국내 기관 리포지터리 콘텐츠 현황분석 및 개선방안에 관한 연구

A Study on the Contents of Institutional Repositories in Korea

백종명, 한국과학기술정보연구원(KISTI), jm100@kisti.re.kr
이은지, 한국과학기술정보연구원(KISTI), ejlee@kisti.re.kr
Jong-Myung, Paek, Korea Institute of Science and Technology Information
Eun Jee, Lee, Korea Institute of Science and Technology Information

본 연구는 국내 D-Space 기반의 11개 리포지터리에 구축된 콘텐츠 현황을 두가지 관점에서 조사하고 분석하였다. 각 리포지터리에 구축된 콘텐츠의 더블린코어(DC) Type 유형을 추출하여 그 현황을 조사하였으며, 2,500여 콘텐츠들을 무작위 샘플링하여 원문 구축현황 및 유형을 살펴보았다. 그 결과 각 리포지터리에는 표준화되지 않은 다양한 Type 유형을 자체적으로 사용하고 있었으며, 원문을 제공하는 방식 또한 다양하게 나타났다. 이에 본 연구에서는 각 기관별로 사용하는 39가지의 Type 유형을 9개 유형으로 병합한 기준을 제시하고, 향후 리포지터리 콘텐츠를 수집하는 측면에서 품질관리 정책 수립을 위한 제언과 시스템 구축을 위한 워크플로우를 제시하였다.

1. 서 론

최근 연구기관에서 구축하고 있는 기관리포지터리(Institutional Repository, IR)는 이제는 초기 단계를 지나 점진적으로 활성화 되어 가는 중기 단계로 들어서고 있으며, 각 기관에서는 자체적으로 리포지터리를 구축하여 외부에 자관의 콘텐츠가 확산되도록 유도하고 있다. KISTI와 문화체육관광부는 OAK (Open Access Korea) 사업을 수행하면서 국가 차원의 D-Space 기반 IR 시스템을 개발하였고, 이를 OAK Repository라고 명명하여 여러 기관들에 보급하였다. 또한 보급한 기관들을 통해 구축된 콘텐츠들을 통합검색할 수 있는 OAK Portal 서비스를 함께 제공하고 있다.

현재 리포지터리 구축의 문제점중 하나는 기관 리포지터리에 구축되고 있는 콘텐츠 및 메타데이터 유형의 다양성과 그 품질문제이다. 지금까지의 국내 리포지터리에 관한 연구는 이용자 관점에서의 연구 및 운영방안, 평가지표 개발 등을 중심으로 수행된 연구가 대부분이며, 리포지터리에 구축되는 메타데이터를 포함한 콘텐츠에 관한 연구는 거의 이루어지지 않고 있다.

이에 본 연구에서는 국내 리포지터리들의 콘텐츠 현황 및 문제점을 파악하여 향후 학술커뮤니케이션 환경에서의 콘텐츠 품질관리의 개선방안을 제시하고자 하였다.

2. 국내 기관리포지터리 현황

국내에서는 서울대와 KAIST, 인하대 등 일부 주요 대학이 자체 예산을 투입해 학내 교수 및 연구자들의 연구성과물을 관리, 보존하거나 또는 국제적 유통을 위한 리포지터리를 개발하여 운영하고 있다. 또한 한국과학기술정보연구원(KISTI)이 문화체육관광부 도서관정보정책기획단의 예산지원으로 2009년부터 추진하고 있는 '국가도서관지식콘텐츠의 창조적 관리 및 확산'사업은 국가차원에서 국가 지식정보 공동

활용 및 선진화된 유통체제 구축을 목적으로 오픈액세스 전반에 대한 기반 시스템 및 사회문화적 기반을 마련하고 있는 것으로, 우리나라의 대표적인 리포지터리 보급 사업이다. 또한 한국교육학술정보원(KERIS)에서 운영하는 dColleciton 사업은 대학의 학술정보 생성 및 유통체제 구축을 목적으로 전국 대학도서관에 리포지터리를 개발 보급하는 사업으로, 학내 연구자들이 생산하는 다양한 지식콘텐츠보다는 학위논문 중심의 리포지터리로 운영되고 있다(한국. 한국교육학술정보원 2012). 대학의학학술지편집인협의회는 의학학술정보유통체제 구축을 목적으로 우리나라의 의학분야 오픈액세스를 선도하고 있다. 2012년 2월 현재, KoreaMed를 통해 국내 의학분야 학술지 188종의 검색서비스와 Synapse를 통해 107종의 의학분야 학술지 아카이빙을 제공하고 있다(대한의학학술지편집인협의회 2012).

국내 오픈액세스 지식콘텐츠의 통합검색서비스를 제공하는 대표적인 서비스로는 OAK Portal을 들수 있다. 이는 국내 기관리포지터리에 구축된 국내 오픈액세스 지식정보의 메타데이터를 수집하여 조직화하고 통합검색서비스를 제공함으로써 이용자들이 보다 편리하고 쉽게 오픈액세스 자료를 이용할 수 있도록 구성되어 있다

3. 연구방법

연구에서는 국내에서 서비스되고 있는 리포지터리 11개 기관의 자료유형 및 원문 구축유형을 조사하였다. 이 중 3개 리포지터리는 기관 자체적으로 구축한 리포지터리이며, 8개 기관은 문화체육관광부와 KISTI가 공동으로 추진하고 있는 Open Access Korea(OAK) 사업으로 구축된 리포지터리이다.

각 리포지터리에서 수집되고 있는 자료유형을 파악하기 위하여 OAK Portal로 수집 중인 26만여건 데이터의 더블린코어(DC)의 “type" 유형을 조사하였으며, Type 유형에 따른 자료유형의 재분류를 진행하였다.

또한 각 리포지터리에 구축된 콘텐츠의 원문 보유 현황을 알아보기 위해 무작위 샘플링을 통해 각 리포지터리에서 원문구축의 다양한 유형들을 조사하였다.

이를 통해 본 연구에서는 리포지터리에 구축된 콘텐츠를 기관에서 활용할 수 있는 콘텐츠 품질관리 전략을 도출하고 이를 통해 실제 OAK Portal의 콘텐츠 품질관리 시스템 구축에 적용하고자 하였다.

4. 콘텐츠 현황 분석

4.1 자료유형 분석 결과

11개 리포지터리에 구축된 콘텐츠를 분석한 결과, 총 콘텐츠의 수는 260,088건을 대상으로 조사가 이루어졌다. 각 콘텐츠들의 Type 유형을 분석한 결과 <표 1>에서 보여지는 바와 같이 KOASAS와 S-Space가 21개의 유형을 사용하고 있었으며, 인하대학교가 15개의 유형을 사용하고 있는 것으로 조사되었다. 반면 고려대학교 화학공학전문정보센터의 경우, 많은 양의 콘텐츠에 비해 단 3개의 Type 유형만을 제공하고 있는 것으로 나타났다.

이렇게 각 유형별 콘텐츠 Type을 모두 살펴보면 11개 리포지터리에서는 <표 2>와 같이 39개의 Type 유형을 사용하고 있다. 가장 많이 수집된 Type 유형 순으로 보면, 'Article'이 7개 기관에서 사용되는 것으로 나타나 가장 많은 기관에서 사용하고 있는 Type 유형으로 조사되었으며, 'Book'이 6개 기관, ‘Thesis'와 ’other'가 5개 기관 순으로 사용 중인 것으로 나타났다. 반면 단지 1개의 기관에서만 사용 중인 Type 유형도 28개로, 현재 사용되는 유형의 2/3 이상을 차지하고

<표 1> 기관별 보유하고 있는 DC 내 Type 유형

기관	DC "Type" 유형	유형수	건수
KAIST	Other, Book chapter, Article, Thesis, Image, Working Paper, Book, Presentation, 학위논문(박사) -, 학위논문(석사) -, 학위논문(석사)-, 학위논문 (석사) -, 학위논문(박사)-, 학위논문(석사) -, 학위논문(석사) -, 학위논문 (박사) -, 학위논문(석사) :, 학위논문(석사) - b한국과학기술원 :, 학위논문 -, 학위논문(석사):, 학위논문(석박사통합과정) -, (없음)*	21	49,083
고려대학교화학공학센터	Article, Proceeing, (없음)	3	79,590
서울대 S-Space	Other, Conference Paper, Book chapter, Preprint, Article, Thesis, SNU Journal, Dissertation, Video, Animation, Working Paper, Book, Presentation, Technical Report, Biographical-Item, Review, Article; Proceedings Paper, Editorial Material, Correction, Meeting Abstract, (없음)	21	63,223
아주대의학문헌정보센터	Other, Article, Thesis, Video, Book, (없음)	6	4,032
인하대학교	Other, Article, Masters Thesis, PhD Thesis, Thesis, Image, Working Paper, Book, Presentation, (없음)	11	22,878
한국과학기술원	Article, Patent, (없음)	3	16,641
한국보건사회연구원	Other, Article, Book, Technical Report, Report, Dataset, (없음)	7	5,459
한국생명공학연구원	Article, Book, Patent, (없음)	4	15,753
한국소비자원	웹진, news, 사례, 논문, (없음)	5	1,352
한국전기연구원	기술보고서(TR), 연구보고서(RR), (없음)	3	1,234
한국표준과학원	특허, (없음)	2	1,443
합계			260,688

* (없음) 의 경우 Type 유형이 지정되지 않은 콘텐츠임

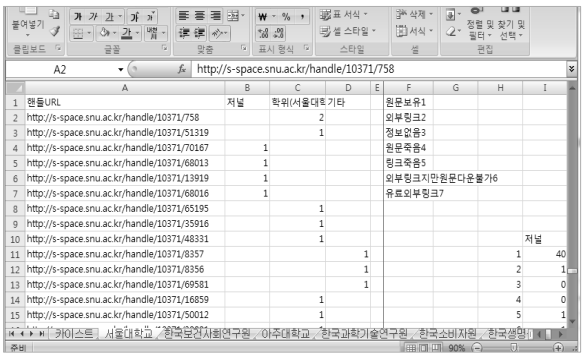
있었다. 자관에서만 쓰이는 독특한 Type 유형을 사용할 경우 수집하는 쪽에서의 자료분류 및 병합에 어려움을 겪고 이는 곧 서비스 및 콘텐츠의 품질에 영향을 미칠 수 밖에 없을 것이라고 분석된다.

<표 2> Type 유형별 보유 기관 순위

순위	Type 유형	기관수
1	Article	7
2	Book	6
3	Other	5
3	Thesis	5
5	Presentation	3
5	Working Paper	3
7	Book chapter	2
7	Image	2
7	Patent	2
7	Technical Report	2
7	Video	2
12	Animation 외 27개 Type	1

4.2 원문유형 분석 결과

본 조사에서는 11개 기관에서 구축된 리포지터리의 콘텐츠 중 <그림 1>과 같이 핸들 URL을 중심으로 무작위 샘플링을 통해 2,540건의 메타데이터를 추출하였고, 각각의 메타데이터의 원문 유무에 대해 직접 리포지터리에 방문하여 그 자료 유형을 분류하였다.



<그림 1> 원문유형 조사를 위한 샘플링 예시

이를 통해 나타난 리포지터리의 원문구축 유형은 아래 <표 3>과 같이 총 7개 유형으로 원문을 제공하고 있는 것으로 나타났다.

<표 3> 리포지터리 콘텐츠 원문제공 유형

구분	유형	상세내용
A	원문보유	· 리포지토리 페이지 상에서 원문을 직접 다운로드 받을 수 있음
B	외부링크	· 원문을 직접 제공하지는 않지만 원문 다운로드가 가능한 페이지 주소를 링크
C	원문정보 없음	· 원문 다운로드도 불가하며, 관련링크 정보도 찾아볼 수 없음
D	원문정보 오류	· 원문 다운로드 버튼을 제공하지만 오류 등으로 인해 다운로드 되지 않음
E	dead link	· 외부링크 정보를 제공하고 있지만, 오류 등으로 인해 링크가 작동하지 않음
F	원문없는 링크	· 외부링크 정보를 제공하고 있지만, 그 링크에서도 원문정보를 찾아볼 수 없음
G	접근불가 링크	· 외부링크 정보를 제공하고 있지만, 실질적으로 원문을 열람하려면 결재나 로그인 등의 절차를 거쳐야 함

유형 A의 경우, 가장 이상적인 리포지터리 콘텐츠 유형이라고 할 수 있다. 이는 이용자가 타 시스템으로 링크를 따라가지고 않고 바로 리포지터리에서 콘텐츠를 다운로드 받을 수 있도록 하는 유형이다. 유형 B의 경우 PDF 파일 등의 원문을 직접 다운로드 받을 수는 없지만, 타 시스템(가령 도서관)으로의 원문링크를 제공하는 유형이다. 유형 C는 원문은 누락된 채 메타데이터만 제공하는 경우이다. 이는 단지 담당자의 실수가 아닌 학술논문의 저작권 문제 등으로 인하여 외부로 노출하지 못하는 경우 등을 비롯한 다양한 환경에서 발생된다. 유형 D, E, F의 경우, 시스템 오류, Dead Link 등의 오류로서 수정을 요하는 경우이다. 유형 G는 원문의 외부링크 정보를 제공하고 있지만 실질적으로 원문을 열람하기 위해서 결재나 로그인 등의 절차를 거쳐야 하는 콘텐츠로서 Free Access 혹은 Open Access로서의 가치를 상실한 자료라고 할 수 있다.

5. 결론 및 제언

본 연구에서는 국내 리포지터리의 콘텐츠 분석을 두 가지 관점에서 조사하였다.

첫째, 자료유형의 관점에서 각 리포지터리에 구축된 Type 유형을 조사함으로써 11개 기관에서 사용하는 39개의 Type 유형을 추출하여 병합하고 재분류 하였으며, <표 4>와 같이 총 9개의 자료유형을 제안한다.

<표 4> Type 병합에 따른 자료유형 재분류

자료 유형	한글 표현명	리포지터리 자료유형
Article	논문 (9)	Article, Article; Proceedings Paper, Conference Paper, Correction, Dissertation,Preprint, Proceeing, SNU Journal, Editorial Material
Thesis	학위 논문 (10)	Masters Thesis, PhD Thesis, Thesis, 논문, 학위논문 -, 학위논문 (박사) -, 학위논문 (석사)-, 학위논문(석박사통합과정) -, 학위논문(석사) - b한국과학기술원 :
Book	단행본(2)	Book,Bookchapter,
Dataset	데이터셋(2)	Biographical-Item,Dataset
Patent	특허(2)	Patent,특허
Report	보고서 (5)	Report, Technical Report, 기술보고서(TR), 연구보고서(RR), Working Paper
Multimedia	멀티미디어(3)	Animation,Image,Video,
Other	기타(6)	Meeting Abstract, news, Other, Review, 사례, 웹진
Presentation	프리젠테이션(1)	Presentation

이렇게 병합된 자료유형 기준을 제안함으로써 OAK Portal과 같은 통합검색서비스에서 수집한 자료들의 서비스를 보다 원활하게 할 수 있을 것으로 생각되나 기 구축된 자료유형 Type을 병합하는 것에는 한계가 존재할 수밖에 없다. 향후 리포지터리는 꾸준히 그 숫자가 늘어날 것으로 예상되기 때문에 리포지터리 콘텐츠 구축 시 한정된 Type 유형을 사용하고 관련 가이드라인

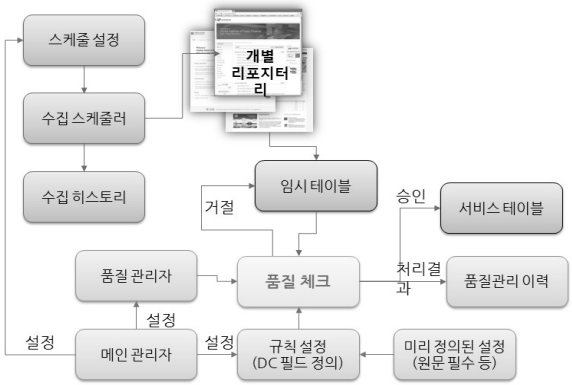
을 수립 및 제공하는 것이 무엇보다 중요하다.

두 번째 관점에서는 각 리포지터리 콘텐츠의 원문 구축유형을 샘플링을 통해서 분석하였다. 오픈액세스의 기본 사상에 따르면, 리포지터리는 기관의 소속 이용자가 자신의 연구물을 리포지터리에 셀프아카이빙으로 구축하고, 등록된 자료를 무료로 이용 가능하도록 제공하는 것이다. 분석결과 리포지터리에 PDF 등 원문 자체가 메타데이터와 함께 등록된 경우가 가장 이상적인 아카이빙 방법이나 기관에 따라 다양한 형태로 원문을 제공하고 있는 것으로 조사되었다. 이는 여러 원인에 의해 발생되었지만, 레거시시스템과의 중복 아카이빙 문제, 저작권 등의 문제가 가장 크게 작용하는 것으로 파악된다. 또한 원문 링크 시 셀프아카이빙보다는 리포지터리 관리자(혹은 도서관 관리자)를 통해 일괄적인 대량의 자료 등록이 이루어지기 때문에 자료 등록자의 원문링크에 대한 확인이 필요하다.

OAK Portal과 같이 여러 리포지터리 콘텐츠를 수집하여 서비스하는 시스템은 자체적으로 콘텐츠 품질관리를 수행하여 양질의 콘텐츠만을 선별적으로 서비스하는 정책이 필요하다. 특히 리포지터리와 같이 비정형화된 다양한 자료를 수집하는 경우 이러한 자체적인 품질관리는 더욱 요구된다.

이러한 결과를 바탕으로 리포지터리 수집 시스템에서의 품질관리 워크플로우를 제시하면 <그림 2>와 같다.

제시된 워크플로우는 각 리포지터리 콘텐츠 수집 시 서비스테이블로 넘어가기 전 임시테이블에서 콘텐츠 품질체크를 수행한 후 관리자 승인 서비스테이블로 넘어가는 과정을 나타낸다. 이때 관리자는 설정된 DC 필드의 규칙들을 정의하고, 재분류된 자료유형으로 병합한 후, 원문 유형을 파악하여(자동 혹은 수동으로) 양질의 콘텐츠만을 서비스테이블로 넘기는 것이 중요하다.



<그림 2> 수집 콘텐츠 관리를 위한 워크플로우

지금까지 수행된 연구 결과는 특정 리포지터리를 선정하여 제한된 콘텐츠들을 미시적으로 살펴본 것으로써 이를 일반화하기에는 어려운 실정이다. 이후 더 많은 리포지터리 사례를 분석하고 파악하는 것을 통해 리포지터리 콘텐츠의 품질을 높이는 개선 방안을 도출할 수 있으며, 향후 표준화된 리포지터리 관련 메타데이터 개발 및 분석 연구로 확대하여 진행할 필요가 있다.

지속적이고 다양한 조사 · 분석을 통해 표준화된 양질의 콘텐츠들이 축척됨으로써 리포지터리를 통한 오픈액세스 콘텐츠가 국내 학술커뮤니케이션 시장에서의 참여가 확산되기를 기대한다.

참고문헌

정연경, 이나니, 이미화(2006). 기관 레포지터리 활성화를 위한 메타데이터 확장방안연구. 한국문헌정보학회지, 41(1), 323-344

대한의학학술지편집인협의회(2012). KoreaMed. [인용 2013.07.26]. http://koreamed.org/.

한국교육학술정보원. (2012). 디지털학술정보유통시스템(dCollection). [인용 2012.02.20]. http://www.dcollection.net/search/ko/intro02_01.do