레코드는 921건 중 369건에 불과했으며 나머지 552건에 대한 서지레코드는 종합목록 DB에서 찾을 수가 없었다. 이러한 결과는 최신자료의 서지레코드를 작성하고자 하는 단위 공공도서관의 편목담당사서에게 KOLIS-NET 종합목록 DB는 더 이상 유용한 도구가 되지 못하고 있음을 실증적으로 보여준다. 즉 KOLIS-NET 종합목록시스템의 구축으로 서지정보의 공동활용을 통하여 이용자 서비스와 도서관 업무의 능률을 향상시키고자 했던 KOLIS-NET의 본래 취지가 기능을 발휘하지 못하고 있음이 실증적으로 나타났다. 연구팀은 최신성을 측정하기에 앞서 이러한 결과를 예상했었다. 왜냐하면 연구팀이 실시한 공공도서관의 편목담당사서들을 대상으로 한 설문조사에서 편목담당사서들은 KOLIS-NET의 유용성을 높이기 위한 방안으로 최신자료의 신속한 반영을 언급하면서 서지레코드의 공동활용을 위해 자관레코드의 업로드와 관련한 노력은 거의 하고 있지 않은 것으로 나타났었기 때문이다. 최신성 측정 결과가 시사하는 것은 KOLIS-NET이 명실상부한 국가자료 종합목록 DB로서의 역할을 수행하기 위해서는 종합목록 DB의 구축주체이자 실질적인 이용주체인 단위 도서관의 편목담당사서들의 인식전환이 절실히 필요하다는 것이다. 아울러 관리주체인 국립중앙도서관은 이들의 협력을 이끌어 낼 수 있는 다양한 정책적 방안을 강구해야 한다.

3.4 정확성 평가

데이터표기의 정확성은 레코드의 데이터 필

〈표 5〉 최신 수서자료리스트를 이용하여 측정한 종합목록DB의 최신성

전체 표본레코드 수	검색된 레코드 수 / 검색되지 않은 레코드 수	최신성
921	369/552	36.2%

드에 기술된 데이터가 原자료에 나타나는 데이터 값을 정확하게 표기하고 있는지를 측정하기 위한 지표이다. 정확성은 原자료의 내용을 정확하게 기술해 줌으로써 이용자가 자료의 유용성을 판단할 수 있는 근거가 되는 지표이다.

KOLIS-NET의 종합목록 DB는 단위 공공도서관에서 작성한 서지레코드를 종합한 DB로서 레코드에 수록되는 데이터의 표기 방식은 단위 공공도서관에서 적용하고 있는 목록규칙과 내부 규정을 따르게 된다. 그러나 국내 도서관에서 생산되는 목록레코드는 표기 방식에 대한 명확한 규정의 부재로 인해, 동일한 데이터 값에 대한 표기가 개별 공공도서관마다 심지어는 개별 사서마다 제각각의 형태로 나타날 수 있다(이제환 외 2001). 이러한 문제적 상황이 그대로 반영될 개연성이 높은 KOLIS-NET 종합목록 DB에서 데이터 표기의 정확성 문제를 논의한다는 것은 매우 어리석은 작업이 될 수도 있다.

그래서 연구팀은 정확성 측정을 위해 〈표본레코드〉를 대상으로 데이터 값이 명백하게 부정확하다고 판단된 표기 사례를 조사하여 '부정확한 데이터 표기의 유형'을 분석하였다. 그리고 이러한 부정확하게 표기된 데이터를 하나라도 가지고 있는 레코드의 수를 모두 계량화하여 종합목록 DB에서 부정확하게 표기된 레코드의 발생비율이 어느 정도인지를 예측해 보고자 하였다. 또한 부정확하게 표기된 데이터의 발생요인을 추적하였으며 발생요인의 유형별 사례는 다음과 같다.

- 철자오류: 목록레코드에 표기되는 데이터는 原자료에 나타나는 문자와 언어를 그대로 표기하는 것이 원칙이다. 그러나 분석을 위해 사용한 표본레코드에서는 다수의 '오자'와 '탈자'가 발견되어 연구팀은 이를 철자오류로 규정하였다. 다만 原자료와 대조하지 않고서는 '오류'로 판단하기 어려운 '표기문자'의 차이(예를 들면, 표제지의 표제명이 한자로 기술되어 있으나 기술은 한글로 표기한 경우)는 측정에서 제외하였다.
- 띄어쓰기 오류: 原자료에 기술된 데이터를 표기하는 과정에서 '띄어쓰기'가 잘못된 사례를 측정하였다. 그러나 原자료와의 대조를 통하지 않고서는 '오류'로 판단하여 계량화하는 것이 불가능하므로 명백하게 드러나는 오류 이외에는 '중복레코드'를 이용한 비교를 통해 측정하였다. 그리고 목록규칙에는 맞지 않지만 모든 서지레코드에서 동일하게 그리고 일관되게 나타나는 오류는 측정에서 제외하였다.[12)]
- 동일한 레코드 내에서 페이지 수의 표기상의 차이: 동일한 중복레코드 내에서 페이지 수의 표기상에 차이가 있거나 페이지 수가 완전히 다른 경우 등을 오류로 분류하였다.
- 중복 기술된 데이터의 값이 서로 일치하지 않는 경우: 단행본 레코드에서 출판사항 필드의 출판년도(260필드의 $c)와 출

12) 예를 들면, KCR4판에 의하면 형태사항의 기술에 있어 쪽·장·권·책의 수와 크기의 단위 명칭을 로마자, 'p.', 'v.', 'leaves', 'cm', 'mm', 등으로 표시할 경우는 숫자와 단위 명칭 사이를 띄어쓰기 할 것을 원칙으로 채택하고 있다. 그러나 KOLIS-NET 종합목록 DB의 모든 레코드는 이와 같은 단위명칭과 숫자사이를 띄어쓰기 원칙을 적용하고 있지 않다.

판년 필드의 출판년도(008필드의 발행년), 출판사항 필드의 출판지(260필드의 $a)와 발행국 데이터 그리고 형태사항 필드의 삽도(300필드의 $b)와 삽도표시 데이터[13] 등은 하나의 레코드에 동일한 데이터가 중복으로 기술된 전형적인 사례이다. 그러나 이들 데이터가 서로 일치하지 않아 둘 중 하나가 잘못 표기되었다고 판단된 경우, 이를 데이터 표기의 오류로 규정하였다.

- 데이터 필드와 수록 데이터의 불일치 사례: 레코드에서 제시하고 있는 데이터 필드명과 기술된 데이터의 내용이 일치하지 않는 레코드를 조사하였다. 가령, 부출표목 또는 기본표목의 사용[14]에 있어서 개인명과 단체명의 태그를 잘못 사용한 경우나 총서사항 필드의 태그 사용에 있어서의 오류 또한 이곳에 분류하였다. 그리고 245 필드의 기술순서에 있어서의 오류도 이곳에 분류되었다.

〈표 6〉은 위에 제시된 절차와 방법에 따라, 표본레코드를 대상으로 데이터의 표기가 부정확한 레코드의 유형을 보여주고 있다. 데이터 표기의 정확성에 있어서 가장 많은 오류가 발생한 경우는 데이터 필드의 사용상의 오류로 16.9%로 나타났다. 구체적 내용을 살펴보면, 기본표목 및 부출표목의 사용에 있어서 단체명임에도 불구하고 필드명을 100 또는 700으로 잘못 표기한 경우가 총 17건에 달했다. 또한 서명저자필드인 245 필드의 기술순서에 있어서 [245 00 $a원서명 = $x대등서명 : $b부서명. $n권차, $p권차서명 / $d저자 ; $e공저자] 순으로 기재 되어야 함에도 불구하고 이의 순서가 지켜지지 않은 경우도 15건으로 나타났다. 다음으로는 原자료를 기술함에 있어 철자상의 오류로 지적된 레코드가 15.6%였으며 그리고 특히 하나의 레코드에 동일한 값이 중복으로 기술될 경우 데이터의 값이 불일치한 경우도 37건으로 11.4%를 기록했다.

이러한 데이터 요소의 정확성 오류의 발생 방지를 위해서는 기술적 요소인 오류처리 알고리즘으로 처리할 수 있는 부분보다는 전적으로 편목담당사서들의 서지레코드 작성에 대한 책임의식 부족과 정성의 문제라고 생각된다. 철자오류 및 띄어쓰기 오류 등은 아무리 효과적인 오류처리 알고리즘으로도 해결 될 수 없는 전적으로 사람에 의해서 판별될 수밖에 없는 요소이다. 한편 동일한 레코드 내에서의 페이지 수의 표기상의 차이는 동일한 레코드를 서로 다른 레코드로 인식하게 만드는 요인이 될 수 있어 중복레코드의 발생의 원인이 되고 있다. 이러한 유형의 오류는 '데이터 입력지침'이 있다면 충분히 예방이 가능한 문제이다. KOLIS-NET 종합목

13) 출판지는 '군포'로 되어 있으나, 발행국에는 '서울'로 표기되어 있는 경우 또는 삽도가 있다고 기술되어 있으나 형태사항 필드의 데이터 요소사항에는 기술되지 않은 경우 또는 그 반대의 경우가 이에 해당된다.

14) 기본표목 1XX의 적용여부는 단위 공공도서관에 따라 적용에 있어 다르게 판단할 수 있는 부분이기 때문에 오류로 취급하지 않았다. 목록업무자동화가 추진되면서 기본기입 레코드와 부기입 레코드에 기재되는 서지정보의 내용에 차이가 없어졌고 전산화에 의해 저작의 검색수단이 다양해짐에 따라 기본기입과 부기입은 동일한 레코드에 대한 검색요소로서 대등한 기능을 가지게 되었다. 이러한 현상을 반영하듯 단위 공공도서관에 따라 기본표목 1XX를 적용하고 있는 기관이 있는가 하면, 기본표목 1XX를 적용하지 않고 부출표목인 7XX만을 사용하고 있는 기관이 있다.

〈표 6〉 데이터 표기별 표기 오류의 발생건수 및 비율

데이터 필드	발생건수	발생비율
철자오류	51	15.6%
데이터 필드와 데이터의 불일치 - 기본표목 필드 오류 - 부출표목 필드의 오류 - 총서사항 필드의 오류 - 245 필드의 기술순서 오류	55 8 9 23 15	16.9%
중복 기술된 데이터 값의 불일치	37	11.4%
동일한 레코드 내에서 페이지 수의 표기 차이	30	9.2%
띄어쓰기 오류	13	4.0%

* 발생비율은 326건의 레코드를 대상으로 산출한 값임.

록 DB에 대한 유용성과 활용성 확보를 위해서는 우선 단위 도서관의 편목담당사서의 책임의식 강화를 위한 정책적 방안이 필요하다. 더불어 종합목록 DB의 품질관리를 담당하고 있는 국립중앙도서관은 단위도서관의 목록 작성을 도와줄 수 있는 '표준 레코드' 작성 지침 등 제도적 장치 마련 및 사서들을 위한 목록규칙의 올바른 이해와 적용을 위한 계속교육을 실시해야 한다.

3.5 일관성 평가

데이터 표기의 일관성은 정보를 제공함에 있어 레코드 구조가 상이하거나 데이터 필드에 관련 데이터를 표기 할 때에 일관된 방식을 따르지 않아 동일한 내용의 데이터가 서로 다른 방식으로 표기된 레코드의 발생비율을 측정하기 위한 지표이다. 데이터의 표기 방식이 목록레코드의 작성기관이나 작성자에 따라 달라질 때, 동일한 자료에 대한 중복레코드의 생성 가능성은 매우 높아진다. 종합목록 DB의 경우 중복레코드의 발생이 이러한 레코드 구조나 데이터표기의 일관성 결여에서 비롯되는 경우가 많다. KOLIS-NET과 같은 종합목록 DB는 많은 수의 공공도서관이 목록 구축주체로 참여하고 있고, 이들이 적용하는 목록규칙이 단위 도서관에 따라 상이하여 레코드 구조와 데이터 표기의 일관성을 일반화하여 계량화하는 것이 쉽지 않다. 만약 '표준' 표기 방식이 존재한다면 표기의 오류를 지적하는 것은 어려운 작업이 아닐 수 있다. 그러나 도서관마다 사서에 따라 목록규칙이 상이한 현실에서 '표준' 표기 방식을 설정하는 것이 용이하지 않다.

기존의 여러 목록규칙들이 내포하고 있는 표기방식의 다양성과 목록 작성자의 자의적인 해석과 적용에 따라 데이터의 표현과 기술이 다를 수밖에 없는 현실을 감안하여 연구팀은 일정한 방식을 표준으로 지정하지 않았다. 대신 데이터 표기의 일관성, 즉 레코드 구조의 일관성을 결여하고 있는 대표적인 사례를 분석하여 일관성 측정의 결과로 제시하고자 한다. KOLIS-NET 종합목록 DB의 일관성의 결여는 결국 중복레코드의 발생을 양산하는 원인으로 지적되고 있기 때문에 연구팀은 KOLIS-NET 종합목록의 일관성 문제가 어느 정도 심각한지를 파악하기 위하여 표집분석을 통한 체계적인 검증작업을

실시하였다. 앞서 설명한 동일한 표본레코드를 활용하였으며, 측정에 있어 경우에 따라서 데이터의 정확성 표기의 오류에서 지적된 사항이 동일하게 레코드 구조의 일관성 오류로 지적되기도 하였다.

레코드 구조의 일관성을 측정하기 위해 다음의 사례를 중점적으로 분석하였다: i) 대부분의 레코드가 공통적으로 지니고 있는 기본 데이터 필드에서 특정 데이터의 필드 자체가 누락되거나 ii) 특정 데이터의 필드에서 데이터가 누락되거나 첨가된 경우, iii) 레코드를 구성하는 기본 데이터 필드의 나열순서가 상이한 경우. 이 작업을 위해서 KOLIS-NET 종합목록 DB를 구성하는 서지 레코드의 '기본 데이터 필드'를 어떻게 볼 것인가를 결정해야한다. 이에 연구팀은 표집분석에 활용한 〈표본레코드〉 집단에 수록된 모든 레코드를 대상으로 하여, 각 레코드에 단 한번이라도 나타나는 데이터 필드를 모두 점검하였다. 그 결과 단행본의 일반자료 레코드에 나타나는 데이터 필드들은 다음과 같이 최대 10개의 항목으로 구성되어 있었다.

① 표제
② 저작사항
③ 발행사항
④ 형태사항
⑤ 총서사항
⑥ 주기사항
⑦ ISBN
⑧ 분류기호(KDC, DDC)
⑨ 저자사항[15)]
⑩ 열람가능도서관

위에서 보듯이, 웹상에서 탐색작업의 결과로 사용자들이 접하게 되는 단행본의 서지레코드는 모두 10개의 데이터 필드가 사용되고 있다. 단행본 서지레코드 구조에서 발견되는 중요한 특징은 레코드에 따라 사용하고 있는 데이터 필드의 수에 있어 많은 차이가 있다는 것이다. 예를 들면 10개의 데이터 필드 중에서 적게는 5개의 필드(①, ②, ③, ④, ⑧)를 사용한 경우가 하면 대부분의 레코드는 7-8개의 데이터 필드(①, ②, ③, ④, ⑦, ⑧, ⑨, ⑩)들로 구성되어 있다. 한편 자료의 특성에 따라, 즉 총서사항이 있거나 번역서의 경우 등, 10개에 해당하는 모든 레코드를 사용하였으며 이는 전체 표본레코드 중 일부에 해당한다.

KOLIS-NET 종합목록 DB에서 제공하는 10개의 서지 데이터 필드의 사용은 목록규칙에서 제시하는 해당시 필수 항목으로 지정된 기술사항이다. 따라서 연구팀은 이들 항목 중 총서사항과 주기사항을 제외한 8개의 데이터 필드를 서지기술의 필수 기술사항임과 동시에 서지 기술의 최소 수준으로 판단하였다. 그렇지만 총서사항과 주기사항도 자료의 성질에 따라 기술되어야 할 주요한 기술요소이므로 비교 가능한 동일레코드가 존재할 경우 레코드 구조의 일관성 측면에서 비교·분석하였다. 즉 연구팀은 KOLIS-NET 종합목록 DB의 일관성 측정을 위해 KOLIS-NET 종합목록이 사용한 10개의 데이터 필드를 모두를 분석대상으로 보았

15) 저자사항 필드는 단위 공공도서관에 따라 적용에 있어 차이가 있다. 가령, 일부 도서관은 부차적 책임표시를 나타내는 부출을 사용할 때 이 필드를 사용한 반면, 일부 도서관은 1XX의 기본 표목 필드를 사용하지 않는 대신 책임사항을 기술할 때 이 필드를 사용하였다.

으며, 이 필드들이 누락된 경우를 조사하여 〈레코드 구조의 일관성〉 정도를 측정하였다.

그러나 실제 측정에 있어서 몇 가지 예외를 두었다. 예를 들면, 웹에서의 검색결과만으로 총서사항 필드의 누락을 확인한다는 것이 어렵기 때문에 이 필드의 누락은 동일한 레코드가 존재할 경우 그 기준이 되는 레코드에 준하여 누락여부를 판단하였다. 그리고 일반주기사항 중 일부의 경우, 기술이 필요함에도 불구하고 누락되어 있다는 사실 확인이 불가능하고, 도서관에 따라 서지기술 수준을 자의적으로 판단할 수 있다고 판단되어 측정에서 제외하였다. 그러나 번역서의 경우는 원저자와 원저자명을 표기하는 것이 원칙이므로 그 기준에 근거하여 누락여부를 판단하였다. 그리고 분류번호 필드는 KDC와 DDC 모두를 표기한 레코드와 KDC 하나만을 표기한 레코드의 비중이 65:35의 비율로 나타났다. 따라서 두 필드 중 모두 누락된 경우에만 분류번호 데이터 필드의 누락으로 간주하였다.

이러한 기준에 의하여 레코드 구조의 일관성을 측정한 결과는 〈표 7〉과 같다. 측정결과를 살펴보면 우선, 레코드를 구성하는 10개의 데이터 필드 중에서 하나의 필드라도 누락된 레코드가 106건(31.3%)이었으며, 이러한 결과는 3개의 레코드 중 한 개의 레코드가 데이터 필드의 기술에 있어서 최소 1개 이상의 누락이 발생하고 있음을 예측가능하게 한다.

구체적으로 살펴보면 레코드의 중복여부를 판단할 수 있는 중요한 데이터 필드인 ISBN 필드의 누락[16]이 12%인 39건으로 가장 많았으며, 다음으로는 총서사항을 기술하지 않은 레코드가 25건, 형태사항필드의 누락이 17건 그리고 번역서의 경우 원서명 필드의 누락이 4.3%인 14건으로 나타났다. 앞서 중복레코드의 발생과 관련지어 설명했듯이 ISBN 필드나 형태사항 필드의 누락은 중복레코드 처리에 있어 주요한 변별 요소에 해당한다. 따라서 이러한 필드들의 누락은 단순하게 레코드 구조의 일관성 차원에서 접근하기보다는 이러한 누락이 결국은 중복레코드의 발생에 기여하고 있으며 이는 결국 종합목록의 유용성을 저하시켜 단위 공공도서관의 KOLIS-NET에 대한 이용기피 현상을 초래하고 있다.

〈표 7〉 레코드 구조의 일관성 측정결과

오류사례	발생건수	발생비율*
형태사항 필드의 누락	17	5.2%
ISBN 필드의 누락	39	12.0%
판사항 필드의 누락	8	2.5%
총서사항 필드의 누락	25	7.7%
원서명 필드의 누락	14	4.3%
저자사항 필드의 누락	1	0.3%
언어 필드의 누락	1	0.3%
발행국 필드의 누락	1	0.3%

* 표본레코드에 포함된 단행본 326건의 레코드에 대한 산출 값임

16) ISBN 필드의 누락은 原자료와의 비교를 통하지 않고서는 누락을 판단하는 것에 한계가 있다. 따라서 동일 레코드가 존재할 경우에 한해서 누락을 계량화하였다. 이처럼 중복레코드 내에서 누락을 측정했음에도 39건이 발견되었는데, 만약 이 필드의 누락을 原자료와의 대조를 통해서 측정했다면 결과는 많이 달라질 것으로 판단된다.

그러나 아이러니하게도 이러한 종합목록 DB의 품질저하현상을 초래한 것은 단위 공공도서관 편목담당사서라고 말할 수 있다. 물론 국립중앙도서관에서 종합목록 DB 구축을 위한 통일되고 일관된 지침을 단위 도서관에 문서화된 형태로 내려준 적은 없지만 편목담당사서가 서지레코드 기술과 관련한 기본적인 규칙과 기준에 대한 충분한 숙지만 있었다면 많은 경우 중복레코드의 발생을 방지 할 수 있다. 레코드 구조의 일관성 측성을 통해서 알 수 있는 것은 단위 공공도서관이 공통적으로 적용할 수 있는 지침의 제정이 절실히 필요하다는 것이다. 또한 정확하고 적절한 이들 지침의 적용을 위해 편목담당사서를 대상으로 한 기술규칙에 대한 지속적인 교육이 필요하며, 현재처럼 현장의 편목 담당자를 업무 순환 차원에서 인사이동시킬 것이 아니라 소속기관이 변경되더라도 동일한 업무를 담당하게 함으로써 국가자료 종합목록 DB를 구축할 수 있는 정예화된 전문가로 육성할 수 있는 정책적 방안이 강구되어야 할 것이다. 기술적으로 해결할 수 있는 부분이 제한적인 상황에서 이러한 정책적 방안이 실현되지 않는 한 KOLIS-NET 종합목록 DB의 중복 및 부실 레코드를 생산은 반복될 수밖에 없다.

3.6 완전성 평가

완전성은 DB를 구성하는 단위 레코드들의 내용이 얼마나 완전하게 기술되었는지를 평가하기 위한 기준이다. DB를 구성하는 레코드는 原자료에 대한 대체물의 역할을 하기 때문에 레코드의 내용은 사용자의 입장에서 볼 때 DB의 유용성을 결정짓는 기본적인 요소가 된다. 따라서 종합목록 DB를 구성하는 레코드는 原자료의 존재를 파악하고 식별하는데 필요한 데이터(서지데이터)와 原자료에 접근하여 이용하는데 필요한 데이터(소장데이터)를 완전한 형태로 포함하고 있어야 한다. 만약 개별레코드의 내용이 불완전하다면, 검색과정에서 누락되는 것은 물론이고 검색되더라도 사용자의 정보요구를 충족시키는데 실패하기 때문이다.

연구팀은 KOLIS-NET 종합목록 DB의 완전성을 측정하기 위하여 레코드를 구성하는 데이터 필드와 그 데이터 필드를 구성하는 관련 데이터 요소가 완전하게 기술되어 있는지에 초점을 뒀다. 즉, KOLIS-NET 종합목록 DB의 완전성을 측정하는 기준으로 데이터 필드의 완전함과 필드를 구성하고 있는 데이터 요소의 누락이 없는지를 정확성 측정의 대상으로 하였다. 따라서 완전성은 레코드를 구성하는 데이터 필드에 반드시 수록되어야 할 데이터가 누락되거나 혹은 불완전하게 수록된 레코드의 발생 비율을 조사하여 측정한다. 이 지표의 측정을 위해서 다음의 요소를 중점적으로 파악하였다.

- 특정 데이터 필드의 누락 사례: 데이터 필드에 데이터가 누락되어 있는 레코드를 조사하였다. 즉 데이터 필드는 있는데 필드를 채울 데이터 내용이 존재하지 않는 경우를 특정 데이터 필드의 누락으로 간주하였다.
- 데이터의 부분기술 또는 일부 데이터의 누락 사례: 하나의 데이터 필드에서 기술하고 있는 데이터의 내용이 복수의 항목을 제시하고 있는 경우거나 데이터 필드에 데이터가 기술되어 있기는 하지만 부분적으로만 기술되어 있는 레코드를 조사하였다.

예를 들면 출판사항필드의 경우, 출판지, 출판사, 출판년도 데이터 중 데이터가 하나라도 누락된 경우는 모두 데이터의 누락으로 간주하였다. 그리고 原자료를 보지 않은 상태에서 누락을 측정하는 것이 어려운 서명필드나 총서사항필드의 데이터 누락은 동일한 중복레코드가 존재할 경우로 한정하여 대등표제, 표제관련정보 및 총서사항이 있는 레코드를 기준으로 누락여부를 측정하였다. 이처럼 부분기술의 사례는 중복레코드가 발생한 경우에 한해서 내용을 객관적으로 비교할 수 있는 레코드를 대상으로 조사하였다.

특정 데이터 필드의 누락 및 데이터 값의 누락 측정은 다음의 기준을 갖고 접근했다. 우선 서명필드와 저자필드는 245필드를 대상으로 데이터 누락을 측정했다. 그리고 형태사항 필드의 데이터 누락은 008필드와 300필드의 데이터 누락을 모두 합친 결과이며, 언어, 출판년, 발행국 필드의 데이터 누락은 008필드에서의 누락을 의미한다. 마지막으로 총서사항 필드의 데이터 누락은 동일레코드가 있을 경우로 한정하여 측정하였으며 편서명이나 총서번호의 누락도 데이터의 누락에 포함하였다.

측정결과 데이터의 누락 및 오류는 두 가지 형태, 즉 필드가 존재함에도 불구하고 데이터 자체가 입력되지 않은 경우와 데이터가 입력되었지만 부분적으로 기술된 경우로 나타났다. 이들 두 가지 사례 -데이터 필드의 누락과 특정필드의 데이터의 부분기술- 를 종합하여 분석한 결과 데이터 자체가 입력되지 않아 데이터 필드의 누락이 발생한 경우는 '서명', '형태사항', '총서사항' 및 '판사항' 필드와 같이 데이터 값을 그대로 기술해야 하는 데이터 필드(Variable Field)에서 대부분 발생한 반면 데이터가 잘못 입력되어 데이터의 누락이 발생한 경우는 '언어'나 '발행국' 필드와 같이 데이터 값을 특정부호로 변환하여 입력해야 하는 제어필드(Control Field)에서 발생하였다. 그러나 후자의 경우는 전자에 비해 누락된 사례가 극히 소수에 불과했다.

필드자체의 누락과 데이터의 누락 정도를 살펴보면 '서명' 필드와 '형태사항' 필드의 누락이 가장 심각했으며 그 다음으로 '저자' 필드의 데이터 누락, '총서사항' 필드의 데이터 누락 순으로 나타났다. 데이터의 누락은 비교 가능한 동일 레코드가 있을 경우로 한정하여 기술사항이 있는 레코드를 '기준' 레코드로 설정하여 측정했음에도 불구하고 '서명' 필드나 '총서사항' 필드에서 데이터 누락이 많이 발생했다. 이 결과

〈표 8〉 데이터 필드의 누락 및 오류의 발생건수 및 비율

오류 사례	발생건수	발생비율
'서명' 필드의 데이터 누락	43	13.2%
'형태사항' 필드의 데이터 누락	43	13.2%
'저자' 필드의 데이터 누락	19	5.9%
'총서사항' 필드의 데이터 누락	16	4.9%
'출판년' 필드의 데이터 누락	3	0.9%
'발행국' 및 '언어' 필드의 데이터 누락	2	0.6%

* 발생비율은 326건을 대상으로 한 것임

는 만약 原자료와의 비교를 통해 데이터 누락을 측정한다면 누락율 측정값이 지금보다 상당히 높아질 수 있다는 것을 의미한다.

4. 결 론

본 연구는 유용성을 중심에 두고 KOLIS-NET 종합목록 DB의 품질을 측정하였다. 종합목록 DB의 구축 목적과 특성을 고려할 때, 사용자에게 유용성을 인정받지 못할 정도의 DB 품질을 보유하고 있다면 품질향상을 위한 전면적인 검토가 필요하다고 판단하였기 때문이다. 본 연구는 KOLIS-NET 종합목록 DB의 품질 평가를 위한 기준으로 포괄성, 최신성, 중복성, 일관성, 완전성, 정확성의 6개 항목을 선정하였다. 유용성의 기준에서 봤을 때 포괄성과 최신성이 종합목록 DB의 상대적 유용성을 평가한다면, 중복성, 일관성, 완전성, 정확성은 종합목록 DB의 절대적 유용성을 평가한다. 그리고 이러한 항목의 평가를 통한 효과로는 먼저, 정확성 측정을 통해 종합목록 DB의 오류 레코드 유형을 파악할 수 있으며, 포괄성, 최신성, 완전성 평가를 통해서는 종합목록 DB의 이용률 저하 원인 및 단위 공공도서관의 종합목록 구축에 대한 책임의식을 엿볼 수 있을 것이다. 그리고 중복성 평가를 통해서는 종합목록 DB가 안고 있는 고질적 문제인 중복레코드의 유형 파악이 가능할 것으로 기대한다. DB 품질평가에 사용된 표본레코드는 연구팀이 개발한 탐색질의를 사용하여 검색된 단행본 레코드 326건을 이용하였다.

6개의 평가기준을 통해 측정한 KOLIS-NET 종합목록 DB의 품질은 전반적으로 매우 열악하였다. 포괄성을 비롯한 모든 평가항목에서 종합목록 DB의 품질은 심각한 결함을 드러내었다. 단위 공공도서관에서 구축된 DB의 평균 포괄성은 73.4%로 매우 낮게 나타났으며, 단위 도서관에서 신규목록 구축시 종합목록 DB의 유용성을 평가할 수 있은 최신성 역시 36.2%로 매우 낮게 나타났다. 다수의 구축주체에 의해 생성되는 종합목록 DB의 특성상 불가피한 문제지만 중복률 또한 44.2%로 상상을 초월하였다. 이러한 측정결과는 단위 공공도서관의 편목담당사서들에게 KOLIS-NET이 더 이상 유용한 서지도구가 될 수 없음을 나타낸다. 그리고 데이터 기술과 표현에 있어서의 일관성, 완전성, 정확성 결여는 중복레코드 생성의 주요 원인으로 작용하고 있음에도 불구하고 품질평가에서 필드의 누락, 특정 필드의 데이터 누락, 필드 사용의 오류, 중복 사용된 필드간의 불일치 등으로 인한 오류레코드의 생성이 심각하였다. 이러한 DB의 열악한 품질문제는 KOLIS-NET이 더 이상 이용자에게 유용하지 않을 뿐만 아니라 단위 도서관의 편목담당사서들에게 유용한 서지도구가 될 수 없음을 의미한다. 결국 DB의 유용성 결여는 이용에 있어서 외면으로 이어진다.

3장에서 각각의 품질평가 항목에서 지적한 문제점과 개선방안을 종합하여 KOLIS-NET 종합목록 DB의 품질향상을 위한 개선방안은 제시하면 〈표 9〉와 같이 요약될 수 있다. 즉, 종합목록 DB의 품질개선 방안은 크게 두 가지 차원에서 추진될 수 있다. 즉 종합목록 DB의 운영주체인 국립중앙도서관의 역할과 구축주체면서 동시에 이용자인 단위 공공도서관의 역할로 구분할 수 있다. 왜냐하면 결국 KOLIS-NET은

이들 두 주체에 의해 운영되고 구축되기 때문에 당연히 문제의 해결점도 이들의 노력과 협력에 의해서만 가능하다. 개선방안을 요약하면, 운영주체인 국립중앙도서관은 일괄업로드 방식의 개선, 사후 품질관리의 강화, 데이터 입력지침의 마련 및 오류처리 알고리즘의 보완을 위해 제도적 장치와 정책적 지원을 강구해야한다. 한편 운영주체이면서 구축주체인 단위 공공도서관은 구축주체로서의 자긍심과 책임감을 갖고 자관의 목록 수준 제고와 함께 명실상부한 국가자료 종합목록 DB 구축을 위해 적극적으로 참여해야한다.

지금까지 연구팀은 품질평가 결과에서 나타난 문제점을 중심으로 개선방안을 제시했다. 그러나 KOLIS-NET 종합목록 DB의 품질이 조악한 데는 여러 가지 요인이 있을 것이다. 보다 근본적인 품질개선 방안 마련을 위해 품질저하의 원인을 명확히 규명하는 작업이 필요하다. 이를 위해서는 종합목록 DB 구축에 참여하고 있는 공공도서관을 대상으로 한 목록 구축 현황에 대한 세밀한 조사와 연구가 절실히 요구된다. 왜냐하면 건강한 종합목록 DB 구축이 결국 이들 공공도서관의 적극적인 참여와 목록 수준의 제고에 달려있기 때문이다. 현상에 대한 정확한 진단을 통해 국립중앙도서관은 미시적 관점에서가 아닌 거시적 관점에서 국가자료 종합목록 시스템 구축을 위한 정책적, 실무적 방안을 강구하는 것이 바람직하다.

〈표 9〉 KOLIS-NET DB의 품질평가 항목 및 개선방안

품질 평가 항목	개선 방안
포괄성	• 일괄업로드 방식의 개선 • 실무자의 참여의식 강화
중복성	• 사후 품질관리 강화 • 참여도서관의 목록 품질수준 제고 • 일괄 업로드 방식의 개선 • 데이터 입력지침 마련
최신성	• 일괄 업로드 방식의 개선 • 실무자의 참여의식 강화
정확성	• 사후 품질관리 강화 • 데이터 입력지침 마련 • 참여도서관의 목록 품질수준 제고 • 오류처리 알고리즘의 보완
일관성	• 사후 품질관리 강화 • 데이터 입력지침 마련 • 참여도서관의 목록 품질수준 제고 • 오류처리 알고리즘의 보완
완전성	• 사후 품질관리 강화 • 데이터 입력지침 마련 • 참여도서관의 목록 품질수준 제고 • 오류처리 알고리즘의 보완

참 고 문 헌

국립중앙도서관, 한국도서관협회. 2005. 『개관 60주년기념 국제심포지엄 자료집: 21세기의 목록과 국가서지정보정책』. 서울: 국립중앙도서관.

국립중앙도서관. 1983. 『한군문헌자동화목록형식(KORMARC): 단행본용』. 서울: 국립중앙도서관.

김성희. 2003. 데이터베이스 평가모델 개발을 위한 해외사례조사연구. 『국회도서관보』, 40(2): 15-33.

노지현. 2003. 유용성의 관점에서 본 도서관목록의 품질. 『한국문헌정보학회지』, 37(2): 107-134

안계성, 함정훈. 2000. 『데이터베이스 품질평가 항목』. 서울: 한국데이터베이스진흥센터.

에바 베로나. 2005. 『저록작성원칙규범』. 김태수, 이재선 역. 서울: 문헌정보처리연구회.

윤정옥. 2003. 연속간행물 종합목록 데이터베이스의 레코드 품질평가. 『한국문헌정보학회지』, 37(1): 27-42.

이만수, 이상열, 신인수. 2003. 서지 데이터베이스의 품질평가에 관한 연구. 『정보관리연구』, 34(3): 21-39.

이용효, 이제환, 노지현, 박홍석. 2001. 『KERIS 서지 DB의 품질관리를 위한 평가모델 개발 및 개선방안 수립』. 서울: 한국교육학술정보원.

이응봉, 조현양, 류범종, 최재황. 2001. 과학기술분야 데이터베이스의 품질향상을 위한 품질 평가 연구. 『한국문헌정보학회지』, 35(2): 109-132.

조순영, 임성희, 임석종, 한혜영, 박연희. 2002. 『해외학술DB 선정 및 평가모델 개발에 관한 연구』. 서울: 한국교육학술정보원.

조순영. 2002. 종합목록 데이터의 오류 유형에 관한 연구: KERIS 종합목록의 학위논문 서지데이터를 중심으로. 『한국문헌정보학회지』, 36(4): 5-19.

조순영. 2002. 『종합목록의 중복레코드 검증을 위한 알고리즘 연구』. 박사학위논문. 성균관대학교 대학원 문헌정보학과.

최석두, 남영준. 1997. 『대학도서관 분담목록시스템의 구축 및 운영방안에 관한 연구』. 서울: 첨단학술정보센터.

한국교육학술정보원. 2005. 『경영・경제 분야 해외전자정보 비교 분석』. 서울: 한국교육학술정보원.

한국도서관협회. 2003. 『한국목록규칙 제4판』. 서울: 한국도서관협회.

한국전산감리원. 2004. 『국립중앙도서관 정보시스템 컨설팅보고서』. 서울: 한국전산감리원.

Barallier, Catherine. 2004. *Le dedoublonnage des notices bibliographiques : pour une mise en oeuvre du control qualité du catalogue commun de la bibliothèque de l'INHA*. Lyon: ENSSIB.

Bolin, Mary K. 2004. "Cataloging design, catalog maintenance, catalog governance." *Library Collections, Acquisitions & Tech-*

nical Services, 24: 80-85.

Brodie, M. 1980. "Data Quality in Information Systems." *Information and Management*, 3: 245-258.

Daniel, E. 1993. "Quality Control of Documents." *Library Trends*, 41(4): 64-66

Falconnet, Veroniaue. 2005. *Les avancées du Catalogue Collectif de France : le futur portail*. Rennes: Maison du Champs de Mars.

Fritz, Deborah A. 2004. *Cataloging with AACR2 & MARC21: For Books, Electronics Resources, Sound Recordings, Videorecordings and Serials*. 2rd ed. Chicago: ALA Editions.

Intner, Sheila S. and Weihs, Jean. 2001. *Standard Cataloging for School and Public Libraries*. 3rd ed. Englewood: Libraries Unlimited.

KCI

KCI