

dCollection의 링크드 데이터 구축에 관한 연구*

A Study on Configuring dCollection as the Linked Data

노 영 희(Young-Hee Noh)**

< 목 차 >

I. 서론	VI. dCollection용 링크드 데이터 실험적 구축
II. 이론적 배경	1. dCollection의 자원유형
1. 관련연구	2. dCollection의 데이터 구조 및 클래스 정의
2. 링크드 데이터 구축도구	3. 링크드 데이터 구축 결과
III. 링크드 데이터로 구축된 사례	VII. 결론 및 제언
IV. 연구질문	
V. 연구설계 및 방법론	

초 록

기관리포지터리와 링크드 데이터는 자원을 공동 활용하자는 정신에 있어 매우 밀접한 관련이 있다고 할 수 있다. 따라서 이 개념을 연계시킴으로써 새로운 접근방법에 의한 정보공동생산과 활용이 가능할 수 있을 것으로 본다. 이에 본 연구에서는 국내 dCollection을 링크드 데이터로 실험적으로 구축하고자 하였으며, 이를 위해 첫째, dCollection의 데이터 구조를 분석하였다. 둘째, dCollection이 구축대상으로 하고 있는 자원유형에 대해서 조사하였다. 셋째, dCollection의 자원유형에 해당하는 자원 중 링크드 데이터로 구축되어 있는 사례를 조사하였다. 그 중에서 가장 많이 구축되어 있는 자원의 사례를 집중 분석하였다. 넷째, dCollection의 메타데이터를 링크드 데이터로 구축하는 방법과 링크드 데이터로 연계될 수 있는 방안을 구체적으로 검토하였다. 마지막으로 링크드 데이터로 갈 경우 발생할 수 있는 문제점 등을 논의하였다.

키워드: 링크드 데이터, 기관리포지터리, 오픈엑세스, 도서관 3.0

ABSTRACT

The institutional repository and linked data share a purpose, co-creation and joint usage of information resources. Therefore, a new approach linking these two concepts can be utilized for co-production and utilization of resources. This study hoped to configure the Korean dCollection repository as linked data. For this purpose, first, we analyzed the current data structure of dCollection. Second, we investigated the resource types which dCollection is targeting. Third, we focused and analyzed a case study of resource types targeted by dCollection constructed as linked data. Fourth, this study examined in detail how to build the linked dCollection data and how to connect this linked data to the linked cloud. Finally, we discussed the problems that might occur in the process of building the linked data.

Keywords: Linked data, Repository, Open access, dCollection, Linked open data, Library 3.0

* 이 논문은 2012년도 건국대학교 학술진흥연구비 지원에 의한 논문임.

** 건국대학교 문헌정보학과 부교수(irs4u@kku.ac.kr)

• 접수일: 2012년 5월 24일 • 최초심사일: 2012년 6월 1일 • 최종심사일: 2012년 6월 28일

I. 서론

오픈엑세스 운동은 연구기관이 연구성과물을 출판하고 있음에도 불구하고 그 연구기관들이 해당 자원을 이용하기 위해 고가의 비용을 지불하고 출판사로부터 다시 사와야 하는 유통구조, 그리고 급격하게 상승하는 저널구독비 문제를 해결하기 위한 것으로, 저자 및 연구기관들이 생산한 자원을 공동으로 구축하고 활용할 수 있도록 하자는 운동이다.

이러한 오픈엑세스 운동의 결과로 나타난 것이 기관리포지터리이다. 즉 기관리포지터리는 정보에 대한 자유로운 접근과 이용을 도모하기 위한 오픈엑세스 정신으로 법적, 경제적, 기술적 장벽 없이 전 세계 이용자 누구라도 정보에 접근할 수 있도록 정보를 공유하고 장기적으로 보존하기 위한 것이다.¹⁾

웹 3.0 및 도서관 3.0의 키워드로 언급되고 있는²⁾ 링크드오픈데이터(Linked Open Data) 또는 링크드 데이터(Linked Data)는 시맨틱 웹이 표방하는 데이터 웹(Data Web)을 구체적으로 구현하는 방법으로, 웹으로 접근가능한 이름(URI)을 붙이고 이를 통해서 RDF 형태의 시맨틱 데이터를 서로 연결함으로써 데이터를 공개하고, 공유하고, 연결하기 위한 방법이다. 즉 시맨틱 웹은 단순히 정의(defined)되고 링크된(linked) 데이터를 보여주기 위한 목적만이 아니라 데이터를 자동화하고 연계하여 재사용하고자 하는 목적을 가지고 있고, 링크드 데이터는 이것을 가능하게 하는 것이다.

이러한 기관리포지터리와 링크드 데이터는 자원을 공동 활용하자는 정신에 있어 매우 밀접한 관련이 있다고 할 수 있다. 따라서 이 개념을 연계시킴으로써 새로운 접근방법에 의한 정보공동생산과 활용이 가능할 수 있을 것으로 보인다.

기관리포지터리의 구축방향도 자원의 공동구축과 연계에 있다고 할 수 있다. 연구기관에서 생산되는 다양한 유형의 정보자료 유형을 최대한 포괄하고자 하며, 기존에 구축된 관련 시스템들(ex, 연구정보시스템, 도서관목록, 강의자료시스템 등)과의 연계를 통해 자원이용 효율성을 최대화 시키고자 하는 노력을 하고 있다.³⁾

세계적으로 기관리포지터리의 수는 2,300여건을 상회하고 있고 우리나라에도 11건이 있으며, 그 중에서 대학도서관 중심으로 구축되고 있는 국내 기관리포지터리의 대표적인 예로 dCollection이 있다. 오픈엑세스를 지향하는 이러한 수많은 기관리포지터리를 개방형의 링크드 데이터로 각각이 구축하고 공동 활용한다면, 세계의 지식공유측면에서 기여하는 바가 클 것이고, 자원효율성도 최대화시킬 수 있을 것이다.

이에 본 연구에서는 국내 dCollection시스템을 링크드 데이터로 실험적으로 구축하고자 하였다.

1) 장금연, "학술정보유통을 위한 레포지터리 적용 방안 연구," 한국문헌정보학회지, 제38권, 제4호(2004), pp.291-310.
2) 노영희, "도서관 3.0의 개념과 서비스 모형에 관한 연구," 정보관리학회지, 제27집, 제4호(2010), pp.283-307.
3) 이나니, 김선미, 이미화, 오픈엑세스 환경에서의 학술·연구저작물 서비스를 위한 라이선스 관리 방안(KERIS 이슈리포트), 서울 : 한국교육학술정보원, 2004.

이를 위해 첫째, 링크드 데이터로 구축할 dCollection의 데이터 구조를 분석하였다. 둘째, 링크드 데이터 구축 사례 중에서 도서관에 적용 가능한 링크드 데이터 구축 사례를 총체적으로 조사하였다. 셋째, dCollection과 같은 기관리포지터리 중 이미 링크드 데이터로 구축된 사례를 조사하였다. 넷째, dCollection의 메타데이터를 링크드오픈데이터로 연계 구축하는 방안을 구체적으로 검토하였다. 마지막으로 링크드오픈데이터로 갈 경우 발생할 수 있는 문제점 등을 논의하였다.

II. 이론적 배경

1. 관련연구

본 연구에서는 선행연구를 살펴봄에 있어 두 가지 방향으로 나누어 간단히 살펴보고자 하였다. 본 연구는 기관리포지터리인 dCollection의 링크드 데이터로의 실험적 구축이므로 dCollection 및 기관리포지터리 관련 선행연구와 링크드 데이터 구축 관련 연구 및 사례를 중심으로 살펴보고자 한다. 특히 dCollection과 관련된 연구는 구축사례연구, 해외사례조사연구, 메타데이터개발 관련 연구, 저작권 관련 연구 등 수없이 많이 있지만, 본 연구 진행에 직접적으로 관련이 있는 논문 등에 대해서만 간단히 살펴보았다.

먼저, 메타데이터 개발사례로서 현재 dCollection 메타데이터의 개발에 기반이 된 연구이다. 정연경, 이나미, 이미화⁴⁾는 국내 대표적인 리포지터리인 dCollection에서 다양한 유형의 학술자원을 기술할 수 있도록 SOMS(Scholarly Object Metadata Set) 메타데이터의 확장안을 제시하였다. 이를 위해 국내외 기관 리포지터리의 동향을 메타데이터와 관련된 측면에서 조사하였으며, DSpace를 개발·운영하고 있는 MIT와 MIT 메타데이터를 확장해 사용하는 브뤼셀 대학의 DSpace 메타데이터를 분석하고 SOMS에서 확장이 필요한 요소를 추출하였다. 연구자들이 제안한 SOMS는 다양한 유형의 자원을 기술할 수 있도록 개발되었다. 이러한 메타데이터 개발과 관련된 또 다른 연구로 dCollection 메타데이터의 활용 실태를 분석하고 시스템 개선방향을 제시한 연구로 김현희 등의 연구가 있다.⁵⁾

다음으로 기관리포지터리에 어떠한 자원유형이 포함되어야 하는지와 관련된 연구 및 사례조사연구로 Lynch와 Lippincott⁶⁾에 따르면, 미국내 기관 리포지터리 구축대상자원은 약 35가지가 넘는

4) 정연경, 이나미, 이미화, “기관 레포지터리 활성화를 위한 메타데이터의 확장 방안 연구,” 한국문헌정보학회지, 제41집, 제1호(2007), pp.323-344.

5) 김현희 외, 국가지식정보 유통체계 구축을 위한 dCollection 활용실태·효과분석 연구, 서울 : 한국교육학술정보원, 2006 ; 김현희, 정경희, 김용호, “사용성 평가를 통한 dCollection 시스템 개선 방안 연구,” 한국도서관·정보학회지, 제37집, 제4호(2006), pp.327-350.

4 한국도서관·정보학회지(제43권 제2호)

것으로 조사되었으며, 그 중 가장 많이 구축된 자원은 심사전논문(e-print), 학위논문, 기술보고서 및 진행중보고서, 전자화된 도서관의 특수장서, 디지털 이미지 등의 순으로 보고하였다. Mark Ware Consulting의 2004년 보고서⁷⁾에서의 자원유형의 순위도 위의 연구와 유사한 순위로 나타나고 있다. 반면 국내의 경우 학위논문의 구축비율이 거의 90%를 상회하고 있는 것이 특징적이다.

또한 기관리포지터리를 구축함에 있어서 저작권 문제와 관련된 연구가 있으며, 이들은 공통적으로 기관리포지터를 오픈엑세스 운동의 결과로서 정보공개를 목적으로 하기 때문에 CCL적용 운동의 필요성을 주장하는 연구,⁸⁾ 일본의 전개양상을 통해서 한국 기관리포지터리가 활성화 될 수 있도록 학·협회 저작권정책 데이터베이스의 구축을 주장한 연구,⁹⁾ 오픈엑세스를 위한 저작권관리 시스템 사례연구를 수행하고 dCollection 라이선스관리시스템을 분석한 연구¹⁰⁾ 등이 있다.

한편, 도서관에 링크드 데이터 적용가능성을 타진한 연구로, 조명대¹¹⁾는 Linked Data를 활용함으로써 전 세계적으로 도서관을 연결하는 것이 가능하다고 주장하고, 실제 링크드 데이터로 발행된 LIBRIS, LCSH 등의 사례분석, FRBR 및 RDA와의 협조방향도 제시하였다. 오삼균, 김성훈, 장원홍¹²⁾은 통합 LLD 설계 및 확산방향에 대한 연구에서 W3C의 특별 작업반 형태로 운영되어 온 도서관 링크데이터(LLD) 관련 국제 활동상황을 분석하고, LLD가 도서관 서비스에 미칠 혜택을 사례연구를 통해 우선 숙고한 후, 보다 의미 있는 LLD를 위한 글로벌 식별체계, 이용 가능한 어휘와 데이터 세트의 연결 문제 및 RDF, SKOS, Microformats, RDFa 등의 LLD 적용 방안을 논하였다. 아울러 LLD 수용 및 LLD 기반 신 정보서비스 구축의 장애 요인들을 분석하고, 그 분석의 바탕 위에 LLD 기반 신 디지털정보시스템 설계 및 LLD 확산을 위한 권고사항을 제시하고 있다.

최근에 수행된 연구로,¹³⁾ 보통 도서관이 목록데이터의 품질을 향상시키기 위해 관련 비용을 추가하기 어렵고 개별기관 또는 단일 분야가 가진 자원의 한계를 완화하기 위한 하나의 방안으로, 즉 서지정보를 확장하기 위한 방안으로 링크드 데이터를 선정하였으며, 링크드 대상 데이터로서 전자

-
- 6) Lynch, C. A., & J. K. Lippincott, "Institutional repository development in the United States as of early 2005." D-Lib Magazine, 11(9)(2005).
 - 7) Mark Ware Consulting, 2004. "Pathfinder Research on Web-based Repositories: Final Report." Publisher and Library/Learning Solutions, [cited 2012, 03, 16]. <<http://www.palsgroup.org.uk>>.
 - 8) 박연희, "디지털 자원으로서 학위논문 이용 변화에 관한 연구," 제18회 한국정보관리학회 학술대회 논문집, 2010, pp.23-28.
 - 9) 조재인, "일본의 전개 양상을 통해서 본 한국 기관 레포지토리의 과제," 정보관리학회지, 제26집, 제1호(2009), pp. 35-55.
 - 10) 박미성, "오픈엑세스를 위한 저작권관리시스템 사례연구를 통한 dCollection 라이선스관리시스템 분석," 한국도서관·정보학회지, 제36집, 제4호(2005), pp.255-284.
 - 11) 조명대, "도서관에서의 Linked Data 활용방안에 관한 연구," 한국문헌정보학회지, 제44집, 제1호(2010), pp.181-198.
 - 12) 오삼균, 김성훈, 장원홍, "도서관 링크데이터(Library Linked Data) 현황분석과 통합 LLD 설계 및 확산 방향에 대한 고찰," 한국도서관·정보학회지, 제42집, 제4집(2011), pp.331-351.
 - 13) 박지영, "링크드데이터 방식을 통한 서지 정보의 확장에 관한 연구," 정보관리학회지, 제29집, 제1호(2012), pp. 231-251.

데이터와 서지데이터를 선정하였다. 서지데이터인 오픈 라이브러리와 BNB, 그리고 전거데이터로서 VIAF와 FAST¹⁴⁾를 프로티지를 통해 반입하여 실험적으로 연계함으로써 도서관의 서지정보를 확장시킬 수 있는 방안을 제안하였다. 목록데이터를 링크드 데이터로 구축하는 방안을 학술논문을 대상으로 실험적으로 구축한 연구도 있다.¹⁵⁾ 그는 Linked Data의 도서관분야 적용현황을 간단히 살펴보고 실제로 학술지 기사를 대상으로 실험적으로 간단히 구축한 사례를 보여 주었다.

2. 링크드 데이터 구축도구

링크드 데이터를 구축하기 위해서는 비구조적 데이터를 구조화하되 시맨틱 형태로 표현해야 한다. 자원을 구조화하는 것은 데이터를 RDF로 표현하는 것이며, 하나의 자원을 주어, 술어, 목적어로 표현하는 것이며, 각각의 자원은 URI를 통해 고유 식별자를 가지고 여러 가지 속성들을 이용하여 자원들 간의 링크를 생산하는 것이다. 링크드 데이터를 저작하거나 편집할 수 있는 시스템 또는 도구는 Loomp, Tabulator, mSpace Data Picker 등이 있다. Loomp는 링크드 데이터 생성 시스템으로서 사용자가 데이터에 대한 유의미한 주석을 달고, 의미적으로 관련된 데이터를 찾아 연결시킬 수 있도록 한다.¹⁶⁾ Tabulator는 시맨틱 웹 브라우저로서 링크드 데이터의 표현 형식인 RDF 문서를 탐색하고 편집할 수 있는 기능을 제공한다.¹⁷⁾ mSpace Data Picker는 개별화된 링크드 데이터 탐색 시스템으로서 사용자의 특정 목적에 맞는 링크드 데이터만 탐색할 수 있도록 하여, 마치 개별화된 링크드 데이터를 생성하여 탐색하는 효과를 제공한다.¹⁸⁾ 또한 D2RQ¹⁹⁾는 오픈소스로서 베를린 대학 연구진에 의해 개발된 것으로 Non-RDF를 RDF로 변환해 주는 시스템이며, 이미 구축된 온톨로지의 개념에 인스턴스를 부착하는 형태이다.

국내에서 개발된 시스템으로 (주)탑쿼드란트코리아가 개발한 OntoBase2.0²⁰⁾은 온톨로지 정보를 트리플 형태로 저장 및 관리하며, SPARQL 질의 성능을 보유한 온톨로지 Repository 구축 솔루션으로서 상업용 시스템이다. 2011년에는 고려대학교 컴퓨터교육학과 연구진에 의해 개발된 링크드 데

14) FAST Linked Data. <<http://experimental.worldcat.org/fast>>.

15) 박지영, "정보공유를 위한 Linked Data의 실험적 구축," 2011년도 한국정보관리학회 추계학술대회, 2011년 11월 11일.

16) Luczak-Rosch, M. & Heese, R. "Linked Data Authoring for Non-Experts," Proceedings of the 2009 Linked Data on the Web Workshop 2009(LDOW09).

17) Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A. & Sheets, D. "Tabulator: Exploring and Analyzing linked data on the Semantic Web," Proceedings of the The 3rd International Semantic Web User Interaction Workshop(SWUI06) workshop, 2006.

18) Smith, D., Popov, I. & Schraefel, "Data Picking Linked Data: Enabling Users to create Faceted Browsers," Proceedings of the 2010 Web Science Conference, m. 2010.

19) D2RQ. <<http://d2rq.org/>>.

20) (주)탑쿼드란트코리아. 2012. OntoBase 2.0. [cited 2012. 04. 13]. <<http://ontobase.topquadrant.co.kr/>>.

이터 생성시스템이 있는데, 이는 URI나 RDF에 대한 지식이나 사용능력이 없는 일반 사용자들이 서로 다른 RDF 문서에 존재하는 데이터를 연결하여 링크드 데이터를 생성할 수 있는 시스템이다.

본 연구자는 향후 지속적으로 링크드 데이터를 구축함에 있어 일반인을 위한 링크드 데이터 생성 시스템들이 많이 개발되고, 이를 활용한 데이터 구축이 바람직하다고 본다. 일반인을 위해 개발된 링크드 데이터 생성시스템의 사용을 권장하는 이유는 RDF로 생성해야 할 문서의 양 때문이다. 웹 3.0 또는 웹 4.0 시대에는 모든 웹이 시맨틱 웹으로 기계가 이해할 수 있는 웹이 구축된 시대라고 예측되고 있다. 그러나 특정 전문가 집단이 지상의 모든 정보를 RDF로 구축, 변환하는 것은 불가능하다. 따라서 이용자 참여형으로 나갈 수밖에 없는 상황이며, RDF 문서 및 링크드 데이터를 매우 쉽게 생성할 수 있는 시스템의 개발방향은 매우 바람직하다고 본다. 이는 현재 웹 2.0의 이용자생성 콘텐츠 구축에서 한 단계 더 나아가 구조화된 이용자생성콘텐츠를 구축하는 것이며, 과거 일반인이 홈페이지를 개발하기 어려웠던 문제를 해결하기 위해 블로그가 나온 것과 같은 맥락에서 이해해도 될 것이다. 그러나 시스템을 확보하는데 어려움이 있어 본 연구에서는 Protégé 4.2 Beta²¹⁾를 사용하고 있다.

Ⅲ. 링크드 데이터로 구축된 사례

본 연구에서는 링크드 데이터를 구축하기 위해 Protégé 4.2 Beta를 사용하였으며, 구축대상 자료는 dCollection의 다양한 자료 유형 중 가장 많이 구축되고 있는 학위논문으로 하였다. 또한 링크드 데이터를 구축하기 위해 참조할 수 있는 관련 사례를 조사하였다. 도서관에서 참조하거나 이용하거나 연계할 수 있는 도서관 링크드 데이터 세트는 대표적으로 W3C와 CKAN²²⁾에서 목록으로서 제공하고 있는 것을 알 수 있다. CKAN은 2012년 4월 20일 현재 56개의 목록을 제공하고 있고 W3C에서도 26개의 목록을 제공하고 있으며, 세 그룹으로 구분하고 있으며, 아래에서 자세히 살펴보고 하겠다.

본 연구에서는 도서관 링크드 데이터로 활용할 수 있는 사례들을 살펴봄에 있어 링크드 데이터 목록을 제공하는 사이트 중심으로 간단하게만 소개하고 있으며, 상세한 내용은 해당 사이트에서 매우 상세하게 발견할 수 있고, 원하는 데이터를 다운로드 받을 수도 있다. 게다가 최근에 수행된 연구들에서 이러한 사이트들을 상세하게 소개하고 있기 때문에 본 연구에서는 dCollection에서 사용될 수 있는 링크드 데이터들을 조사하여 분석하는데 중점을 두고 있다.

21) Protégé. <<http://protege.stanford.edu/download/registered.html>>.

22) CKAN은 데이터허브로서 2012년 4월 현재 3,414건의 데이터 세트를 제공하고 있으며, 도서관과 관련된 데이터 세트를 따로 분류하여 제공하고 있다.

1) W3C의 도서관 링크드 데이터

단순히 링크드 데이터로 구축된 사례는 링크드 데이터 클라우드²³⁾에서 볼 수 있듯이 데이터 세트의 수가 2011년 9월 현재 295건에 이르는 것을 알 수 있다. 그 중에서 주로 도서관에서 이용할 수 있는 링크드 데이터를 종합적으로 정리한 단체가 W3C도서관링크드데이터인큐베이터그룹(W3C Library Linked Data Incubator Group)이다. 시맨틱 웹 전문가들 중 특히 링크드 데이터 전문가들로 구성된 이 그룹은 웹상의 도서관 자원의 상호운용성을 향상시키기 위해 RDF와 같은 표준을 사용해서 구축된 도서관과 관련되고 도서관에서 활용 가능한 리스트를 1년 3개월 동안 개발하여 제시하였다. 도서관 링크드 데이터에 가장 관련이 깊은 데이터 세트 목록은 다음 <표 1>과 같다.

<표 1> Library Linked Data Set

데이터 세트	설 명	트리플 양
Bibbase	· 인터넷상에서 과학출판물을 용이하게 배포	200,000
British National Bibliography (BNB)	· BNB는 VIAF, LCSH, Lexvo, GeoNames, MARC의 국가 및 언어코드, Dewey.info, RDF Book Mashup을 포함한 외부자원을 링크하기 위해 링크드 데이터로 출판됨	84,961,180
Calames	· ABES에 의해 운영되는 아카이브와 매뉴스크립의 프랑스대학종합목록	2,000,000
Chronicle America	· 역사적 신문에 대한 정보에 접근하게 하고, 디지털화 된 신문페이지를 선정 · 현재 140,000건의 신문기사, 320만 페이지 수록	20,000,000
Cambridge University Library dataset #1	· COMET은 JISC가 지원한 Cambridge University Library와 CARET간 협력프로젝트의 결과	-
data.bnf.fr - Bibliothèque nationale de France	· 저작과 저자에 대한 웹페이지를 RDF로 구축한 것	6,330,000
Scottish Mountaineering Council Journals Issues 1-36	· <i>Scottish Mountaineering Club Journal</i> 의 1권부터 36권(1890-1901)까지의 디지털아카이브	15,801
CrossRef DOI Resolver	· Digital Object Identifiers (DOI)는 문헌 특히 학술출판물을 식별하기 위해 약 3,000개의 출판사가 사용하고 있는 영구적자료식별체계	1,000,000,000
Europeana Linked Open Data	· 링크드오픈데이터로 이용가능한 메타데이터를 만들기 위한 노력 · 350만 아이템에 대한 메타데이터 구축	117,000,000
Freebase	· 시맨틱 웹기반의 협력지식베이스	337,203,427
Hungarian National Library (NSZL) catalog	· LOD로 구축된 OPAC, 디지털도서관, corresponding authority데이터	19,300,000
Linked Periodicals Database	· CrossRef, Highwire Press, and the National Library of Medicine에 의해 제공되는 저널메타데이터를 수집해 놓은 Data Incubator로 구축된 데이터 세트	100,000,000
lobid. Index of libraries and related organisations	· lobid-organisations는 International Standard Identifier for Libraries and Related Organizations (ISIL)를 기반으로 하여 도서관에 대한 URI를 제공	635,464
lobid. Bibliographic Resources	· lobid-resources는 서지자원(책, 아티클, pdf등)에 대한 메타데이터에 대한 접근을 제공하는 서비스(현재 7백만레코드)	311,427,107
Medline	· Mdeline목록의 RDF표현 · 약 1,900만개의 아티클이 http://dx.doi.org/ (논문기사식별자) 및 http://crossref.org/ (저널식별자)에 링크되어 있음	1,016,652,223
NTNU special collections	· 디지털화된 역사 단행본(Norwegian University of Science and Technology (NTNU))	3,569
The Open Library	· 인터넷 아카이브의 공개프로젝트로서 모든 책에 대해 하나의 웹페이지 가짐 · 현재 개인 또는 대규모 목록자로부터 약 2,000만 레코드 수집, 보유	400,000,000
English Language Books listed in Printed Book Auction Catalogs from 17th Century Holland	· 학자 및 성직자가 소장했던 인쇄도서경매카탈로그에 게재된 영어버전의 도서들	400,000,000
ePrints3 Institutional Archive Collection (RKBExplorer)	· ePrints3 아카이브의 링크드 데이터 버전	8,417,840
ECS Southampton EPrints	· RKB 탐색기에서 제공하는 서비스와는 차별화된 ePrints서버에서 생산된 라이브 데이터	1,000,000

23) The Linking Open Data cloud diagram. <<http://lod-cloud.net>>.

8 한국도서관·정보학회지(제43권 제2호)

Sudoc bibliographic data	· Sudoc은 ABES의 의해 운영되는 프랑스대학종합목록, 약 천만개의 서지레코드 포함	350,000,000
Open Library data mirror in the Talis Platform	· Open Library의 데이터 미러사이트 · SPARQL endpoint와 OpenSearch (with RSS 1.0 output) 제공	541,396,055
theses.fr	· ABES가 운영하는 프랑스학위논문탐색사이트	300,000
Linked Data Service der Universitätsbibliothek Mannheim	· RDF로 출판되는 서지데이터	53,415,605
University of Sussex Reading Lists	· 대학의 독서리스트로 이용되는 자원의 링크드 데이터 버전	1,762,492
20th Century Press Archives	· 개인, 회사, 공장, 생상품, 그리고 경제관련 주제의 3천만건 이상의 문헌	12,000,000

또한 <표 2>와 <표 3>에서 보는 것은 도서관 링크드 데이터로 활용가능한 유용한 어휘시스템이나 체계 및 RDF 형태로 표현된 메타데이터 요소집합으로, 이를 잘 활용하면 도서관 데이터의 확장 가능성이 매우 높아질 수 있음을 알 수 있다.

<표 2> Value vocabularies

분 류		Vocabularies
출판된 것	Classification systems	Dewey Decimal Classification (DDC) Universal Decimal Classification (UDC)
	Subject headings/ subject authority files	Library of Congress Subject Headings (LCSH) Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié (RAMEAU) Schlagwortnormdatei (SWD) National Diet Library List of Subject Headings (NDLSH)
	Name authority data	Virtual International Authority File (VIAF) Getty Union List of Artist Names (ULAN) Library of Congress Name Authority File (LC/NAF) GeoNames
	Thesauri	STW Thesaurus for Economics AGROVOC Eurovoc Thesaurus for Graphic Materials (TGM)
	Other controlled vocabularies	DCMI Type Vocabulary MARC Code List of Relators (also in element sets) PRONOM Creative Commons (CC) License set Preservation vocabularies from LoC
	Additional sources	Wordnet Freebase (also in datasets) DBpedia
공식 또는 비공식적으로 출판 진행중인 것	Aquatic Sciences and Fisheries Abstracts (ASFA) thesaurus Fisheries Reference Metadata National Agricultural Library's Agriculture Thesaurus and Glossary Getty Art and Architecture Thesaurus (AAT) Medical Subject Headings (MeSH) Iconclass Getty Thesaurus of Geographic Names (TGN)	
기타 LLD에 적합한 것	New York Times subject headings MARC List for Countries MARC List for Languages MARC List for Geographic Areas	

〈표 3〉 Metadata Element Sets

Set Type	Metadata
RDF 어휘사전으로 출판된 것	Dublin Core and DCMI Metadata Terms Open Archives Initiative – Object Reuse and Exchange (OAI-ORE) Simple Knowledge Organization System (SKOS) SKOS eXtension for Labels (SKOS-XL) MARC Code List of Relators (also in value vocabularies) CIDOC Conceptual Reference Model (CRM) Dublin Core Collection Description vocabularies Functional Requirements for Bibliographic Records (FRBR) and related ontologies International Standard Bibliographic Description (ISBD) Metadata Authority Description Schema in RDF (MADS/RDF) Gemeinsame NormDatei (GND) vocabulary Friend of a Friend (FOAF) Vocabulary of Interlinked Datasets (VoID) Bibliographic Ontology (BIBO) Upper Mapping and Binding Exchange Layer (UMBEL) Vocabulary vCard Lexvo.org ontology Exchangeable Image File Format (EXIF) Open Provenance Model (OPM) Music Ontology Creative Commons Rights Expression Language (CC REL) A Citation Type Ontology (CiTO) Description Of A Project (DOAP) W3C Geo vocabulary Semantically-Interlinked Online Communities (SIOC) Schema.org vocabulary Open Graph W3C Ontology for Media Resources
RDF 사전으로 출판 진행중인 것	General International Standard Archival Description (ISAD(G)) Europeana Data Model (EDM) Encoded Archival Context – Corporate bodies, Persons, and Families (EAC-CPF) MARC21 Preservation Metadata: Implementation Strategies (PREMIS) Encoded Archival Description (EAD) and other archive-oriented element sets
RDF 사전이 사용될 수 없는 것	Categories for the Description of Works of Art (CDWA) EBU P/Meta Semantic Metadata Schema (P/META) SPECTRUM Metadata Object Description Schema (MODS) Text Encoding Initiative (TEI) Guidelines
RDF어휘사전이 없지만 도서관 분야에 적합한 기타의 것	Visual Resources Association Core Categories (VRA Core) Public Broadcasting Metadata Dictionary (PBCore)

2) 데이터 허브

데이터 허브(Data Hub)는 2012년 4월 현재 3,414건의 데이터 세트를 제공하고 있으며, 커뮤니티 기반 인터넷상의 목록집합을 제공한다. 이 사이트는 원하는 데이터 세트를 검색하거나 다운로드 받을 수 있도록 하고 있으며, 기본적인 시각화 도구들도 제공하고 있다.

또한 이 사이트는 오픈소스 데이터목록 소프트웨어인 CKAN(Open Knowledge Foundation에 의해 개발 및 운영)을 운영하고 있다. CKAN의 각각의 데이터 세트 내 레코드는 데이터에 대한 기술과 기타의 유용한 정보, 예를 들어 사용될 수 있는 포맷, 무료로 이용할 수 있는 자료의 소장처, 데이터 세트가 속하는 주제영역 등에 대한 정보를 포함하고 있다. CKAN은 인터넷의 데이터목록의 수를 늘리는데 크게 기여하고 있다.

데이터 허브는 위키피디아 스타일의 누구나 편집할 수 있는 오픈데이터목록이다. 영국 정부는 CKAN을 이용하여 data.gov.uk를 운영하고 있으며 현재 8,000건의 정부데이터세트를 목록으로 제공하고 있고, 그 외 유럽나라들의 정부공개데이터인 publicdata.eu, 세계 여러나라의 목록리스트인 datacatalogs.org 등도 CKAN에 의해 지원되고 있다.

이 사이트는 도서관과 관련된 사이트를 별도로 분류하여 제공하고 있으며, 서지데이터(Bibliographic Data)와 도서관 링크드 데이터(Library Linked Data)가 있다. 서지데이터 부분은 주로 오픈서지데이터원칙을 준수하는 오픈서지데이터들로 구성되어 있으며, 현재 72개의 데이터 세트가 제공되고 있다. 도서관 링크드 데이터 부분은 RDF형태로서 링크드 데이터로 도서관에서 활용, 연계할 수 있는 데이터 세트로 구성되며 56개의 데이터 세트를 제공하고 있다.

3) SIMILE

본 연구에서 링크드 데이터 구축대상으로 하고 있는 dCollection과 유사한 시스템은 SIMILE(Semantic Interoperability of Metadata and Information in unLike Environments)²⁴⁾이다. SIMILE 프로젝트는 MIT의 디지털 저장소인 DSpace²⁵⁾에 시맨틱 개념을 적용하고자 하는 것이며, 특히 전문 도구 영역에는 의미기반 디지털 자원의 생성과 연계, 온톨로지 구현 및 구축 등 RDF 온톨로지의 구축과 온톨로지 기반 정보 탐색을 가능하게 하는 것이 주 목적이다. 시스템 구성요소로 RDF 온톨로지 구축도구인 RDFizer, Screen Scrapers, Gadger, Referee 등이 있다. SIMILE은 digital assets, schemata/vocabularies/ontologies, metadata, and services 간의 상호운용성을 향상시키고자 하고 있으며, 특히 개인, 기관, 커뮤니티 등에 분산되어 있는 컬렉션의 연계를 통해 상호운용성을 향상시키고자 하고 있다.

24) SIMILE Homepage. <<http://simile.mit.edu>>.

25) DSpace는 HP와 MIT 도서관의 협력으로 2002년에 개발된 시스템으로서 다양한 전공영역을 가진 연구공동체 조직에서 생산하는 디지털형식의 지적생산을 수집, 보존, 배포할 수 있는 동적 저장소이다.

IV. 연구질문

본 연구에서는 dCollection을 링크드 데이터로 구축하기 위한 기초 연구로서, 현재 dCollection에 가장 많이 구축되어 있는 학위논문을 대상으로 링크드 데이터를 실험적으로 구축하고자 하였다. 이를 위해 dCollection의 데이터 구조 분석, dCollection이 구축대상으로 하고 있는 자원유형조사, dCollection에 연계가능한 자원 중 링크드오픈데이터로 구축되어 있는 사례 조사 등의 과정을 거쳤으며, 이 과정에서 연구자가 제기한 연구질문은 다음과 같다.

RQ 1: dCollection을 링크드 데이터로 구축하는 것이 가능한가?

RQ 2: dCollection의 데이터 필드에 연계 가능한 링크드 대상 데이터는 어느 정도 구축이 되어 있는가?

RQ 3: dCollection으로 연계시킴으로써 얻을 수 있는 효과는 어느 정도 인가?

본 연구에서는 dCollection의 링크드 데이터로의 실험적 구축 외에, 위의 연구질문에 대한 답을 찾아가는 과정에서 시스템적 접근과 시스템 구축을 통한 파급효과, 문제점 해결방안 등을 찾아보고자 하였다.

V. 연구설계 및 방법론

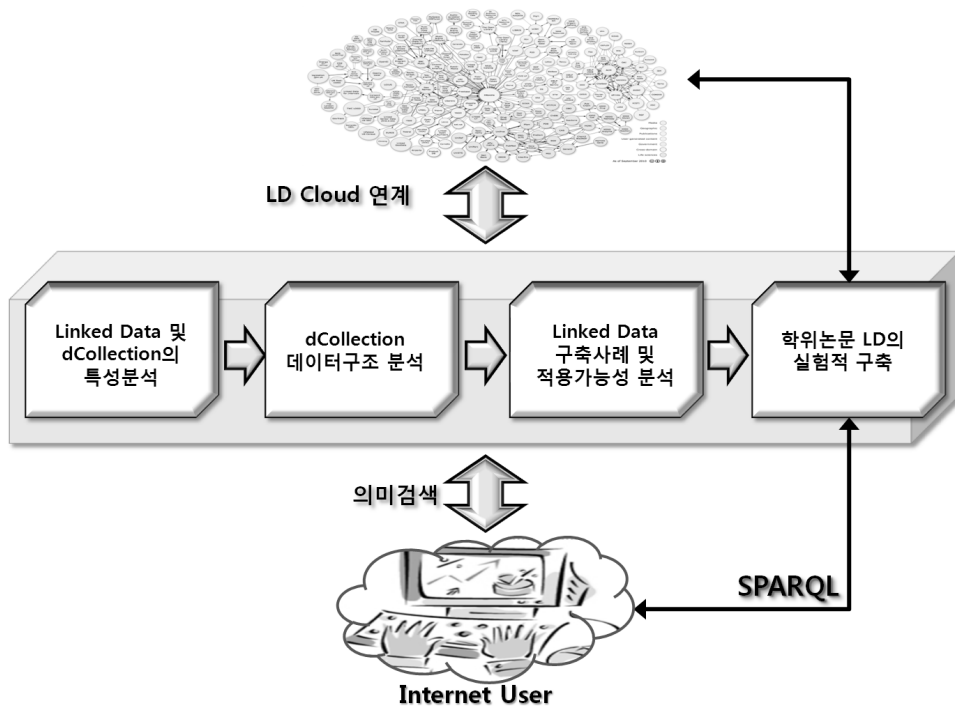
폭발적으로 증가하는 비구조적인 웹정보 및 기관생산자료의 활용성을 증대하기 위해 지식정보를 관리하는 세계 각국의 기관들은 RDF/OWL 온톨로지 기반의 시맨틱 웹 기술을 활용함으로써 데이터를 구조화하고 스마트화 하고 있으며, 이를 통해서 데이터 및 데이터베이스간의 연동성 및 확장성을 확보하고자 하고 있다. 연동성과 확장성은 링크드 데이터를 통해서 더욱 가속화되고 있으며, 전 세계적으로 3천 4백만건의 아이템을 포함하는 2,300여건의 기관리포지터리²⁶⁾를 링크드 데이터로 구축하는 것은 매우 의미 있다고 본다. 본 연구에서는 이를 실험적으로 구축하지만 적용 가능성은 전 세계적이라 할 수 있다.

실험적 링크드 데이터를 구축하기 위해 사용한 도구는 링크드 데이터 구축에 많이 사용되고 있는 Protégé 4.2 Beta이다.

본 연구에서는 dCollection을 링크드 데이터 구축대상으로 하고 있으며, 그 중 실험적 데이터 구축 대상 자료는 학위논문으로 하였다. 현재 dCollection 내에 구축되어 있는 자원의 유형 및 건수는 2012년 5월 14일 현재, 학위논문 1,563,990건, 학술논문 3,125,114건, 공개강의 172,965건, 기타자료

26) Repository66.org Repository Maps, <<http://maps.repository66.org/>>.

1,437,712건이다²⁷⁾. 이중 실험대상 자원은 학위논문으로 하고 있는데, 이는 학위논문이 기관생산물로 가장 높은 비율을 보이고 있고, 학술논문의 경우 DBPia 등 학술DB에서 제공하는 자료를 포함한 통계로서 실제로는 그다지 많지 않은 것으로 조사되었다. dCollection의 링크드 데이터로의 구축과정은 첫째, 구축대상 자원들의 특성을 분석한다. 둘째, dCollection의 데이터 구조를 분석하며, 셋째, Linked Data 구축사례 분석 및 dCollection 데이터의 링크드 데이터로의 구축가능성을 분석한다. 마지막으로 학위논문을 링크드 데이터로 실험적으로 구축한다. 이를 위해 기존에 구축된 도메인 온톨로지 및 도메인 사전을 참조하여 시맨틱 태깅을 수행하고, 온톨로지 변환 규칙을 적용하여 인스턴스화를 수행하여 트리플구조로 저장하며, RDF에 링크 가능한 관련 링크를 표현한다. 그 절차를 그림으로 표현하면 <그림 1>과 같다. 한편, 본 연구는 dCollection 데이터 중 학위논문만을 대상으로 실험적으로 링크드 데이터를 구축한 것이며, 다양한 유형의 데이터를 대상으로 링크드 데이터를 생성해야 할 것이다.



<그림 1> dCollection의 링크드 데이터 구축과정 및 활용

27) 건국대학교 글로컬캠퍼스 dcollection. <<http://kku.dcollection.net/jsp/index.jsp>>.

VI. dCollection용 링크드 데이터 실험적 구축

최근 기관리포지터리 구축의 방향은 연계와 확장을 들 수 있다. 즉 기존에 구축되어 있는 다양한 정보시스템과의 연계와 다양한 자원유형의 확장을 위해 노력하고 있다는 것이다.

첫째, 연구업적정보시스템과 연계를 하고 있는 Tokyo Tech STAR(Science and Technology Academic Repository)나 DSpace@ULB는 연구자가 제출한 원문을 연결하거나 저작권 등의 이유로 직접 연결이 어려운 경우 발행처에서 제공하는 원문 등으로의 링크를 제공하고 있다. 둘째, MIT는 DSpace 데이터 구축시 도서관목록이나 강의지원시스템의 강의자료 메타데이터와의 연계를 통해 자원이용의 효율성을 극대화하고자 하고 있다.

이들이 자원연계를 할 때 사용하는 방법은 구조화된 데이터 구조(예, MARC, Metadata)를 기반으로 하고 있다. 즉 MIT의 경우 DSpace에 데이터를 입력할 때 해당 자료가 도서관 목록에 있을 경우 MARC 레코드를 반입하는 방법, 그리고 강의자료 메타데이터가 있을 경우 크로스워크 테이블을 이용하여 강의자료를 DSpace에 연계한다. 링크드 데이터도 연계되는 대상이나 대상자원의 유형을 제한하지 않고 있기 때문에 기관리포지터리인 dCollection으로 구축하는 것이 용이할 것으로 보인다.

본 절에서는 dCollection의 데이터를 실험적으로 구축하기 위한 dCollection의 데이터 구조 분석, 클래스 정의 및 관계 정의, dCollection에 활용할 수 있는 도서관링크드 데이터 구축사례조사, 그리고 링크드 데이터의 실험적 구축결과를 제시하고자 한다.

1. dCollection의 자원유형

dCollection의 구축대상 자원유형은 해외와 비교해서 그다지 많지 않다. Lynch와 Lippincott²⁸⁾는 미국에서 기관 내 리포지터리로 구축하고자 하는 대상 자원은 40여 가지에 가깝다고 했으며, 학위논문, 학술지논문, 심사전논문, 기술보고서를 포함하여 다음 <표 4>와 같이 매우 다양함을 알 수 있다.

국내의 경우 KERIS의 dCollection에서도 현재는 주로 학위논문 및 연구소 발간자료 등을 중심으로 하고 있지만, SOMS의 'Type'을 보면, 학위논문, 심사전논문, 회의자료, 회의발표자료, 기술보고서, 연구보고서, 워킹페이퍼, 저서, 편저, 번역, 번역, 전자도서, 저널, 신문, 전자화된 특수컬렉션, 대학간행물, 대학전자기록, 학과자료, 학생논문, 서평 등의 자료를 포함할 예정임을 알 수 있다²⁹⁾. 또한 한국과학기술정보원(KISTI)은 과학기술 및 산업분야의 정보유통을 담당하고 있으며 구축대상

28) Lynch, C. A., & J. K. Lippincott, 2005. "Institutional repository development in the United States as of early 2005." D-Lib Magazine, 11(9). [cited 2012. 03. 15].

<<http://www.dlib.org/dlib/september05/lynch/09lynch.html>>.

29) 한국교육학술정보원. 2012. dCollecton. [cited 2012. 04. 13]. <<http://www.dcollection.net/>>.

〈표 4〉 미국 내 기관 레포지터리의 구축 대상 자원 유형

자원의 유형	구축중인 기관 수	구축 예정 기관 수	자원의 유형	구축중인 기관 수	구축 예정 기관 수
학위논문	21	15	전시	0	14
심사전문/e-print	24	9	공연	2	14
학술회의 회의록	14	14	인터뷰 필기록	1	14
학술회의 발표자료	15	13	지도	9	12
기술보고서, 진행 중 보고서	20	12	계획/청사진	1	10
e-Book	4	18	소프트웨어	1	10
저널	11	13	강의 자료	5	13
신문(디지털)	0	8	학습 객체	6	15
데이터 세트	4	26	학위논문 외 학생 논문	9	14
전자화한 도서관의 특수장서	19	13	E-포트폴리오	0	11
전자화한 박물관의 장서	3	6	기타		
대학 출판물	8	17	캠퍼스 블로그	1	0
대학의 전자기록물	2	13	신문	1	0
확과의 자료나 기록	6	17	실험 기록	1	0
디지털 이미지	19	15	전시회 가이드	1	0
디지털 오디오	10	19	도서 필사본	1	0
디지털 동영상	6	20	웹 페이지	1	0
디지털 악보	2	12			

출처: Lynch and Lippincott, 2005.

으로 국가 연구보고서, 학회지원문, 정보분석 보고서, 과학기술 기자재 정보, 인력정보 및 사실정보 등 각종 무료 포털사이트를 구축하여 운영하고 있다.³⁰⁾

2. dCollection의 데이터 구조 및 클래스 정의

dCollection이 채택하고 있는 데이터구조는 SOMS(Scholarly Object Metadata Set)로서 대상자원은 학위논문, 학술자원, 강의자료 및 기타 정보원이다. SOMS는 학술자원에 대한 일관성 있는 설명을 도출 수 있는 일반적인 명명법을 제공하는데 목적이 있으며, 학술분야의 기관, 단체가 추진하는 정보시스템 구축시 데이터의 표현, 프로그램 구현 등을 위한 지침으로 활용할 수 있다. SOMS 메타데이터의 세부적인 적용범위는 1) 학술정보를 생산, 저장, 관리하는 기관 및 단체, 2) 학술정보 수집을 통한 정보공유유통을 지원하는 기관 및 단체, 3) 학술정보 메타데이터의 교환을 원하는 기관 및 단체 등이다.

SOMS의 인용규격은 Dublin Core Metadata Element Set(Version 1.1)이며, 규격기술에 필요한 스킴은 ISO 639(1998 Code for the representation of names of languages), MIME types based on

30) 황혜경, 김혜선, 최선희, “오픈액세스기반 지식정보저장소 구축에 관한 연구,” 비블리아학회지, 제15집, 제1호(2004), pp.91-116.

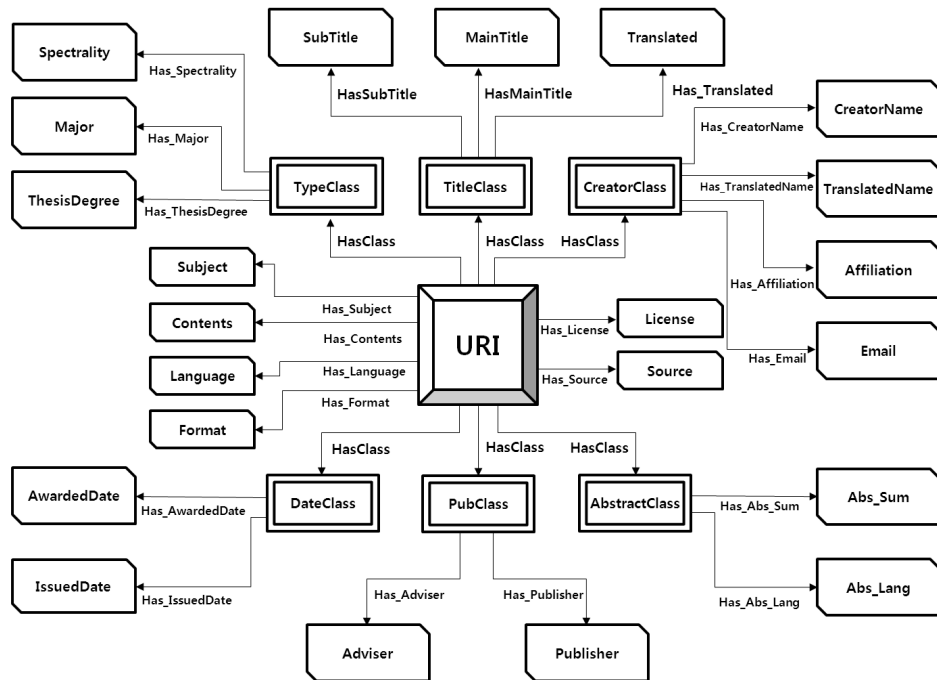
IANA registration, DCMI Type Vocabulary, ISO 8601(Date and Time)이다. SOMS 메타데이터 정보모델은 SOMS에 적합한 메타데이터를 기록하는 방법을 정의하는 데이터 요소이다.

SOMS는 더블링크어를 매우 기본적인 데이터 세트로 선택하고 있기 때문에 기존에 구축되어 있는 RDF 형식으로 전환하는 것이 매우 용이하며, 이후에 추가적으로 데이터를 구축하는 것도 매우 쉽기 때문에 공동 구축 및 공동활용의 효과는 매우 높을 것이다.

본 연구에서 RDF 구축대상으로 삼고 있는 학위논문의 클래스와 속성을 간단히 다음 <표 5>와 같이 정리하였으며, 이를 도식화하여 <그림 2>와 같이 표현하였다. 클래스와 속성 등의 관계는 현재 dCollection에서 취하고 있는 학위논문 입력지침을 참조하여 이를 가장 잘 표현할 수 있도록 하였다. 여기서 URI는 링크드 데이터로 구축이 되는 사이트 및 객체에 부여되는 URIs로서 Class와 객체를 갖는다. Predicate는 매우 다양하게 표현될 수 있으나 여기서는 최소한의 관계만 보여주고 있다.

<표 5> 주요 클래스와 속성(관계테이블)

Predicate / Property	Subject	Object	설명
HasClass	URI	TitleClass	Title 관련
HasMainTitle	TitleClass	MainTitle	본서명
HasSubTitle	TitleClass	SubTitle	부서명
Has_Translated	TitleClass	Translated	제2언어 서명
HasClass	URI	CreatorClass	Creator 관련
Has_CreatorName	CreatorClass	CreatorName	저자명
Has_TranslatedName	CreatorClass	TranslatedName	제2언어 저자명
Has_Affiliation	CreatorClass	Affiliation	소속기관
Has_Email	CreatorClass	Email	이메일
Has_Subject	URI	Subject	주제어
HasClass	URI	AbstractClass	초록 관련
Has_Abs_Sum	AbstractClass	Abs_Sum	초록, 요약, 개요
Has_Abs_Lang	AbstractClass	Abs_Lang	초록기술언어
Has_Contents	URI	Contents	목차
Has_Class	URI	PubClass	Publication 관련
Has_Publisher	PubClass	Publisher	발행기관
Has_Adviser	PubClass	Adviser	지도교수
Has_Class	URI	DateClass	Date 관련
Has_IssuedDate	DateClass	IssuedDate	발행년도
Has_AwardedDate	DateClass	AwardedDate	학위수여연월
Has_Class	URI	TypeClass	Type 관련
Has_ThesisDegree	TypeClass	ThesisDegree	학위명
Has_Major	TypeClass	Major	주전공
Has_Spectrality	TypeClass	Spectrality	세부전공
Has_Language	URI	Language	본문의 언어
Has_Format	URI	Format	페이지 수
Has_License	URI	License	공개정책



<그림 2> URI를 중심으로 해서 본 서지관계도

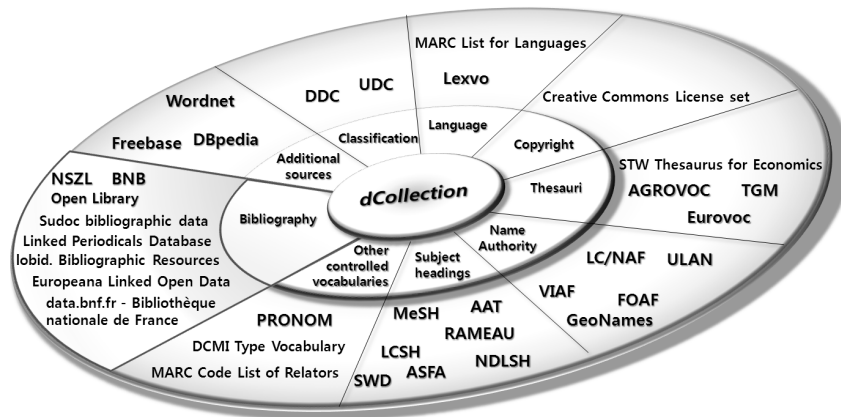
3. 링크드 데이터 구축 결과

본 연구에서는 링크드 데이터를 구축하기 위해 Protégé 4.2 Beta를 사용하였으며, 구축대상 자원은 앞 절에서 언급된 dCollection의 다양한 자원 유형 중 가장 많이 구축되고 있는 학위논문으로 하였다. 본 절에서는 dCollection의 메타데이터를 링크드오픈데이터로 구축하는 방법과 링크드 데이터로 연계될 수 있는 방안을 구체적으로 검토하였다.

학술자료메타데이터 세트(SOMS)는 학술분야 지식정보자원에 대한 시스템 구축, 자원의 공동활용 및 유통체제를 확립하기 위한 것으로 학술 관련 기관들이 학술정보자원을 구축하고자 할 때 지침으로 활용할 수 있도록 하고자 함이며, dCollection의 메타데이터로 사용되고 있다. SOMS의 인용규격은 Dublin Core Metadata Element Set이며, 17개의 데이터요소로 구성되어 있다. 가장 일반적인 기술대상 자원유형은 대학에서 생산되는 석·박사 학위논문, 대학 및 연구기관에서 생산되는 교내발간 학술지에 수록된 학술논문, 그리고 대학 및 연구기관에서 생산되는 연구보고서이며, 기타 프리프린트 자료, 회의보고서, 컨퍼런스발표 자료 등으로 추가되고 있다. 본 연구에서는 현재 가장 많은 건수로 구축되어 있는 순으로 학위논문을 대상으로 링크드 데이터를 실험적으로 구축하고 있다.

dCollection을 링크드 데이터로 구축함에 있어, 17개 데이터요소의 각각에 대한 연계 대상 링크드

데이터를 제시하고자 시도하였으며, 이는 샘플로 구축되는 것은 아주 일부에 불과하고 이 데이터의 확장을 위해 고려해야 하는 것이 매우 많기 때문이다. 구축대상자원의 URI는 <http://www.semanticweb.org/younghee/ontologies/2012/4/dCollection>이며, 본 연구에서 검토했던 dCollection 연계를 고려해 볼 수 있는 LD를 나열하였다.



〈그림 3〉 dCollection 연계 고려 대상 Linked Data

1) 저자 및 기관명 전거데이터

링크드 전거데이터는 RDF 형태의 링크드 데이터로 구조화시킨 전거데이터이며, 저자전거 및 기관명전거데이터가 적용될 수 있는 항목은 Creator, Advisor 항목이다. 대표적으로 Virtual International Authority File (VIAF), Getty Union List of Artist Names (ULAN), Library of Congress Name Authority File (LC/NAF), GeoNames³¹⁾ 등을 들 수 있다. 특히 VIAF는 OCLC를 중심으로 한 여러 국가도서관의 협력으로 구축된 개인명 및 단체명 전거파일이지만, 표제, 저작과 표현형에 대한 전거도 공개함으로써 전거파일의 유용성을 높이고자 하고 있다³²⁾.

또한 RDF로 출판되는 FOAF(Friend of a Friend)³³⁾는 사람에 대해, 그리고 사람 사이의 관계를 기술하는 온톨로지로서, Personal Profile을 통해서 성별, 생일, 위치, 취미, 이미지, 이메일, 온라인서비스 계정, 직장 및 학교 등에 대한 정보를, Relation(Knows)을 통해서 사람과 사람 사이의 관계를 보여준다. W3C에서는 RDF로 출판된 메타데이터요소세트로 분류하고 있으나, Creator나 Advisor, Publisher, Creator/Affiliation 정보를 링크드 데이터로 확장하는데 활용할 수 있다. 다음은 Creator Name에 대한 데이터생성 예이다.

31) GeoNames. <<http://www.geonames.org>>.

32) OCLC. 2012. "Virtual International Authority File." [cited 2012. 04. 13]. <<http://viaf.org>>.

33) FOAF-a-matic. <<http://www.ldodds.com/foaf/foaf-a-matic.ko.html>>.

```

-- CreatorClass/CreatorName(Younghee Noh와 관련된 전거데이터)
  <owl:NamedIndividual rdf:about="http://id.loc.gov/authorities/names/n2005066881">
    <rdf:type
rdf:resource="http://www.semanticweb.org/younghee/ontologies/2012/4/dCollection#CreatorName"/>
    </owl:NamedIndividual>
  <owl:NamedIndividual rdf:about="http://viaf.org/viaf/46173195">
    <rdf:type rdf:resource="http://www.semanticweb.org/younghee/ontologies/2012/4/dCollection#CreatorName"/>
  </owl:NamedIndividual>
  <owl:NamedIndividual rdf:about="http://viaf.org/viaf/46173195/rdf.xml">
    <rdf:type rdf:resource="http://www.semanticweb.org/younghee/ontologies/2012/4/dCollection#CreatorName"/>
  </owl:NamedIndividual>
  <owl:NamedIndividual rdf:about="http://viaf.org/viaf/46173195/#No__Yo%CC%86ng-hu%CC%86i__1969">
    <rdf:type rdf:resource="http://www.semanticweb.org/younghee/ontologies/2012/4/dCollection#CreatorName"/>
  </owl:NamedIndividual>
  <owl:NamedIndividual
rdf:about="http://www.semanticweb.org/younghee/ontologies/2012/4/dCollection#Younghee__Noh">
    <rdf:type rdf:resource="http://www.semanticweb.org/younghee/ontologies/2012/4/dCollection#TranslatedName"/>
  </owl:NamedIndividual>
  <owl:NamedIndividual rdf:about="http://www.semanticweb.org/younghee/ontologies/2012/4/dCollection#노영희">
    <rdf:type rdf:resource="http://www.semanticweb.org/younghee/ontologies/2012/4/dCollection#CreatorName"/>
  </owl:NamedIndividual>
  <owl:NamedIndividual rdf:about="http://www.worldcat.org/wcidentities/lccn-n2005-66881">
    <rdf:type rdf:resource="http://www.semanticweb.org/younghee/ontologies/2012/4/dCollection#CreatorName"/>
  </owl:NamedIndividual>

-- -- FOAF 데이터와 연계된 예 -- --
- <rdf:Description rdf:about="http://viaf.org/viaf/46173195">
- <!-- The primary entity (formerly a viaf:NameAuthorityCluster) -->
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person" />
  <rdf:type rdf:resource="http://rdvocab.info/uri/schema/FRBRentitiesRDA/Person" />
  <foaf:name>No, Yǒng-hŭi, 1969-</foaf:name>
  <foaf:name>노영희, 1969-</foaf:name>
  <foaf:name>노 영희, 1969-</foaf:name>
  <foaf:name>Noh, Younghee, 1969-</foaf:name>
  <rdaGr2:dateOfBirth>1969-11-25</rdaGr2:dateOfBirth>
  </rdf:Description>

```

2) 주제명 링크드 데이터

주제명 표목 및 주제명전거파일로서 링크드 데이터로 구축된 것은 Library of Congress Subject Headings (LCSH), Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié (RAMEAU), Schlagwortnormdatei (SWD), National Diet Library List of Subject Headings

(NDLSH) 등이며, 주제명 표목에 대한 자유로운 접근이 가능하고 다양한 포맷의 다양한 통제어휘를 연계하거나 다운 받아 활용할 수 있다. 이 링크드 데이터들은 Subject, Content, TitleClass, AbstractClass 정보를 링크드 데이터로 확장하는데 활용할 수 있다. 또한 현재 RDF 형태의 링크드 데이터로 출판중인 것으로 Getty Art and Architecture Thesaurus (AAT), Medical Subject Headings (MeSH), National Agricultural Library's Agriculture Thesaurus and Glossary, Aquatic Sciences and Fisheries Abstracts (ASFA), Thesaurus 등이 있다. 다음은 주제명 생성을 위해 LCSH를 검색하여 연계한 예이다.

```

<!-- 주제명: Information Retrieval => http://id.loc.gov/authorities/subjects/sh85066148 -->
<owl:NamedIndividual rdf:about="http://id.loc.gov/authorities/subjects/sh85066148">
  <rdf:type rdf:resource="http://www.semanticweb.org/younghee/ontologies/2012/4/dCollection#Subject"/>
</owl:NamedIndividual>
<owl:NamedIndividual rdf:about="http://id.ndl.go.jp/auth/ndlsh/00575010">
  <rdf:type rdf:resource="http://www.semanticweb.org/younghee/ontologies/2012/4/dCollection#Subject"/>
</owl:NamedIndividual>

<!-- 주제명: Semantic networks => http://id.loc.gov/authorities/subjects/sh92004914 -->
<owl:NamedIndividual rdf:about="http://id.loc.gov/authorities/subjects/sh92004914">
  <rdf:type rdf:resource="http://www.semanticweb.org/younghee/ontologies/2012/4/dCollection#Subject"/>
</owl:NamedIndividual>

<!-- 주제명: Knowledge => http://id.loc.gov/authorities/subjects/sh2002011410 -->
<owl:NamedIndividual rdf:about="http://id.loc.gov/authorities/subjects/sh2002011410">
  <rdf:type rdf:resource="http://www.semanticweb.org/younghee/ontologies/2012/4/dCollection#Subject"/>
</owl:NamedIndividual>

```

3) 언어

각국의 언어데이터가 링크드 데이터로 구축된 사례로 MARC List for Languages (<http://id.loc.gov/vocabulary/languages.html>)가 있으며, 이는 ISOs 639-1, 639-2, and 639-5를 적절히 참조하고 있다. 이 링크드 데이터는 Language 정보를 다음 예시와 같이 링크드 데이터로 확장하는데 활용할 수 있다. 다음은 Language를 링크드 데이터로 확장한 예이며, 실제 데이터 값을 입력하지 않고 'kor'에 대해 링크드 데이터 URI 값으로 연계한 예이다.

```

<owl:NamedIndividual rdf:about="http://www.lexvo.org/page/iso639-3/kor">
  <rdf:type rdf:resource="http://www.semanticweb.org/younghee/ontologies/2012/4/dCollection#Abs_Lang"/>
  <rdf:type rdf:resource="http://www.semanticweb.org/younghee/ontologies/2012/4/dCollection#Language"/>
</owl:NamedIndividual>

```

4) 저작권

dCollection에 등록된 다양한 학술자원의 저작권과 관련된 권리사항을 CCL(Creative Commons License)을 통해 지원하고 있다. 또한 dCollection에 학위논문 및 학술논문 등록 시 CCL이 적용된 자료는 dCollection을 통해 국내외 연계서비스를 제공하고 있으며 해당 대학 및 RISS에서 서비스 제공 시 저작권 정보를 안내하고 있다. 논문 제출자가 직접 권리정보를 설정할 수 있는 기능을 제공함으로써 논문제출자의 권리를 지원하고 있다. 다음은 저작권 권리관계 정보를 링크드 데이터로 적용해 본 예이다.

```

<rdf:Description rdf:about="http://web.resource.org/cc/permits">
  <owl:equivalentProperty rdf:resource="http://creativecommons.org/publicdomain/zero/1.0/" />
</rdf:Description>

```

5) 원문정보 등 기타

본 논문에서는 dCollection에 있는 모든 필드에 대한 링크드 데이터로의 확장가능성을 보기 위해 모든 필드에 대해 검토하였으며, 페이지정보 등을 제외하고 외부데이터의 링크주소를 통해 확장을 함으로써 서지데이터를 매우 풍부하게 관리할 수 있을 것으로 확인하였다. 특히 원문정보로의 링크는 저작권 문제 때문에 원문을 직접 소장할 수 없는 경우에 링크를 제공함으로써 이용자를 원문으로 연결해 줄 수 있다.

```

-- -- 자료유형정보(Type) -- --
<rdf:type rdf:resource="http://purl.org/ontology/bibo/Dissertation"/>

-- -- 원문정보(Resource) -- --
<owl:NamedIndividual rdf:about="http://naver.nanet.go.kr/SearchDetailView.do?cn=KDMT1200004264">
  <rdf:type rdf:resource="http://www.semanticweb.org/younghee/ontologies/2012/4/dCollection#Source"/>
</owl:NamedIndividual>

```

5) 서지데이터

서지데이터의 링크드 데이터로의 확장을 위해 완전한 서지데이터를 링크드 데이터로 구축하고 있는 사례는 많이 있다. 선행연구를 포함한 W3C에서 제시하고 있는 서지데이터를 참조할 수 있으며, 이러한 서지데이터로의 연계활용 가능성은 박지영 연구(2012)³⁴)에서 실험적으로 검증된 바 있다. 그는 대표적으로 BNB와 Open Library를 실험대상으로 하였다. BNB는 8천여 개의 트리플이 있는 영국국가도서관 서지데이터이고, Open Library는 인터넷아카이브 프로젝트의 서지데이터이다. 또한 스웨덴 국가도서관의 서지데이터로 LIBRIS는 세계 최초로 발행된 국가서지로 LCSH, DBpedia 등으로 다양한 링크를 제공하고 있다. 그 외 data.bnf.fr - Bibliothèque nationale de France, Europeana Linked Open Data, Linked Periodicals Database, Hungarian National Library (NSZL) catalog, lobid, Bibliographic Resources, Sudoc bibliographic data 등이 있다.

VII. 결론 및 제언

기관리포지터리와 링크드 데이터는 자원을 공동 활용하자는 정신에 있어 매우 밀접한 관련이 있다고 할 수 있다. 따라서 이 개념을 연계시킴으로써 새로운 접근방법에 의한 정보공동생산과 활용이 가능할 수 있을 것으로 본다.

이에 본 연구에서는 국내 dCollection시스템을 링크드 데이터로 실험적으로 구축하고자 하였다. 이를 위해 링크드 데이터로 구축할 dCollection의 데이터 구조 분석, 링크드 데이터로 구축된 도서관에 적용가능한 링크드 데이터로의 구축사례 조사, dCollection과 같은 기관리포지터리 중 이미 링크드 데이터로 구축된 사례 조사, dCollection의 메타데이터를 링크드오픈데이터로 구축하는 방법과 링크드 데이터로 연계될 수 있는 방안 검토를 수행하였다. 그 결과를 간단히 정리하면 다음과 같다.

첫째, 오픈 액세스 운동과 링크드 데이터의 개발은 자원을 공동으로 구축하고 공동으로 활용함으로써 지식정보유통을 활성화하자는 정신에 있어서 공통점이 있다고 할 수 있다. 따라서 공통의 개념을 전제로 한 두 원리를 융합할 경우 그 효과가 증폭될 수 있을 것이라 생각한다.

둘째, 기관리포지터리를 활성화할 수 있는 방안에 대한 연구는 많이 수행되어 있고, 세계적으로 기관리포지터리가 1,500 건에 이르고 있지만, 이를 링크드 데이터로 연계하기 위해 데이터 구조를 RDF로 전환한 사례는 DSpace2 뿐인 것으로 조사되었다.

셋째, 도서관에서 데이터의 확장을 위해 활용 가능한 링크드 데이터에 대한 W3C의 조사³⁵)에 따

34) 박지영, "링크드데이터 방식을 통한 서지 정보의 확장에 관한 연구," 정보관리학회지, 제29집, 제1호(2012), pp.231-251.

35) W3C, 2011. "Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets." [cited 2012. 04. 13]. <<http://www.w3.org/2005/Incubator/ld/XGR-ld-vocabdataset-20111025/>>.

르면, 2011년 현재 100여건의 데이터 세트가 이미 링크드 데이터로 구축되어 있거나 구축중이며, RDF 구조로 변환하고 있다고 하였다. 따라서 이러한 데이터를 집중적으로 분석하여 국내외 도서관 데이터를 확장하는데 활용할 경우 세계적인 도서관 네트워크가 구축될 수 있다고 본다. 또한 오픈소스 데이터 목록 소프트웨어인 CKAN이 운영하는 데이터 허브에서도 RDF형태의 링크드 데이터로서 도서관에서 활용할 수 있는 56개의 데이터 세트에 대한 목록을 제공하고 있다.

넷째, 링크드 데이터 전용 구축도구는 비구조적 또는 구조적 데이터를 RDF형태로 생산하는 것이며, 링크드 데이터를 저작하거나 편집할 수 있는 도구로 이미 Loomp, Tabulator, mSpace Data Picker, D2RQ, OntoBase 2.0 등이 있으며, 기타 연구자들에 의해 실험적으로 구축된 링크드 데이터 구축도구도 있다. 이러한 도구들을 활용하여 연구자는 물론 일반 이용자들도 매우 쉽게 링크드 데이터를 생산할 수 있을 것으로 본다.

다섯째, 본 연구에서는 dCollection의 학위논문을 대상으로 링크드 데이터를 실험적으로 구축하였으며, 이를 위해 W3C의 도서관 링크드 데이터를 집중적으로 분석하였다. 그 결과 SOMS 기반의 dCollection 메타데이터 개개 항목을 연계할 수 있는 데이터 세트가 상당히 많이 있는 것으로 분석되었다. 즉, 5장의 <그림 3>에서 보는 바와 같이 저자명 및 기관명 전거데이터 확장에 사용할 수 있는 VIAF, ULAN, LC/NAF, GeoNames 등이 있고, 주제명 전거데이터로 활용될 수 있는 LCSH, RAMEAU, SWD, NDLSH 등이 있다. 그 외 언어관련 링크드 데이터로 Lexvo, MARC List for Languages, 저작권 관련으로 Creative Commons License set가 있다. 그리고 이미 링크드 데이터로 구축된 도서관 서지데이터로 BNB와 Open Library, data.bnf.fr, NSZL, Linked Periodicals Database 등이 있다. 즉, 이미 도서관 서지데이터의 링크드 데이터로의 구축활동은 매우 활발하게 이루어지고 있음을 알 수 있다.

여섯째, 해외 사례에서 보았듯이 해외의 경우 도서관에서 활용 가능한 링크드 데이터들이 100여건에 이르고 있음에도 불구하고 국내의 경우 연구단계에 있는 것이 현실이다. 따라서 본 연구에서는 dCollection의 링크드 데이터로의 연계 가능성을 해외에서 개발된 사례로 제시할 수밖에 없었다. 그러나 링크드 데이터에 대한 연구가 활성화 되고 많은 기관에서 자관의 데이터를 링크드 데이터로 개발하여 개방한다면, 그 활용효과는 매우 높아질 것으로 보인다.

종합적으로 보았을 때, dCollection을 링크드 데이터로 구축함으로써 서지데이터를 매우 풍부하게 구축하는 것이 가능하고, dCollection의 데이터 필드에 연계 가능한 링크드 대상 데이터는 매우 많다는 것을 알 수 있었다. 이렇게 dCollection을 링크드 데이터로 연계시킴으로써 얻을 수 있는 효과는 첫째, 개개의 도서관의 데이터가 연계되어 전 세계적인 도서관 데이터 세트가 구축되는 결과를 낳을 수 있다. 둘째, 웹 포털이나 인터넷 서점 등에 구축된 자원으로의 연계도 가능하기 때문에 풍부한 도서관의 서지데이터를 이용자에게 제공할 수 있을 것으로 보인다. 그 외 오삼균, 김성훈, 장원홍³⁶⁾

36) 오삼균, 김성훈, 장원홍, “도서관 링크데이터(Library Linked Data) 현황분석과 통합 LLD 설계 및 확산 방향에

의 연구에서는 W3C의 보고서에 제시된 링크드 데이터를 도서관에 도입할 때의 장점을 이용자 측면의 혜택, 도서관 조직 측면의 혜택, 도서관 협업 종사자 측면의 혜택, 도서관 개발자 측면의 혜택으로 나누어 상세히 기술하고 있다.

마지막으로 dCollection을 링크드오픈데이터로 구축할 경우 고려해야 할 사항을 보면, 첫째, 기존에 비구조화된 형태로 구축된 이질적인 기관리포지터리를 변환하는 문제가 있다. 즉 Non-RDF 데이터(TopicMap, XML, Spreadsheet 등)를 RDF 형식으로 변환해 주어야 링크드 데이터 세트로 연계될 수 있다. 개개의 비구조화된 데이터를 RDF로 변환하는 것도 쉽지 않은 일이다. 현재 비상업적인 솔루션으로 베를린 대학의 D2RQ가 기존의 Non-RDF 데이터를 RDF 형식으로 변환해 주는 매핑 언어이며, 오픈소스로 제공되고 있다. 국내의 경우 (주)탐쿼드란트코리아가 이와 유사한 기술을 보유하고 있으나 상업적인 솔루션이다.

둘째, 기관리포지터리를 링크드 데이터로 오픈함에 있어 언급하지 않을 수 없는 것이 저작권 및 라이선스문제이다. 기관리포지터리는 원래 기관의 내부생산물을 DB로 구축해서 공동활용하는 것으로서, 국내 연구로 오픈엑세스 기반 기관리포지터리 시스템에서 적용할 수 있는 라이선스를 범주화한 후 라이선스 개념을 도입한 관리모델을 제시한 연구가 있다³⁷⁾. 그러나 자원구축 및 이용에 있어 국내 학회, 출판사, 연구기관 및 대학 내 기관의 저작권 이양 관계, 저자나 학회 및 출판사가 셀프아카이빙시의 라이선스 문제 등을 링크드 데이터로 연계할 경우에는 다시 한번 검토할 필요가 있다고 본다.

참고문헌

- (주)탐쿼드란트코리아, 2012. OntoBase 2.0. [cited 2012. 04. 13].
 <<http://ontobase.topquadrant.co.kr/>>.
- 김현희 외. 국가지식정보 유통체제 마련을 위한 평가모형 및 표준 업무 모델 개발 연구, 서울 : 한국교육학술정보원, 2005.
- 김현희 외. 국가지식정보 유통체제 구축을 위한 dCollection 활용실태·효과분석 연구, 서울 : 한국교육학술정보원, 2006.
- 김현희, 정경희, 김용호. “사용성 평가를 통한 dCollection 시스템 개선 방안 연구,” 한국도서관·정보학회지, 제37집, 제4호(2006), pp.327-350.

대한 고찰,” 한국도서관·정보학회지, 제42집, 제4호(2011), pp.331-351.

37) 이나니, 김선미, 이미화. 오픈엑세스 환경에서의 학술·연구저작물 서비스를 위한 라이선스 관리 방안(KERIS 이슈리포트), 서울 : 한국교육학술정보원, 2004.

- 노영희. “도서관 3.0의 개념과 서비스 모형에 관한 연구,” 정보관리학회지, 제27집, 제4호(2010), pp.283-307.
- 박미성. “오픈엑세스를 위한 저작권관리시스템 사례연구를 통한 dCollection 라이선스관리시스템 분석,” 한국도서관·정보학회지, 제36집, 제4호(2005), pp.255-284.
- 박연희. “디지털 자원으로서는 학위논문 이용 변화에 관한 연구,” 제18회 한국정보관리학회 학술대회 논문집, 2010, pp.23-28.
- 박지영. “정보공유를 위한 Linked Data의 실험적 구축,” 2011년도 한국정보관리학회 추계학술대회, 2011년 11월 11일.
- 박지영. “링크드데이터 방식을 통한 서지 정보의 확장에 관한 연구,” 정보관리학회지, 제29집, 제1호(2012), pp.231-251.
- 오삼균, 김성훈, 장선홍. “도서관 링크데이터(Library Linked Data) 현황분석과 통합 LLD 설계 및 확산 방향에 대한 고찰,” 한국도서관·정보학회지, 제42집, 제4호(2011), pp.331-351.
- 이나니, 김선미, 이미화. 오픈엑세스 환경에서의 학술·연구저작물 서비스를 위한 라이선스 관리 방안 (KERIS 이슈리포트). 서울 : 한국교육학술정보원, 2004.
- 장금연. “학술정보유통을 위한 레포지터리 적용 방안 연구.” 한국문헌정보학회지, 제38집, 제4호(2004), pp.291-310.
- 정연경, 이나미, 이미화. “기관 레포지터리 활성화를 위한 메타데이터의 확장 방안 연구.” 한국문헌정보학회지, 제41집, 제1호(2007), pp.323-344.
- 정효숙, 김희진, 박성빈. “일반인을 위한 링크드 데이터 생성시스템 개발 및 활용.” 컴퓨터교육학회 논문지, 제14집, 제2호(2011), pp.47-59.
- 조명대. “도서관에서의 Linked Data 활용방안에 관한 연구.” 한국문헌정보학회지, 제44집, 제1호(2010), pp.181-198.
- 조재인. “일본의 전개 양상을 통해서 본 한국 기관 레포지토리의 과제,” 정보관리학회지, 제26집, 제1호(2009), pp.35-55.
- 한국교육학술정보원. dCollection, 2012. <<http://www.dcollection.net/>> [cited 2012. 04. 13].
- 황혜경, 김혜선, 최선희. “오픈엑세스기반 지식정보저장소 구축에 관한 연구.” 비블리아학회지, 제15집, 제1호(2004), pp.91-116.
- Berners-Lee, T., et al. “Tabulator: Exploring and Analyzing linked data on the Semantic Web.” Proceedings of the The 3rd International Semantic Web User Interaction Workshop(SWUI06) workshop, 2006.
- CKAN. “Library Linked Data.” 2012. <<http://thedatahub.org/group/ld>> [cited 2012. 04. 13].
- Luczak-Rosch, M. & R. Heese. “Linked Data Authoring for Non-Experts.” Proceedings of

- the 2009 Linked Data on the Web Workshop (LDOW09), 2009.
- Lynch, C. A., & J. K. Lippincott. 2005. "Institutional repository development in the United States as of early 2005." *D-Lib Magazine*, Vol.11, No.9(2005).
〈<http://www.dlib.org/dlib/september05/lynch/09lynch.html>〉 [cited 2012. 03. 15].
- Mark Ware Consulting. "Pathfinder Research on Web-based Repositories: Final Report." Publisher and Library/Learning Solutions, 2004. 〈<http://www.palsgroup.org.uk/>〉 [cited 2012. 03. 16].
- OCLC. "Virtual International Authority File." 2012. 〈<http://viaf.org>〉 [cited 2012. 04. 13].
- Smith, D., I. Popov & M. Schraefel. "Data Picking Linked Data: Enabling Users to create Faceted Browsers." *Proceedings of the 2010 Web Science Conference*, 2010.
- W3C. "Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets." 2011. 〈<http://www.w3.org/2005/Incubator/ld/XGR-ld-vocabdataset-20111025/>〉 [cited 2012. 04. 13].
- D2RQ. 〈<http://d2rq.org/>〉.
- FAST Linked Data. 〈<http://experimental.worldcat.org/fast>〉.
- FOAF-a-matic. 〈<http://www.ldodds.com/foaf/foaf-a-matic.ko.html>〉.
- GeoNames. 〈<http://www.geonames.org>〉.
- Ontotext . 〈<http://www.ontotext.com/>〉.
- Protégé. 〈<http://protege.stanford.edu/download/registered.html>〉.
- SIMILE Homepage. 〈<http://simile.mit.edu>〉.