

국내 기관 리포지토리에서 DSpace 메타데이터 요소의 활용 현황

The Uses of DSpace Metadata Elements in the Institutional Repository

이 은 주(Eun-Ju Lee)*

〈목 차〉

I. 서론	2. 메타데이터의 입력 현황
II. DSpace 개요	IV. DSpace 검색시스템의 활용 현황
1. DSpace의 개발 배경 및 특징	1. 검색에 활용되는 메타데이터 요소
2. DSpace의 메타데이터 요소	2. 검색결과에 대한 분석
3. DSpace의 요소 확장	V. DSpace 메타데이터 요소의 활용 전략
III. DSpace 메타데이터 요소의 활용 현황	VI. 결론
1. 메타데이터 요소의 설계	

초 록

이 연구는 DSpace를 기반으로 구축한 기관 리포지토리 메타데이터 요소의 활용 현황을 살펴보고, 향후 DSpace를 사용하여 기관 리포지토리를 구축하고자 하는 기관을 위한 메타데이터의 구축 방향과 활용 전략을 제시하는 것을 목적으로 삼았다. 이를 위해, 이 연구에서는 (1) DSpace의 개발 배경과 특징, 메타데이터 요소와 활용 방법에 대해 살펴본 후, (2) DSpace를 기반으로 한 국내 기관 리포지토리 중 사례 기관을 선정하여 모든 기관에서 공통적으로 활용되는 메타데이터 요소는 무엇이며, 개별 기관에서 추가 또는 확장하여 사용하는 요소는 무엇인지를 구분하여 분석하였다. 이에 더해, (3) 입력된 메타데이터가 검색에 어떻게 활용되고 있는지 검색 방식, 검색 요소 및 디스플레이 요소 등으로 나누어 검토한 후, 마지막으로, (4) 연구결과를 토대로 기관 리포지토리에서 메타데이터 구축시 기관이 고려해야 할 과제와 추진전략에 대해 논의하였다. 이러한 논의를 수행함에 있어 필요한 데이터는 문헌연구와 사례분석을 통해 수집하였다.

키워드: 기관 리포지토리, 메타데이터, 검색시스템, 오픈액세스

ABSTRACT

The purpose of this research is to investigate the metadata elements of institutional repository using DSpace and to provide the directions and strategies for constructing metadata. To the end, discussed in details are (1) understanding the background of development, features, elements of metadata, (2) figuring out the common metadata elements and the differentiated metadata elements for the institutional repository, (3) examining the search methods, search elements, and display elements of the institutional repository, and (4) based on theses, suggested are directions and strategies for metadata of institutional repository using DSpace. The research methods were used such as literature review, the analysis of 4 domestic institutional repository.

Keywords: DSpace, Institutional repository, Metadata, Search system, Open access

* 동의대학교 문헌정보학과 조교수(ejulee@deu.ac.kr)

•논문접수: 2016년 11월 20일 •최초심사: 2016년 11월 29일 •개재확정: 2016년 12월 19일

•한국도서관정보학회지 47(4), 465-487, 2016. [<http://dx.doi.org/10.16981/kliss.47.201612.465>]

I. 서론

학술정보를 둘러싼 생산 환경이 변화하고 있다. 대학에서 진행되는 강의를 온라인상에서 공유하는 서비스가 개발되면서 기존에 없었던 새로운 유형의 정보자원이 등장하는가 하면, 다양한 디지털 프로젝트의 수행으로 인해 디지털자원이 폭발적으로 증가하는 등, 정보자원의 생산 양식이 혁신적으로 달라지고 있다. 이러한 시대적 변화에 따라 이용자들은 전통적인 정보자원인 학술논문, 학위논문뿐만 아니라 강의자료, 리포트, 데이터셋, 학생발표자료, 멀티미디어 콘텐트에 이르기까지 광범위한 자원을 필요로 하고 있다. 그러나 이용자들에게 학술정보를 제공해야 할 도서관은 예산감소라는 차가운 현실에 직면해 있으며, 학술DB의 가격인상으로 자원수집에 더욱 어려움을 겪고 있다. 그 결과, 도서관과 사서는 위기의 상황 속에서 보다 다양한 유형의 자원을 수집하고 보존해야 하는 새로운 역할에 대해 고민하기 시작하였다.

때마침 1990년대 초반에 등장한 새로운 학술커뮤니케이션 모델인 ‘저자 셀프 아카이빙 서비스’(author self-archiving services)는 도서관의 새로운 역할을 정립하는데 단초를 제공하고 있다. 본인의 연구 성과물이 널리 읽히기 바라는 연구자들의 희망과 이를 가능케 하는 기술의 진보는 셀프 아카이빙 서비스를 창출하였고, 이는 도서관이 직면하고 있는 어려움을 해결하기 위한 방안 중 하나인 기관 리포지토리(institutional repository) 구축의 밑거름이 되었다(Van de Sompel and Lagoze 2002, 147). 즉, 저자 스스로 자신의 연구 결과물을 등록하는 셀프 아카이빙 서비스는 교수나 학생에 의해 생산된 디지털 형태의 학술 자료를 개별 도서관이나 연구소가 새로운 방식으로 수집·관리·보존·서비스하는 기관 리포지토리의 구축에 동력이 된 것이다. 이렇게 구축된 기관 리포지토리는 결과적으로 다양한 유형의 정보자원을 확보하기 원하는 대학 및 연구소의 요구에 부응하는 디딤돌이 되었다.

그런데 국내의 기관 리포지토리 구축 과정을 살펴보면 다른 국가와 조금 다른 양상을 보인다. 가령, 미국과 유럽에서는 대학이나 연구소가 자발적으로 기관 리포지토리를 구축한 것과 달리, 국내에서는 정부 차원에서 수행한 특정 사업의 결과물로 구축되고 있다. 물론 초기에 구축된 리포지토리의 경우 자발적으로 개발한 경우도 존재하지만, 대부분의 국내 기관 리포지토리는 정부에서 먼저 사업의 추진 기관을 선정하고 시스템을 개발한 뒤 이를 회원 기관에 보급하는 방식으로 구축되었다. 그 결과, 현재 국내에서 구축된 기관 리포지토리의 대부분은 KERIS의 dCollection 혹은 국립중앙도서관의 OAK 사업과 연계되어 있다고 볼 수 있다. 따라서 국내의 기관 리포지토리에 대한 충분한 이해를 위해서는 두 중심축에 대한 면밀하고도 고른 이해가 수반되어야 한다.

그러나 기관 리포지토리에 대한 국내 연구는 주로 KERIS의 dCollection에 한정되어 있을 뿐, OAK 리포지토리와 관련한 연구는 상대적으로 적은 편이다. 구체적으로 살펴보면 KERIS

의 dCollection의 경우, 대학 리포지터리 개발을 위한 다양한 전략을 모색하는 거시적인 관점에서부터 콘텐트 유형 및 원문제공 유형을 분석하고, 메타데이터 요소의 확장을 고찰해 보는 등의 미시적인 관점에 이르기까지 다양한 연구들이 진행되어 왔다(장금연 2004 ; 김현희, 정경희, 김용호 2006 ; 정연경, 이나니, 이미화 2007 ; 백종명, 이은지 2013 등). 이에 비해, OAK 리포지토리에 관한 연구는 구축 사례를 소개하는 수준에 불과하며, OAK 리포지토리의 운영 소프트웨어인 DSpace에 대한 논의조차 개괄적인 안내 수준에 그치고 있다. 특히, 리포지토리의 메타데이터에 관한 연구는 dCollection의 SOMS에 관한 연구(정연경 등, 2007)가 거의 유일하며, OAK의 DSpace에 대해서는 아직 다뤄지지 않고 있다.

이에, 이 연구에서는 기존 연구에서 다뤄지지 않았던 DSpace를 중심으로 메타데이터 요소의 분석 및 활용에 대해 논의해보고자 한다. 구체적으로, DSpace의 메타데이터에 대한 기본적인 이해를 바탕으로, DSpace를 사용하는 기관 리포지토리의 메타데이터 요소를 면밀하게 비교분석하고, 나아가 기관 리포지토리에서 구축한 메타데이터가 검색에 어떻게 활용되고 있는지를 살펴봄으로써 향후 DSpace를 활용하여 메타데이터를 구축하고자 할 때 고려해야 할 사항에 대해서 논의해 보고자 한다. 이처럼 이 연구는 DSpace를 기반으로 한 기관 리포지토리 메타데이터의 구축 방향을 제안하고, 구축을 위한 실행전략을 제시하는 것을 목적으로 삼고 있다.

이러한 목적을 위해 이 연구에서는 (1) DSpace의 개발 배경과 특징, 메타데이터 요소와 활용 방법에 대한 충분한 이해를 시도하였다. 이어, (2) DSpace를 기반으로 한 국내 기관 리포지토리 중 사례 기관을 선정하여 모든 기관에서 공통적으로 활용되는 메타데이터 요소는 무엇이며, 개별 기관에서 추가 또는 확장하여 사용하는 요소는 무엇인지를 구분하여 분석함으로써 실제 메타데이터 요소가 어떻게 활용되고 있는지 주의 깊게 살펴보았다. 이에 더해, (3) 입력된 메타데이터가 검색에 어떻게 활용되고 있는지 검색 방식, 검색 요소 및 디스플레이 요소 등으로 나누어 검토한 후, 마지막으로, (4) 이상의 연구결과를 토대로 기관 리포지토리에서 메타데이터 구축시 기관에서 고려해야 할 과제와 추진전략에 대해 논의하였다. 이러한 논의를 수행함에 있어 필요한 데이터는 문헌연구와 사례분석을 통해 수집하였다.

II. DSpace 개요

1. DSpace의 개발 배경 및 특징

기관 리포지토리의 개념과 존재 목적을 정립할 때 가장 많이 인용되는 키워드는 ‘기관’, ‘학술자료’, ‘보존’, 그리고 ‘상호운용성’이다. 즉, 기관 리포지토리는 주최 ‘기관’의 지적 산출물을 대상으로 하며, 콘텐트의 내용은 ‘학술’적이거나 교육, 연구 등에 해당되어야 하며, 연구자가

제출한 자료는 기관 리포지토리 콘텐트로서 장기간 ‘보존’해야 할 의무를 지니며, 마지막으로, 기관의 지적 산출물을 배포하기 위해 ‘상호운용성’을 갖추고 오픈액세스를 제공해야 한다 (Dora and Maharana 2008, 250). 이를 정리하면 기관 리포지토리의 일차적 목적은 기관 소속의 연구자들이 생산한 여러 형태의 학술자원을 수집하여 관리하고 장기보존하는데 있다. 이러한 목적은 “기관 리포지토리의 핵심은 기관 구성원들이 생산해 낸 자원을 디지털 형태로 보존하는 장기보존정책”이라는 Lynch의 지적과도 맥락을 같이 한다(Lynch 2003, 227). 이에 더해 기관 리포지토리는 수집한 자원을 오픈액세스 정신에 입각하여 타 기관과 공유하고 연구 결과물을 널리 확산시켜 학술커뮤니케이션을 증진시키고자 하는 거시적인 목적도 동시에 가지고 있다.

기관 리포지토리는 기관 리포지토리용 소프트웨어를 활용하여 구축되는데, 오픈 리포지토리 리스트를 제공하는 ROAR과 OpenDOAR로부터 받은 데이터를 종합하여 구축한 Repository map에 따르면 2014년 10월 현재 전 세계적으로 구축된 3,045개의 기관 리포지토리 중 DSpace를 활용한 기관이 1,225개, EPrints가 468개, BEPress가 168개, ETD-db가 44개의 순으로 나타나 기관 리포지토리용 소프트웨어 중 DSpace가 가장 많이 활용되고 있음을 알 수 있다.

DSpace는 미국의 MIT(Massachusetts Institute of Technology) 대학을 주축으로 개발되어 왔는데, 개발 배경을 살펴보면 전자자원의 생산과 유통구조의 변화, 그리고 그에 대한 이용자들의 요구가 시발점이었음을 알 수 있다. 먼저, MIT 소속 연구자들은 자신의 연구결과를 다양하고도 복잡한 형식으로 출판하게 되면서 해당 자료들을 수집, 관리, 보존할 필요성을 느끼게 되었다. 이에 도서관에서 보다 다양한 유형의 학술자원을 수집, 관리해달라는 요구가 빛발쳤고, 그 결과, MIT 도서관에서는 이러한 요구를 수용하기 위한 방안을 적극적으로 모색하였다. 이러한 과정에서 MIT 도서관에서는 2000년부터 정보자원을 전문적으로 관리하는 기관 리포지토리를 구현하기 위해 HP연구소(Hewlett-Packard)와 공동으로 DSpace의 개발에 착수하기 시작하였고 2002년 11월에 개발을 완료하였다.

DSpace가 세상에 알려지기 한 달 전 DSpace를 활용한 기관 리포지토리는 MIT 도서관의 새로운 서비스로 먼저 소개되었다. DSpace는 학술논문, 학위논문, 각종 이미지, 오디오 파일 등 다양한 정보자원을 수집, 활용할 수 있는 시스템으로 소개되었으며, 모든 유형의 자원을 관리할 수 있는 가능성을 열었다는 점에서 여러 연구에서 긍정적인 평가를 받았다. 실제, DSpace를 활용한 MIT의 사례를 살펴보면 초기에는 교수들의 연구 성과물만을 대상으로 수집하였으나, 현재는 교수 이외에 연구원, 대학원생 등이 생산한 유용한 자료들 중 저작권 문제가 없는 자료까지 수집 대상으로 삼는 등, 다양한 자원의 유형을 공격적으로 수집하여 활용하고 있다.

또한 DSpace는 소스 코드를 공개하는 오픈 소스 소프트웨어라는 특징으로 인해 전 세계에서 가장 많은 개발자 커뮤니티를 가지고 있다. 이러한 특징은 커스터마이징을 강조하는 DSpace의 기조로 인해 더욱 커다란 강점으로 부각되었고, 다양한 개발자 커뮤니티들의 협업을 통해 더 나은 시스템으로 발전하는 디딤돌이 되었다. 그 결과, DSpace를 활용한 기관 리포지토리 구축 시 기관의 특성과 규모, 자원의 유형과 속성에 따라 여러 기능을 변경할 수 있게 되었으며, 이러한 특성은 메타데이터 요소의 설계나 데이터의 구축에서도 동일한 강점으로 인정받고 있다.

2. DSpace의 메타데이터 요소

DSpace에서 제공하고 있는 메타데이터 요소는 DC(Dublin Core)의 15개 요소를 사용하고 있으며 52개의 한정어를 활용하여 의미를 세분화하고 있다(<표 1> 참조). 이들 요소와 한정어는 기관의 목표와 특성, 자원의 유형과 특징에 따라 수정 및 확장이 가능한데, 가령, 자원의 유형(학술논문, 프리프린트, 보고서, 회의자료, 학위논문, 이미지, 오디오, 비디오, 데이터세트 등)에 따라 메타데이터 요소를 변경하거나 자원의 유형별로 상이한 요소 및 한정어를 설정하여 사용할 수도 있다.

<표 1> DSpace 메타데이터 요소

요소	한정어	요소명 (한글)	요소	한정어	요소명 (한글)
title		표제		medium	원문매체
	alternative	대체표제		mimetype *	MIME 유형
creator		저작자	type		유형
subject		주제키워드	identifier		식별자
	classification	분류번호		citation	표준서지인용기술
ddc	주제(DDC)		govdoc		정부문서번호
lcc	주제(LCC)		isbn		국제표준도서번호
lcsh	주제(LCSH)		issn		국제표준연속간행물번호
mesh	주제(MESH)		sici		SICI
other	기타 주제		ismn		국제표준음악번호
description		설명문(일반)		other	기타번호
	abstract	초록		uri *	식별자
provenance *	관리정보기술		source		정보원
sponsorship	후원		uri		식별자
statementof responsibility	책임사항		language		언어
tableofcontents	목차		iso		언어(ISO)
uri	식별자		relation		관계
publisher	발행자		isformatof		타포맷 아이템
contributor	기여자		ispartof		하위 아이템
	advisor	지도교수	ispartofseries		총서의 하위 아이템
	author	저자	haspart		포함 아이템

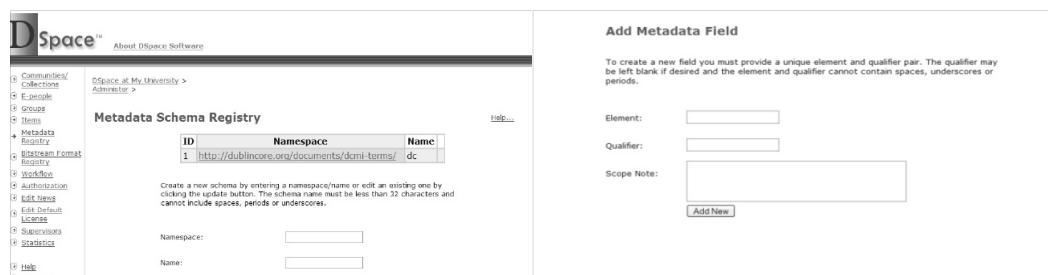
6 한국도서관·정보학회지(제47권 제4호)

요소	한정어	요소명 (한글)	요소	한정어	요소명 (한글)
	editor	편집자		isversionof	이전 버전
	illustrator	삽화가		hasversion	최근 버전
	other	기타		isbasedon	제출원본
date		날짜		isreferencedby	참조정보원
	accessioned *	입수일		requires	필요자원
	available *	이용가능일		replaces	선행자원
	copyright	저작권일		isreplacedby	후속자원
	created	제작일		uri	식별자
	issued *	발행일	coverage	spatial	지역범위
	submitted	제출일		temporal	연대범위
format		파일형식	rights		저작권
	extent *	원문크기		uri	식별자

* 표시는 시스템에서 입력값을 자동 생성하는 요소임

3. DSpace의 요소 확장

DSpace는 메인 홈페이지로부터 필요한 프로그램을 일괄 다운받는 턴키방식(turn-key approach)을 적용하고 있다.¹⁾ 따라서 다른 시스템에 비해서 상대적으로 쉽게 설치할 수 있는 소프트웨어라는 평가를 받고 있다(Castagné 2013, 2). DSpace 설치 후에는 자관이 필요로 하는 메타데이터를 커스터마이징 할 수 있다. 가령, <그림 1>과 같이 새로운 메타데이터 요소를 추가하고자 할 경우 네임 스페이스와 이름을 입력하면 바로 요소가 생성이 되는데, 이 때, 해당 시스템에서 활용할 요소명과 필요하다면 한정어와 주기를 추가할 수 있다.



<그림 1> DSpace 메타데이터 요소 확장 화면

이렇듯 커스터마이징의 용이성은 검색기능에 대해서도 유사하게 적용된다. 즉, 검색기능 설정 화면에서 기존에 설정되어 있는 검색필드 값 하단에 새롭게 추가하고자 하는 검색요소

1) DSpace를 설치하기 위해서는 사전에 자바개발도구(Java Development Kit, JDK)를 설치한 후, 자바 프로그래밍(Apache Ant)을 위한 도구와 자바용 프로젝트 관리 도구(Apache Maven)를 설치해야 한다. 또한 데이터베이스를 생성하기 위한 데이터베이스관리시스템(PostgreSQL)을 설치하고 새 데이터베이스를 생성한 뒤 웹 어플리케이션 서버(Apache Tomcat)를 설치한다. 사전에 필요한 다양한 프로그램을 설치 후, DSpace를 설치할 수 있다(Patel 2011, 5-16).

명(<그림 2>)에서 “search.index.5 =abstract:dc.description.abstract”)을 추가하면 기존에 설정된 검색요소 외에 추가한 요소를 검색에 적용할 수 있다.

이처럼 자관의 상황에 맞추어 환경설정을 변경하는 등의 과정을 거치고 나면 실제 메

타데이터를 입력할 수 있다. 이 때, 절차 및 입력 주체 역시 기관별로 상이하게 운영할 수 있다. 가령, MIT에서는 커뮤니티별로 코디네이터를 두고 독자적인 지침에 따라 기관 리포지토리를 운영하도록 하고 있다. 따라서 커뮤니티 소속의 연구자가 연구 성과물을 제출하면 제출된 자원을 해당 커뮤니티의 코디네이터와 기관 리포지토리 담당자(메타데이터 편집자)가 적절성 여부를 검토한 뒤 승인을 하여야만 비로소 수집이 결정된다(곽승진, 신재민 2008, 74–75). 이와 달리, DSpace를 활용하는 국내의 포항공과대학교 도서관에서는 연구자의 행정 부담을 최소화하기 위해 도서관의 관리 부서에서 직접 자료를 수집하고 메타데이터를 작성한 뒤 업로드까지 전담하고 있다(이진솔 2015, 33). 이처럼 DSpace는 DSpace에서 설정된 메타데이터 요소는 물론이고 검색기능 및 데이터 구축 주체 등에 이르기까지 개별 기관의 목적이나 특성에 맞춰 유연하게 활용할 수 있는 환경을 제공하고 있다.

```
search.index.1 = author:dc.contributor.*  
search.index.2 = author:dc.creator.*  
search.index.3 = title:dc.title.*  
search.index.4 = keyword:dc.subject.*  
search.index.5 = abstract:dc.description.abstract //추가
```

* 요소는 디폴트로 정의된 검색필드임

<그림 2> DSpace 검색요소 추가 화면

III. DSpace 메타데이터 요소의 활용 현황

지금까지 DSpace의 개발 배경 및 특징, 기본적인 메타데이터 요소에 더해 실제 DSpace를 활용하여 기관 리포지토리를 구현하고자 할 때의 메타데이터 확장 방법에 대해 살펴보았다. 지금부터 국내에서 DSpace를 활용하여 자원을 수집·서비스하는 사례들을 대상으로 실제 어떠한 요소를 적용하고 있으며, 특히, 기관별로 어떻게 추가 또는 확장하고 있는지에 대해서 분석해 보고자 한다. 분석 대상은 DSpace를 활용하고 있는 국내 기관 리포지토리의 성격별로 구분(공공기관, 대학도서관, 연구소, 의학도서관)한 후, 각 성격별 리포지토리 중에서 데이터의 구축량이 가장 많은 리포지토리 1개관씩 총 4개관을 선정하였다. 또한 메타데이터 요소 분석은 대표적인 정보자원인 ‘학술논문’에 한정하였다.²⁾

2) 2016년 현재, 37개에 이르는 기관에서 DSpace를 활용하고 있는 것으로 조사되었다. 이들 기관 중 접속이 불가능한 기관과 메타데이터를 공개하지 않는 기관을 제외하면 22개의 기관 리포지토리를 실제 확인할 수 있는데, 22개 기관 리포지토리를 기관의 성격과 수집 자원의 유형에 따라 구분하면 다음 표와 같다.

1. 메타데이터 요소의 설계

학술논문에 대한 메타데이터 요소의 활용 현황을 살펴보기 위해 임의선정한 4개의 사례 기관을 조사한 결과, 기관에 따라 적게는 8개에서부터 많게는 10개의 요소를 사용하고 있는 것으로 나타났다(<표 2> 참조). 요소별로 살펴보면 title(표제), description(설명문), contributor(기여자), date(날짜), type(유형), identifier(식별자) 총 6개의 요소는 4개 기관에서 공통적으로 사용하고 있는 반면, subject(주제), publisher(발행자), format(파일형식), language(언어), relation(관계), rights(저작권) 총 6개의 요소는 기관별로 달리 사용하고 있었다. 결과적으로 학술논문에서 creator(저작자), source(정보원), coverage(범위) 3개 요소는 전혀 활용하지 않는 요소로 조사되었다.

<표 2> 학술논문 기술을 위한 메타데이터 요소 설계

기관 성격	공공기관 A		대학도서관 B		연구소 C		의학도서관 D	
	요소	한정어	요소	한정어	요소	한정어	요소	한정어
DSpace 제공 요소	title		title		title		title	
							title	alternative
		subject		subject				
		description	abstract				description	abstract
	description	provenance	description	provenance	description	provenance	description	provenance
							description	statementof responsibility
	publisher		publisher		publisher			
	contributor	author	contributor	author	contributor	author	contributor	author
	date	accessioned	date	accessioned	date	accessioned	date	accessioned
	date	available	date	available	date	available	date	available

기관 성격	기관 수	수집 자원의 유형
공공기관	3개	학술논문, 보고서, 뉴스, 분쟁조정결정사례, 전국자연환경조사자료
대학도서관	10개	학술논문, 학위논문, 보고서, 회의자료, 특허, 단행본, 연구용역보고서, 정책보고서, 강의자료, 연감, 백서, 뉴스레터, 행사자료집, 세미나자료, 멀티미디어, 홍보자료, 통계, 신문, 사진, 출판앨범, 업무보고자료
연구소	6개	학술논문, 학위논문, 보고서(연구보고서, 정책보고서, 기술보고서 등), 회의자료, 특허, 단행본, 행사자료집, 정책자료, 회의자료, 멀티미디어, 홍보자료
의학도서관	3개	학술논문, 학위논문, 특허, 포스터자료
계	22개	

기관 성격	공공기관 A		대학도서관 B		연구소 C		의학도서관 D	
	요소	한정어	요소	한정어	요소	한정어	요소	한정어
자관 추가 요소 또는 한정어	identifier	other			identifier	other	identifier	other
	identifier	uri	identifier	uri	identifier	uri	identifier	uri
		language					relation	ispartof
							rights	
	rights	uri					rights	uri
			type	rims			contributor	college
			identifier	pid			contributor	department
			identifier	rimsid			contributor	alternative author
			identifier	wosid			contributor	googleauthor
			identifier	scopusid			identifier	controlno
			identifier	doi			relation	volume
		citation	volume				relation	issue
		citation	issue				rights	isfree
		citation	beginning page					
		citation	endingpage					
		citation	publicationname					

한정어 또한 9개부터 23개에 이르기까지 기관에 따른 활용도의 차이가 나타났는데, 요소에 비해 한정어의 경우 그 편차가 상대적으로 큰 것으로 나타났다. 가령, 의학도서관 D에서는 alternative title(대체표제)을 한정어로 구분하여 활용하고 있는 반면, 공공기관 A에서는 alternative title(대체표제)을 title(표제) 요소 아래에 해당 데이터를 함께 기술하며 한정어를 사용하지 않고 있었다. 또한 format(파일형식)의 기술에 있어서도 extent(파일크기)(의학도서관 D)나 mimetype(매체유형)(공공기관 A) 한정어를 추가적으로 적용하는 기관이 있는 반면, 파일형식 자체를 적용하지 않는 기관(대학도서관 B와 연구소 C)도 있는 등, 기관에 따라 적용 요소 또는 요소의 세분화 정도에서 분명한 차이가 존재하였다.

또한 DSpace에서 제공하는 요소와 한정어 이외에 자관에서 자체적으로 추가 및 확장한 메타데이터 요소에서도 기관에 따른 편차가 역력하게 드러났다. 가령, 공공기관 A와 연구소 C에서는 요소를 전혀 확장하지 않고 있는 반면 대학도서관 B와 의학도서관 D에서는 다양한 요소를 확장하여 활용하고 있었다. 특히, 대학도서관 B에서 요소의 확장이 두드러지게 나타났는데, <표 2>의 하단에 제시되어 있듯이 기존 요소인 type(유형)과 identifier(식별자) 하위에 새로운 한정어를 추가하고 있었으며, 이에 더해 citation(인용)이라는 새로운 요소도 추가하고 있었다. 구체적으로 살펴보면 identifier(식별자) 요소를 세분화하여 저자의 식별력을 특히 강조하는 한정어를 활용하고 있었으며, citation(인용) 요소를 새롭게 추가하여 학술지

권, 호, 페이지, 출판사명 등 서지인용정보에 대한 자세한 기술을 시도하고 있었다.

반면 의학도서관 D에서는 새로운 요소의 추가 없이 기존의 요소에 한정어를 추가적용하고 있었다. 이는 contributor(기여자), identifier(식별자), relation(관계), rights(저작권) 4가지 요소에서 발견되는데, college(단과대학 구분), department(학과명), alternativeauthor(영문저자명), googleauthor(구글저자명) 등과 같이 주로 contributor(기여자) 요소에서 한정어를 추가하고 있어, 비록 사용된 한정어는 다르지만 대학도서관 B와 유사하게 저자에 대해 보다 자세한 기술을 시도하고 있음을 알 수 있었다. 또한 대학도서관 B와 달리 의학도서관 D에서는 relation(관계) 요소에 학술논문의 권, 호 정보를 기술하고 있었다.

이처럼 DSpace의 15개 요소 중 학술논문을 기술하기 위해 실제 활용되는 요소는 약 10개 정도였으며, 그 중 모든 기관에서 공통으로 활용되는 요소는 6개(표제, 설명문, 기여자, 날짜, 유형, 식별자) 정도였고, 3개의 요소(저작자, 정보원, 범위)는 아예 사용되지 않고 있었다.³⁾ 또한 한정어는 DSpace에서 제공하는 것의 절반 정도만 실제 활용하고 있을 뿐, 기존의 요소 아래 새로운 한정어를 추가하거나 아예 새로운 요소와 한정어를 추가하여 사용하는 경우가 많이 나타났다.

2. 메타데이터의 입력 현황

지금까지 DSpace를 활용하여 구축한 기관 리포지토리에서 학술논문에 대한 메타데이터 요소의 설계 현황을 세밀하게 살펴보았다. 그 결과, 모든 기관에서 공통적으로 활용하는 주된 메타데이터 요소가 존재하는 것은 분명하지만 이에 못지않게 기관의 상황에 따라 여러 요소와 많은 한정어가 커스터마이징 되고 있음을 확인할 수 있었다.

지금부터는 이들 요소에 입력된 ‘실제’ 데이터에 대한 상세한 분석을 시도해 보고자 한다. 먼저, 4개 기관에서 사용하는 메타데이터 요소는 대부분 동일한 의미로 활용되고 있었지만, 일부 데이터의 경우에는 동일한 의미에 대해 기관별로 상이한 요소명을 활용하고 있음을 발견할 수 있었다. 가령, 아래의 예시와 같이 학술논문이 게재된 학술지명을 기술하기 위해서 대학도서관 B에서는 완전히 새로운 요소인 citation(인용) 요소를 신설하여 그 하위에 pulicationname(출판사명)이라는 한정어를 사용하고 있는 반면, 의학도서관 D에서는 기존의 relation(관계) 요소를 적용하되 한정어로 ispartof(하위아이템)를 사용하고 있었다. 뿐만 아니라 저널의 권호정보를 기입함에 있어서도 relation(관계) 요소의 volume(권), issue(호) 한정어를 활용하고 있는 의학도서관 D와는 달리, 대학도서관 B에서는 새로 추가한 citation(인용) 요소에

3) 북미의 기관 리포지토리를 대상으로 한 Kurtz의 연구결과에 따르면 북미에서는 5개의 요소(표제, 날짜, 유형, 식별자, 언어)를 조사 대상 기관에서 공통적으로 사용하고 있었으며, 설명문과 기여자 요소를 일괄 적용하는 국내와 달리 언어 요소를 공통적으로 적용하는 특징을 보였다(Kurtz 2010, 44).

volume(권)과 issue(호)라는 한정어를 추가하여 기술함으로써, 동일한 의미를 표현하기 위해 상이한 요소를 사용하고 있음을 확인할 수 있었다.

<대학도서관 B 사례>

```
<dcvalue element="citation" qualifier="publicationname">ASIA-PACIFIC JOURNAL  
OF FINANCIAL STUDIES</dcvalue>  
<dcvalue element="citation" qualifier="volume">40</dcvalue>  
<dcvalue element="citation" qualifier="issue">2</dcvalue>
```

<의학도서관 D 사례>

```
<dcvalue element="relation" qualifier="ispartof">Korean Journal of Medical  
Education (한국의학교육학회지)</dcvalue>  
<dcvalue element="relation" qualifier="volume">18</dcvalue>  
<dcvalue element="relation" qualifier="issue">3</dcvalue>
```

또한 동일 요소 아래 한정어의 적용 여부에 따른 차이도 발견되었다. 가령, 3개 기관에서는 title(표제) 요소 아래 alternative title(대체표제) 한정어를 사용하고 있었으나 공공기관 A에서는 아래의 예시와 같이 title(표제) 요소 아래 표제 정보와 대체표제를 같이 기입하고 있었다.

```
<dcvalue element="title" qualifier="none">정부부처의 리콜정보 제공체계 개선방안 =  
How to improve Government websites on product recall Information–Analysis  
using a perspective of information systems–</dcvalue>
```

이에 더해, 동일한 요소 및 한정어에서 기술방법에 따른 차이도 발견되었다. 가령, 학술논문의 citation(인용) 정보를 기술한 아래의 사례를 살펴보면 대학도서관 B와 의학도서관 D에서는 ‘저널명, 권호정보, 페이지’를 기술하는 반면 연구소 C에서는 ‘권호정보, 페이지’만 기술하고 있었다. 또한 대학도서관 B에서는 영문학회지명만을 기술한 반면 의학도서관 D에서는 영문학회지명에 더해 한글학회지명까지 괄호로 병기하고 있어 기관에 따른 기술 방식의 차이가 존재함을 알 수 있었다.

<대학도서관 B 사례>

```
<dcvalue element="identifier" qualifier="citation">ASIA-PACIFIC JOURNAL OF  
FINANCIAL STUDIES, v.40, no.2, pp.317 – 346</dcvalue>
```

<연구소 C 사례>

```
<dcvalue element="identifier" qualifier="citation">VOL E81-D, NO 9, 968–975</dcvalue>
```

<의학도서관 D 사례>

<dcvalue element="**identifier**" qualifier="**citation**">Korean Journal of Medical Education (한국의학교육학회지), Vol.18(3) : 239~247, 2006</dcvalue>

심지어는 메타데이터 요소 및 한정어가 가지는 본래의 의미와 다르게 사용되는 경우도 있었다. 이러한 의미의 오용 혹은 해석의 차이는 특히 format(파일형식) 요소에서 많이 나타났다. 가령, format 요소 아래의 한정어(extent, mimetype) 사례를 살펴보면 다음과 같다.

<대학도서관 B 사례>

<dcvalue element="**format**" qualifier="**extent**">239~247</dcvalue>

<공공기관 A 사례>

<dcvalue element="**format**" qualifier="**mimetype**">2482068.pdf</dcvalue>

<dcvalue element="**format**" qualifier="**mimetype**">계속적_계약관계에서의_약관의_개정.pdf</dcvalue>

format(파일형식) 요소의 extent(원문크기) 한정어는 파일크기나 재생시간 등을 기술하는 요소이기 때문에 자원의 전체 페이지 수를 기술해야 하며, mimetype 한정어는 원문이 어떤 종류의 파일인지를 MIME 유형에 따라 설명하는 한정어이기 때문에 application/pdf 등으로 기술해야 한다.

이와 같은 데이터 요소의 부정확한 사용에 더해, 기술의 오류도 심심찮게 볼 수 있었다. 가령, 아래 공공기관 A의 사례에서는 저자사항과 출판사에 입력된 데이터가 동일하였다. 이와 같은 오류는 출판사와 저자사항을 혼돈하여 기술하거나 중복기술 하는 등, 다양한 원인으로 잘못 기술되고 있는 사례라 볼 수 있다.

<공공기관 A 사례>

<dcvalue element="**contributor**" qualifier="**author**">지광석</dcvalue>

<dcvalue element="**publisher**" qualifier="**none**">지광석</dcvalue>

이처럼 학술논문 기술을 위해 입력된 실제 데이터를 확인한 결과, 단순하게 요소와 한정어의 활용 유무를 떠나 훨씬 더 복잡한 문제들이 저변에 깔려 있음을 알 수 있었다. 이러한 문제는 개별 기관의 입장만을 생각하면 큰 문제가 아닐지도 모른다. 가령, 동일한 의미가 기관별로 상이하게 활용되고 있더라도 개별 기관에서 일관성을 갖춰 구축되고 있으면 해당 기관 내에서는 문제가 발생하지 않을지도 모른다. 그러나 기관 리포지토리의 거시적인 관점을 고려해 본다면 상호운용성을 보장하기 위해 이와 관련된 추가적인 논의가 보다 더 요구된다고 하겠다.

IV. DSpace 검색시스템의 활용 현황

1. 검색에 활용되는 메타데이터 요소

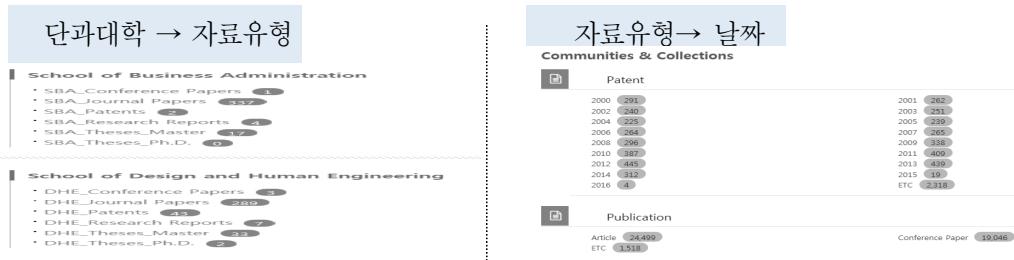
메타데이터의 구축은 자원의 검색과 식별로 직결된다. 따라서 데이터의 구축현황과 더불어 입력한 데이터가 검색에 어떻게 활용되고 있는지 함께 살펴볼 필요가 있다. 이를 위해 기관 리포지토리의 검색시스템을 ① 검색 방식, ② 검색 요소, ③ 디스플레이 요소로 나누어 살펴보았다.

흔히 검색 방식은 특정 검색 능력을 보유하고 있느냐의 여부로 결정된다. 이러한 검색 능력의 대표적인 요소는 키워드 검색과 브라우징 검색으로 대별되는 탐색질의 방식이 있다(김효정 2000, 56). 이러한 방식에 따라 살펴보면 조사 대상인 4개 기관 리포지토리 모두에서 키워드 검색과 브라우징 검색 기능을 기본적으로 제공하고 있었다(<그림 3> 참조). 그러나 키워드 검색 중 상세검색 기능의 제공여부는 기관 리포지토리 간에 차이가 존재하였는데, 조사 대상 중 3개 기관에서는 상세검색 기능을 제공하고 있는 반면, 1개 기관에서는 간략검색 기능만을 제공하고 있었다.



<그림 3> DSpace 검색 방식

그런데 <그림 3>의 브라우징 검색에서 보듯이 'Communities & Collections'이라는 DSpace만의 특징적인 브라우징 기능이 4개 기관에서 공통적으로 나타나는 것을 볼 수 있다. Communities & Collections란 커뮤니티 > 하위 커뮤니티 > 컬렉션 > 아이템의 순차적 구조에 따라 접근할 수 있는 DSpace의 브라우징 검색서비스의 이름이다. 실제 Communities & Collections 검색서비스가 구현된 사례를 살펴보면 대학도서관 B에서는 대체로 단과대학과 같은 조직구조로 브라우징 한 후 그 하위에 자료유형별로 재검색 하는 '단과대학별 → 자료유형 방식'을 제공하고 있었고(<그림 4>의 좌측 참조), 나머지 3개 기관(공공기관 및 연구소, 의학도서관)에서는 자료유형별로 브라우징 한 후 그 하위에 날짜, 기여자, 표제별로 재브라우징할 수 있는 '자료유형 → 날짜, 기여자, 표제 방식'으로 해당 서비스를 제공하고 있었다(<그림 4>의 우측 참조).



<그림 4> DSapce의 Communities & Collections 검색서비스

다음, 검색에 직접적으로 사용되는 검색 요소를 살펴보았다. 그 결과, 키워드 검색 중 간략 검색에서는 ‘표제’, ‘기여자’, ‘주제’ 요소를 검색에 활용하고 있었다. 또한 상세검색에서는 기관에 따라 적게는 3개(표제, 기여자, 날짜)에서 많게는 5개(표제, 기여자, 날짜, 자료유형, 주제)의 요소를 검색에 활용하고 있었다. 또한 필드제한 검색기능도 제공하고 있었는데, 대학도서관 B에서는 ‘단과대학별 혹은 하위 기관’에 따라 그리고 나머지 3개 기관에서는 ‘자료유형’에 따라 검색하도록 하고 있었다(<그림 5> 참조).



<그림 5> DSpace의 상세검색 요소와 필드제한 사례(대학도서관 B)

브라우징 검색에서는 <표 3>과 같이 모든 기관에서 표제, 기여자, 날짜 세 가지 검색항목에 따른 브라우징을 제공하고 있었다. 반면 주제, 키워드, 간행물명 등은 기관에 따라 선택적으로 적용하고 있었다.

<표 3> 기관별 브라우징 검색시 활용되는 검색 요소

	공공기관 A	대학도서관 B	연구소 C	의학도서관 D
표제	✓	✓	✓	✓
기여자	✓	✓	✓	✓
주제		✓		
날짜	✓	✓	✓	✓
키워드			✓	
간행물				✓

마지막으로, 검색된 결과화면에서 이용자들에게 실제 제공하는 요소를 살펴보기 위해 디스플레이 요소를 검토해 보았다. 결론부터 말하자면 학술논문의 경우 5개의 요소(표제, 기여자,

발행일, URI, 계층구조)가 공통적으로 디스플레이에 활용되고 있었으며, 3개 기관에서 발행자와 ISSN, 서지인용기술을, 2개 기관에서 초록과 키워드를, 1개 기관에서 대체표제, 언어, 학술지명, DOI를 디스플레이에 활용하고 있었다.

〈표 4〉 검색결과의 디스플레이 요소

	공공기관 A	대학도서관 B	연구소 C	의학도서관 D
표제	✓	✓	✓	✓
대체표제				✓
키워드		✓	✓	
초록		✓		✓
발행자	✓	✓	✓	
기여자	✓	✓	✓	✓
발행일	✓	✓	✓	✓
서지인용기술		✓	✓	✓
ISSN		✓	✓	✓
URI	✓	✓	✓	✓
언어		✓		
학술지명				✓
계층구조	✓	✓	✓	✓
DOI		✓		

앞서 살펴본 것을 종합해 보면 4개 기관 리포지토리는 키워드 및 브라우징 검색에서 일부 차이가 존재하긴 하지만, 편차가 가장 두드러지게 나타난 부분은 검색결과에 대한 디스플레이 요소였다. 물론 동일한 소프트웨어인 DSpace를 활용하고 있기 때문에 시스템의 외형적 큰 틀은 유사하게 나타났다. 그러나 디스플레이 요소를 면밀히 분석해 본 결과, 개별 기관의 특성이 반영되어 커스터마이징되고 있는 것을 곳곳에서 발견할 수 있는데, 학술논문의 검색 결과 디스플레이에 6개의 요소만을 제공하는 기관이 있는 반면, 그 2배에 해당하는 12개의 요소를 제공하는 기관도 존재하여 기관별 편차가 매우 뚜렷하였다. 이처럼 검색 방식 및 검색 요소, 디스플레이 요소에 이르기까지 기관 리포지토리 검색시스템의 활용 현황 역시 기관에 따라 매우 큰 편차를 보였다.

2. 검색결과에 대한 분석

다음으로 실제 몇 가지 사례를 들어 검색을 시도해 본 후, 검색결과에 대한 세부적인 분석을 시도해 보았다. 먼저, 4개의 기관을 대상으로 ‘정보’ 혹은 ‘정보시스템’으로 표제검색한 결과, 가장 눈에 띄는 특징은 검색결과의 중복적 나열이었다. 가령, 연구소 C의 사례를 살펴보면 다음 <그림 6>과 같이 동일 논문이 반복적으로 제시되는 경우를 종종 확인할 수 있었다.

Issue Date	Title	Author(s)
200409	Web-based information system for controlling membrane bioreactor	조진우; 안규홍; 송경근; 맹승규
200410	Web-based information system for controlling membrane bioreactor	조진우; 안규홍; 송경근; 맹승규
199909	Development of organic NLO materials for optical information technology	김낙중
198211	Conversion algorithm of decision table using information theory.	권용무; 박용진
201004	iLight: Information flashLight on Objects using Handheld Projector	김선준; 정재우; Alice Oh; Chris Schmandt; 김의재
200505	Web-based information system for controlling membrane bioreactor	조진우; 안규홍; 송경근; 맹승규

<그림 6> 기관 리포지토리의 검색결과 중복 사례

이러한 현상의 원인은 실제 입력된 데이터를 확인해 보면 쉽게 유추할 수 있다. 다음 <그림 7>과 같이 데이터 입력의 오류로 인해 두 아이템이 각기 다른 아이템으로 인식되어 별도의 검색결과로 디스플레이 되었던 것이다.

Web-based information system for controlling membrane bioreactor	Web-based information system for controlling membrane bioreactor
◎ Title Web-based information system for controlling membrane bioreactor	◎ Title Web-based information system for controlling membrane bioreactor
◎ Authors 조진우; 안규홍; 송경근; 맹승규	◎ Authors 조진우; 안규홍; 송경근; 맹승규
◎ Keywords membrane bioreactor; fouling; pressure; remote control; monitoring; internet	◎ Keywords membrane bioreactor; fouling; pressure; remote control; monitoring; internet
◎ Issue Date 2004-09	◎ Issue Date 2004-10
◎ Publisher Journal of Environmental Instrumental Control and Automation (Journal of EICA)	◎ Publisher Proceedings of the 16th EICA Conference
◎ Citation VOL 9, NO 2, 159-162	◎ Citation , 159-162
◎ URI http://pubs.kist.re.kr/handle/201004/25877	◎ URI http://pubs.kist.re.kr/handle/201004/26945

<그림 7> 부정확한 데이터 입력으로 인한 검색결과 중복 사례

또한 동일 연구자에 대한 다양한 표기방식으로 기여자검색한 결과, 이 역시 문제가 있음을 확인할 수 있었다. 가령, 대학도서관 B에서 활용하는 저자표기법인 국문저자명, 영문저자명, 영문축약저자명에 따라 ‘강문영’, ‘Kang, Moon Young’, ‘Kang, MY’으로 기여자검색을 시도해 보았다. 그 결과, 검색어에 따른 검색결과의 차이(‘강문영’으로 검색했을 시 1건, ‘Kang, Moon Young’으로 검색했을 시 51건, ‘Kang, MY’으로 검색했을 시 75건)가 명확하게 드러났다. 이처럼 이름의 형식이나 문자 표기방식이 검색효율을 떨어뜨리고 있음을 알 수 있었다.

이러한 현상의 원인은 실제 구축되어 있는 데이터에서 확인할 수 있었다. 가령, ‘강문영’과 ‘Kang, Moon Young’으로 검색한 결과에 해당되는 레코드를 살펴보면 아래와 같이 입력된 데이터 값의 차이가 존재함을 알 수 있었고, 이러한 차이가 실제 검색결과에 영향을 미치고 있음을 알 수 있었다.

<대학도서관 B 사례>

```
<dcterms:contributor element="contributor" qualifier="author">Kang, MY[Kang, Moon Young]
```

</dcvalue>

<dcvalue element="contributor" qualifier="author">강문영</dcvalue>

문제는 이러한 현상이 사례로 살펴본 4개의 기관에서 모두 비일비재하게 일어나고 있다는 점이다. 이러한 문제에 더해, 한 기관 내에서 검색결과로 보여주는 데이터 요소 역시 매우 큰 편차를 보이고 있었다. 가령, 대학도서관 B의 사례를 살펴보면 4개의 회의 자료의 저자사항을 기술할 때, ‘회의명과 회의연도’ 혹은 ‘회의명과 회의연도, 페이지’ 등으로 개별기간 내에서 동일한 정보자원의 유형을 다른 방식으로 표기하고 있음을 알 수 있었다(<그림 8>의 좌측 참조). 이러한 검색결과의 상이한 디스플레이 방식은 4개의 기관을 상호비교해보면 더욱 심각해진다. 가령, <그림 8>의 우측을 보면 4개의 기관별로 저자사항을 제시하는 방식을 확인할 수 있다. 구체적으로 공공기관 A와 연구소 C의 경우 ‘저자명’만을, 대학도서관 B에서는 ‘영문축약저자명(영문저자명)’의 형식으로, 의학도서관 D에서는 ‘저자명과 소속’까지 제시하고 있음을 알 수 있었다.

<p>Information Systems Education in Korea Lim, JK[Lim, Joo-Kyul] ; The Japan Society for Management Information, 1997</p> <p>Information systems management and development: a three-schema hierarchical architecture Lim, JK[Lim, Joo-Kyul] ; PRIISMProceedings of the PRIISM 96 International Conference, pp.167 - 176, 1996-01</p> <p>Ambient Information Systems Lim, YK[Lim, Youn-kyung] ; Marshall, P[Marshall, Paul J]; Hazlewood, W.R. [Hazlewood, William R.]; Rogers, Y[Rogers, Yvonne] ; UbicompUbicomp2008, 2008-09-21</p> <p>Multimedia Information Systems Whang, KY[Whang, Kyu-Young] ; Park, G[Park, Gyeong] ; Whang, KY[Whang, Kyu-Young] ; Park, S[Park, Seung-hwan] ; First Joint Workshop on Multimedia Communication(JW-MMC), pp.1 - 10, 1994-10</p>	<p>A 류미현 이수범 최승재 염기원; 박지형 강선구; 강무진; 박면웅; 박지형 C 고희동; 유병현; 서대일</p>	<p>B Lee, JN[Lee, Jae-Nam] Lim, YK[Lim, Youn-kyung] CATHEY, WT; ISHIHARA, S; Whang, KY[Whang, Kyu-Young] 정우진; 조우현; Graduate School of Public Health (보건 대학원); Graduate School of Public Health (보건대학 원); Chung, Woo D</p>
---	---	--

<그림 8> 동일한 요소의 상이한 디스플레이 방식

V. DSpace 메타데이터 요소의 활용 전략

이 연구의 목적은 DSpace를 사용하여 구축한 기관 리포지토리의 메타데이터 요소의 활용 현황을 면밀히 조사하고, 검색시스템의 활용 현황을 살펴 메타데이터가 검색에 어떻게 활용되는지를 확인한 뒤, 향후 DSpace를 활용하여 기관 리포지토리를 구축하고자 하는 기관들이 데이터 구축시 어떤 점을 주의해야 하는지 등의 가이드라인을 제공하는 것에 있다. 이러한 목적을 달성하기 위해 4개 기관을 사례로 살펴본 결과, DSpace에서 제공하는 기본적인 요소 및 한정 어는 기관에서 필요로 하는 사항을 모두 충족하지 못함을 확인하였다. 즉, DSpace를 활용하여 기관 리포지토리를 구축하고자 할 때 메타데이터의 확장은 불가피하다는 것을 알 수 있었다.

이러한 필요성에 따라 실제 기관 리포지토리에서는 메타데이터 및 검색시스템에서 DSpace

요소를 다양한 방식으로 확장하고 있었다. 그 결과, 메타데이터의 적용에서 동일한 의미가 기관에 따라 상이한 메타데이터 요소에 포함되기도 하고, 동일한 요소를 상이한 방식으로 표현하기도 하는 등 다양한 문제가 발생하고 있었다. 또한 검색에서도 동일한 자원유형에 대해 상이한 요소를 검색에 활용하거나 동일한 요소와 한정어를 상이한 방식으로 디스플레이하는 등, 메타데이터 요소의 설계 및 활용을 둘러싼 여러 문제점을 발견할 수 있었다. 이는 DSpace 소프트웨어의 강점이라 할 수 있는 커스터마이징에서 기인하는 것이라 볼 수 있는데, 자유로운 DSpace의 커스터마이징이 역설적으로 데이터의 공유와 오픈액세스를 위한 국가 차원의 아카이브 구축시 걸림돌이 될 수 있음을 의미하는 것이라고도 볼 수 있다.

그렇다고 해서 모든 기관에서 DSpace 요소를 표준화되고 일관된 방식으로 써야한다는 의미는 결코 아니다. 하지만 2007년 KAIST가 국내 최초로 DSpace를 도입한 이후 10년이 경과하였고, 2010년부터 DSpace를 활용한 여러 기관 리포지토리의 자원에 대한 통합서비스 체계를 구축하고 있는 현 상황에서, 타 기관 리포지토리의 구축 상황을 완전히 배제한 채 개별 기관의 목적과 용도에만 맞춘 기관 리포지토리의 구축은 근시안적 태도라 할 수 있다. 따라서 향후 DSpace를 활용하여 기관 리포지토리를 구축하고자 할 경우 자관의 상황을 고려하여 요소 및 한정어를 커스터마이징하되, 타 기관과의 데이터 공유에 대한 가능성 역시 고려할 필요가 있으며, 이와 관련하여 모든 기관에서 공통적으로 적용 가능한 원칙을 세울 필요가 있다.

따라서 이 연구에서는 DSpace를 활용하여 기관 리포지토리를 구축하고자 하는 기관들이 메타데이터 입력과 활용시 고려해야 할 몇 가지 사항을 제안하고자 한다.

첫째, DSpace에서 제공하는 기본 요소를 최대한 활용하는 방법을 모색하여야 한다. 기관 리포지토리 개발시 오픈 소스 소프트웨어를 활용하는 이유는 우선적으로 개발 비용의 절약에 있지만 이보다 상호운용성을 확보하고자 하는 목적이 더 크다고 할 수 있다. OAK 사업추진을 위해 DSpace를 선택한 것 역시 DSpace가 전 세계적으로 가장 많이 활용되고 있는 오픈 소스 소프트웨어이기 때문에 여러 기관별 상호운용성을 확보하는 것이 용이할 것이라는 이유가 중요하게 작용하였다. 따라서 향후 기관 리포지토리간의 데이터 공유와 나아가 전 세계적인 공유와 협력체제의 구축을 고려한다면 DSpace에서 제공하는 ‘기본’ 요소와 한정어를 우선적으로 활용하는 것이 필요할 것이다.

둘째, DSpace에서 제공하는 기본 요소로 부족할 경우 불가피하게 새로운 요소의 추가 · 확장을 검토해야 하는데, 이 때 새로운 요소를 추가하기보다는 가능한 현재 제공되는 요소 아래 새로운 ‘한정어’를 확장하는 것을 우선적으로 검토해야 할 것으로 보인다. 이는 적절한 요소 아래 새로운 한정어를 추가함으로써 기존의 요소를 보다 상세하게 기술할 수 있고 덤다운 원칙에도 위배되지 않는 방법이기 때문이다. 덤다운 원칙은 메타데이터의 표준화를 위한 기본 원칙으로써 한정어 때문에 오히려 그 뜻이 모호해지는 것을 방지하기 위해 한정어를 삭제하

고 상위 요소만으로도 의미전달을 가능하게 하는데 목적이 있다. 따라서 기관 리포지토리에서는 새로운 요소로 확장하기보다는 새로운 한정어를 추가함으로써 향후 기관 리포지토리들 간의 상호운용성을 보장하는 방법을 채택하여야 할 것이다.

셋째, 요소를 기술할 때 가급적 한정어를 활용하여 내용을 구분지어서 기술할 필요가 있다. 앞서 ‘표제’ 요소에 ‘대체표제’에 해당하는 데이터를 한정어로 구분하지 않고 함께 기술했던 사례를 상기해보자. 해당 한정어를 활용하지 않고 요소 안에 구분 없이 기술할 경우, 향후 개별 기관의 상이한 요소를 대상으로 한 매핑작업시 걸림돌이 될 수밖에 없다. 따라서 이 같은 경우는 데이터 입력시부터 의미에 해당하는 정확한 한정어로 구분지어 기술하여야 한다.

넷째, 데이터 설계와 입력시 개별 데이터를 검색에 어떻게 활용할 것인가에 대한 충분한 고려도 필요하다. 주지하다시피 검색과 검색결과의 디스플레이에서 활용하고자 하는 요소는 입력된 데이터의 수준과 품질에 따라 달라질 수밖에 없다. 앞서 살펴본 바와 같이 동일 저자의 다른 표기방식에 따라 검색결과가 달리 나타난 이유도, 기관에 따라 ‘저자명’, ‘저자명과 소속’, ‘저자명과 영문저자명’ 등으로 디스플레이가 상이하게 제공되고 있는 이유도 결국 입력된 데이터의 차이에서 기인하는 현상이라 볼 수 있다. 따라서 기관 리포지토리의 운영에 앞서 메타데이터를 어떻게 검색에 활용할지에 대한 충분한 논의가 요구된다.

이와 같은 메타데이터 요소 설계시 고려해야 할 원칙이 제대로 지켜지려면 기관 리포지토리의 상황에 적합한 운영 전략 또한 필요하다. 이와 관련하여서는 서두에서 밝힌 것처럼 기관 리포지토리의 일차적 목적과 거시적인 목적에 따라 추진전략을 나누어 제안해보고자 한다.

기관 리포지토리를 통해 기관의 연구자들이 생산해 낸 자원의 수집과 영속적인 관리라는 미시적인 목적을 달성하기 위해서는 가장 먼저 구성원들을 대상으로 한 철저한 교육이 요구된다. 기관에 따라 다르긴 하지만 기본적으로 기관 리포지토리는 저자 셀프 아카이빙을 전제로 구축되기 때문에 저자의 자발적인 참여와 그 과정에서 생산해 내는 메타데이터가 주축이 된다. 그러나 기관 구성원들은 메타데이터의 기본 개념과 중요성에 대한 인식이 부족할 수밖에 없다. 따라서 그들을 대상으로 한 우선적이고도 철저한 교육을 통해 저자가 메타데이터 입력시 발생할 수 있는 문제를 미연에 방지할 필요가 있다.

이러한 교육은 비단 기관 구성원들만을 대상으로 해서는 안 된다. 실제 메타데이터의 추가 및 검증작업은 기관 리포지토리 담당자의 몫이다.⁴⁾ 따라서 개별 기관의 리포지토리 품질은 결국 운영자들의 메타데이터에 대한 철저한 검토 작업을 통해서만 보장될 수 있다. 가령,

4) 저자가 기관 리포지토리의 데이터를 입력하는 것은 사실이지만 실제 그들에게 요구하는 메타데이터 입력 수준은 그리 높지 않다. 그들이 자원을 등록하려고 하면 먼저 해당 자료가 속하는 계층구조(가령, 대학별 학과, 연구소의 연구실 등)를 입력한 후, 메타데이터 중 일부 요소만을 입력한다. 가령, 통제어휘 등을 활용한 ‘subject’가 아닌 ‘keywords’ 리스트를 입력하게 되어 있어 가능한 한 간략화 된 메타데이터를 입력하도록 하고 있다. 따라서 실제 데이터의 품질과 관련된 최종 책임은 기관 리포지토리 담당자의 몫이라 볼 수 있다.

DSpace를 커스터마이징하여 제공하는 서울대의 S-Space에서는 저자가 직접 메타데이터를 입력하거나 리포지토리 담당자에게 생산한 저작물을 제출하는 두 가지 방법을 활용하고 있는데, 두 경우 모두 업무 담당자가 메타데이터를 보완하고 저작권 정책을 확인한 후 승인하는 과정을 거쳐야만 데이터를 업로드하도록 규정화하고 있다. 이처럼 기관 리포지토리 담당자의 철저한 검증작업과 이에 대한 노력은 저자 셀프 아카이빙을 근거로 운영되는 기관 리포지토리의 데이터 품질을 유지하는데 필수적인 과정이라 할 수 있다.

이에 더해, 기관 리포지토리가 가지고 있는 거시적인 목적을 달성하기 위해서는 타 기관 메타데이터와의 상호운용성을 보장하기 위한 협력체계의 역할이 무엇보다 중요하다. 물론 여기서 이야기하는 협력체계의 역할은 다양한 측면을 포함할 수 있다. 가령, “의미적 상호운용성을 보장하기 위해서는 단순히 같은 시스템을 쓰는 것만으로 가능한 것이 아니라 요소의 선정과 이에 대한 활용 및 생성 방침, 활용 수준에 대한 구체적인 가이드라인이 필요하다(정은경 2008, 207)”는 지적처럼, DSpace를 활용하고 있는 기관 간에 데이터 구축을 위한 가이드라인의 개발부터 데이터의 실질적 공유를 위한 기반의 마련 등 협력체계의 역할이 보다 중요해지고 있다. 현재, DSpace의 국내 운영주체는 국립중앙도서관이 담당하고 있기 때문에 이러한 협력체계의 구축은 비교적 용이한 구조라 할 수 있다. 따라서 국립중앙도서관은 DSpace를 활용하는 기관 리포지토리들 간의 상시 소통채널을 가동하면서 개별 리포지토리가 당면한 문제에 공동으로 대응하는 구조를 갖추어야 할 것이다. 더불어, 국가 차원의 OAK 구축 및 운영을 효과적으로 진행하기 위해 개별 리포지토리의 협력을 유도하는 구심체 역할을 수행해야 할 것이다.

VII. 결론

지금까지 DSpace를 기반으로 구축된 기관 리포지토리의 메타데이터 요소의 설계 및 구축 사례를 살펴보고, 구축된 메타데이터가 검색에 어떻게 활용되는지 살펴보았다. 이에 더해, 조사 결과를 바탕으로 기관 리포지토리의 메타데이터 구축을 위한 원칙과 추진전략을 제안하였다. 물론 일부 기관을 선정하여 살펴본 결과이기에 일반화시켜 기관 리포지토리의 현 실태를 논하는 것은 매우 위험하고도 섣부른 일일지도 모른다. 하지만 현재 DSpace를 기반으로 하는 국내 기관 리포지토리의 메타데이터를 둘러싼 현황을 가감 없이 살펴보았다는 측면에서 이 연구의 의미를 두고자 한다.

조사 결과에 따르면 DSpace를 활용한 기관 리포지토리는 기관의 연구자들이 생산한 다양한 학술자원을 대상으로 구축되고 있으며, 그 결과 기존의 도서관에서 수집하는 자원의 유형

을 뛰어 넘는 범주까지 관리함으로써 도서관의 역할을 확장시킬 수 있는 가능성을 확인하였다. 이는 기관 리포지토리의 활성화를 통해 도서관이 또 다른 기회를 얻을 수 있다는 것을 의미한다. 그러나 구축되어 있는 메타데이터와 검색시스템을 면밀히 살펴본 결과, 현재 기관 리포지토리는 개별 기관의 데이터 구축에 관심이 집중되어 있으며 향후 상호운용성을 갖춘 오픈액세스의 전초기지로써의 준비는 아직 미흡한 것으로 확인되었다.

구체적으로 살펴보면 DSpace에서 현재 제공되고 있는 메타데이터 요소는 자유롭게 사용하거나 개별 기관의 상황에 맞추어 수정·확장하여 쓸 수 있는데, 데이터 분석 결과 DSpace의 이러한 장점은 양날의 검처럼 작용하였다. 즉, 개별 기관에서 자유롭게 커스터마이징하는 과정에서 여러 문제를 발생시켰는데, 이러한 제한점은 검색시스템의 성능으로 직결되는 것 또한 확인할 수 있었다. 이에 더해, 검색 방식 및 검색 요소, 디스플레이 요소에 이르기까지 기관별 편차는 매우 크게 나타났으며, 검색결과에서 나타나는 문제점 역시 입력된 메타데이터의 품질에서부터 비롯되는 것임을 알 수 있었다.

따라서 향후 DSpace를 활용하여 기관 리포지토리를 구축하고자 할 경우, 일차적으로 개별 기관의 데이터 구축 시 데이터의 정확성 및 일관성을 갖춰야 되는 것은 기본이고, 나아가 기관 리포지토리의 궁극적인 목적인 오픈액세스를 실현하기 위해 여러 주체간의 협력을 기반으로 한 메타데이터의 구축 및 관리가 요구된다. 또한 개별 기관에서 구축한 기관 리포지토리 메타데이터의 상호운용성을 보장하기 위한 다양한 방책에 대한 연구 역시 절실히 필요하다. 이에 더해 도서관계에서는 작금의 상황에 대해 분명히 인지하고 개별 기관에서의 충실한 데이터 관리는 물론 향후 데이터 공유를 위해 보다 더 큰 관심을 쏟아야 할 것이다. 나아가 구심체 역할을 수행하고 있는 국가도서관에서 보다 적극적이고 공격적인 협력체계의 구축 및 활성화를 위한 노력 역시 필요할 것이다. 이 연구가 이와 같은 여러 주체들의 관심과 적극적인 참여를 촉발하고, 궁극적으로 기관 리포지토리가 진정한 정보공유의 장으로 거듭나는데 있어 긍정적인 촉매제가 되기를 기대한다.

참고문헌

- 곽승진, 신재민. 2008. 오픈액세스 리포지터리의 사용성 평가에 관한 연구. 『정보관리연구』, 39(4): 67–96.
- 김미향, 나혜란. 2010. 기관 리포지터리 S-Space 구축과 활성화 전략. 『도서관보』, 132: 25–44.
- 김현희, 정경희, 김용호. 2006. 대학 기관 리포지토리의 운영 현황 분석 및 개선 방안에 관한 연구. 『정보관리학회지』, 23(4): 17–39.

- 김효정. 2000. 이미지정보 탐색을 위한 웹 검색엔진의 비교 평가.『정보관리연구』, 31(4): 50–70.
- 백종명, 이은지. 2013. 국내 기관 리포지터리 콘텐츠 현황분석 및 개선방안에 관한 연구.『한국정보 관리학회 학술대회 논문집』. 2013년 8월 19일. 서울: 한성대학교, 83–87.
- 이진솔. 2015. 연구중심대학 도서관의 기관 오픈 액세스 리포지터리 구축·운영 사례.『디지털도서관』, 80: 31–39.
- 장금연. 2004. 학술정보유통을 위한 레포지터리 적용 방안 연구.『한국문헌정보학회지』, 38(4): 291–310.
- 정연경, 이나니, 이미화. 2007. 기관 레포지터리 활성화를 위한 메타데이터의 확장 방안 연구.『한국문헌정보학회지』, 41(1): 323–344.
- 정은경. 2008. 디지털 도서관의 의미적 상호운용성 향상을 위한 메타데이터 요소와 활용에 관한 연구.『한국문헌정보학회지』, 42(1): 193–211.
- OAK Portal 홈페이지. <<http://www.oak.go.kr>> [인용 2016. 10. 20].
- Castagné, Michel. 2013. “Institutional repository software comparison: DSpace, EPrints, Digital Commons, Islandora and Hydra.” *LIBR*, 596: 1–15.
- Chapman, John W., David Reynolds and Sarah A. Shreeves. 2009. “Repository Metadata: Approaches and Challenges.” *Cataloging & Classification Quarterly*, 47: 309–325.
- Dora, Mallikarjun and Bulu Maharana. 2008. Scholarly Communication through Institutional Repositories. <<https://core.ac.uk/download/pdf/11889660.pdf>> [cited 2016. 10. 25].
- DSpace Home Page. <<http://dspace.mit.edu/>> [cited 2016. 10. 10].
- Herbert, Van de Sompel and Carl Lagoze. 2002. “Notes from the interoperability front: A progress report on the Open Archives Initiative.” *International Conference on Theory and Practice of Digital Libraries*, 144–157.
- Kurtz, Mary. 2010. “Dublin Core, DSpace, and a Brief Analysis of Three University Repositories.” *Information Technology and Libraries*, 29(1): 40–46.
- Lynch, Clifford A. 2003. “Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age.” *ARL*, 3(2): 327–336.
- MacColl, John, Richard D. Jones and Theo Andrew. 2006. *The institutional repository in the digital library*. <<https://www.era.lib.ed.ac.uk/bitstream/1842/858/1/Chapter1.pdf>> [cited 2016. 10. 10].
- Patel, Yatrik. 2011. *DSpace installation Guide*. <<http://dlib.ionio.gr/portal/.../15-installing-dspace-on-windows.html>> [cited 2016. 11. 5].
- Repository map Home Page. <<http://maps.repository66.org/>> [cited 2016. 10. 10].

국한문 참고문헌의 영문 표기

(English translation / Romanization of reference originally written in Korean)

- Chung, Eun-Kyung. 2008. "The Utilization of Metadata Elements and Content Designation for Improving Semantic Interoperability in Context of Digital Libraries." *Journal of the Korean Society for Library and Information Science*, 42(1): 193–211.
- Chung, Yeon-Kyoung, Na-Nee Lee and Mihaw Lee. 2007. "A Study of an Extension of Metadata for Institutional Repository." *Journal of the Korean Society for Library and Information Science*, 41(1): 323–344.
- Jang, Kum-Yeoun. 2004. "A Study on Repository Construction Scheme for Scholarly Communication." *Journal of the Korean Society for Library and Information Science*, 38(4): 291–310.
- Kim, Hyo-Jung. 2000. "Comparison and Evaluation of Web-based image search engines." *Journal of Information Management*, 31(4): 50–70.
- Kim, Hyun-Hee, Kyoung-Hee Joung and Yong-Ho Kim. 2006. "A Study on the Improvement of University Institutional Repositories (dCollection) based on its Current State." *Journal of the Korean Society for Information Management*, 23(4): 17–39.
- Kwak, Seumg-Jin and Jae-Min Shin. 2008. "A Study on the Usability Evaluation of Open Access Repository." *Journal of Information Management*, 39(4): 67–96.
- Lee, Jinsol. 2015. "A Case of Research-Centered University Library: Development of Open Access Institutional Repository." *Digital Library*, 80: 31–39.
- Paek, Jong-Myung and Eun Jee Lee. 2013. "A Study on the Contents of Institutional Repositories in Korea." *Proceedings of Korea Society for Information Management*, 20: 83–87.

