

DCAT을 활용한 디지털도서관 데이터셋 관리와 서비스 설계

Designing Dataset Management and Service System for Digital Libraries Using DCAT

박진호 (Jin Ho Park)*

목 차

- | | |
|---------|------------------------|
| 1. 서론 | 4. DCAT 기반 디지털도서관 데이터셋 |
| 2. 선행연구 | 서비스 설계 |
| 3. DCAT | 5. 결론 |

초 록

본 연구는 새로운 지식정보자원으로 중요성이 높아지고 있는 데이터셋을 관리·서비스하기 위해 W3C 표준인 DCAT 활용방안을 제시하고자 하였다. 이를 위해 먼저 DCAT을 구성하는 8개의 클래스 중 핵심 클래스 4가지를 중심으로 클래스와 속성을 분석하였다. 또한 디지털도서관에서 DCAT을 기반으로 다양한 데이터셋을 관리·서비스할 수 있는 시스템을 모델링하여 제시하였다. 이 시스템은 원천데이터, 데이터셋 관리, 링크드 데이터 연결, 이용자 서비스로 구분하여 구성하였으며 특히 데이터셋관리에서는 DCAT 매핑 기능을 제시하여 다양한 데이터셋 서비스와 상호운용성 확보가 가능하도록 하였다.

ABSTRACT

The purpose of this study is to propose a W3C standard, DCAT, to manage and service dataset that is becoming increasingly important as new knowledge information resources. To do this, we first analyzed the class and properties of the four core classes of DCAT. In addition, I modeled and presented a system that can manage and service various data sets based on DCAT in digital library. The system is divided into source data, data set management, linked data connection, and user service. Especially, the DCAT mapping function is suggested in dataset management. This feature can ensure interoperability of various datasets.

키워드: DCAT, 데이터셋 검색, 데이터셋 관리, 데이터셋 목록

DCAT, Dataset Retrieval, Dataset Management, Dataset Catalog

* 주식회사 리스트 사업개발본부장(jinhopark.lis@gmail.com / ISNI 0000 0004 7641 0372)

논문접수일자: 2019년 4월 22일 최초심사일자: 2019년 5월 9일 게재확정일자: 2019년 5월 16일

한국문헌정보학회지, 53(2): 247-266, 2019. [http://dx.doi.org/10.4275/KSLIS.2019.53.2.247]

1. 서론

새로운 형식을 가진 지식정보자원의 출현은 이를 관리, 서비스하는 기관에 커다란 변화를 요구한다. 변화가 필요한 부분은 새로운 자원을 해당 자원의 속성에 맞게 관리, 서비스, 보존하는 전 과정에 해당한다. 여기에 해당하는 대표적인 기관은 도서관을 포함한 기억기관(memory institute)이다. 이 범주를 좀 더 확장하면 구글(Google), 야후(Yahoo)와 같은 검색 엔진, 검색 서비스들도 해당한다고 볼 수 있다. 현대 정보화 사회에서 이런 새로운 유형의 지식정보자원 중 하나가 데이터셋(dataset)이며, 데이터셋은 새로운 관리, 서비스 대상으로 중요성이 증대하고 있다. 일반적으로 데이터셋은 데이터의 집합을 의미한다. 여기서 데이터 집합이란 하나의 데이터베이스 테이블 혹은 특정 변수에 대응하는 컬럼(column)을 포함하는 통계 처리가 가능한 데이터 매트릭스(data matrix)를 말한다. 학술적 관점에서 데이터셋은 특정 실험, 연구 진행, 결과와 관련한 특정 데이터의 집합을 의미한다.

이런 데이터셋에 대한 관심은 빅 데이터(big data)의 중요성, 오픈 데이터(open data)의 확산, 정부 데이터(Government Data), 연구 데이터(research data)의 개방과 탐색에 대한 요구가 늘어나면서 본격적으로 대두하기 시작했다. 또한 웹이라는 문서(document) 중심 정보 생태계가 웹 3.0(Web 3.0) 혹은 시맨틱 웹(semantic web)을 목표로 하는 데이터 중심의 웹(web of data)으로 변화하면서 데이터셋에 대한 필요성이 범용적으로 확대하고 있다. 이제 데이터는 단순한 정보의 참고자료를 넘어서

유용한 지식정보를 담고 있는 활용성 높은 정보자원으로 중요성이 날로 증가하고 있다.

이처럼 정부기관, 연구기관, 도서관과 같은 기관들이 데이터셋을 개방하기 시작하면서 산재해 있는 데이터셋들을 검색하고 활용할 수 있는 방식에 대한 고민을 시작하기에 이르렀다. 고민의 결과물은 데이터셋 검색을 지원하는 다양한 사례들에서 확인할 수 있다. 먼저 공적 영역에서 보면 미국(data.gov), 영국(data.gov.uk), 한국(data.go.kr) 등 정부기관들이 운영하고 있는 정부 데이터 포털 검색과 Science.gov와 같이 미국 연방정부가 과학, 엔지니어링, 연구개발 과제와 해당 데이터를 함께 제공하고 있다. 연구 데이터 영역으로 보면 DATA CITATION INDEX(WEB OF SCIENCE), DataCite(search.datacite.org/ui), ICPSR(www.icpsr.umd.edu/icpsrweb/ICPSR/) 등 전통적인 데이터 제공 서비스가 상당하다. 민간 영역으로 넘어가면 가장 대표적인 결과물이 구글(Google)의 데이터셋(dataset) 검색(google dataset search beta, <https://toolbox.google.com/datasetsearch>)이다. Noy(2018)에 따르면 구글 데이터셋 검색은 학술 검색(Google Scholar)과 같은 전문 데이터 검색 서비스이다. 이 서비스는 데이터 과학자, 분석가, 연구자들을 대상으로 전 세계에 흩어져있는 정부, 공공 데이터를 쉽게 검색해서 활용할 수 있도록 해준다. 초기 구글의 목표 데이터셋은 각 국 정부가 발행한 공공 데이터 서비스 사이트와 연구 데이터 사이트였고 현재는 디지털도서관, 개인 홈페이지까지 대상을 확대하고 있다.

지금까지 살펴본 것처럼 데이터셋은 새로운 지식정보자원 유형으로 중요성이 날로 증대하

고 있다. 더불어 데이터셋 관리, 서비스에 필요한 정책, 운영 플랫폼(platform), 메타데이터 개발 등 다양한 연구와 노력이 이어지고 있다. 이런 연구에 있어서 핵심 중 하나는 데이터셋 관리와 서비스를 위한 표준 메타데이터의 활용에 있다. 앞서 예로 들었던 구글의 경우도 데이터셋 관리와 서비스를 위한 표준 메타데이터로 schema.org 용어집을 기본으로 사용하고 있으며 가이드라인을 발행해 데이터 제공자들의 준수해야하는 방법을 안내하고 있다. 또한 유럽 연합(European Union)의 데이터 포털인 European Open Data Portal(<https://www.europeandataportal.eu/en/homepage>)은 유럽 각국의 데이터를 수집하고 관리하기 위해 DCAT(Data Catalog Vocabulary)을 기반으로 개발한 DCAT-AP(Application Profile)을 채택하고 관련 메타데이터를 표준으로 데이터를 등록, 관리, 서비스하고 있다.

지식정보자원을 관리하는 기관 입장에서 새로운 유형의 지식정보자원을 다루어야할 때 가장 필요한 요소 중 하나는 메타데이터로 볼 수 있다. 전자자원, 디지털자원이 출현하고 양적, 질적으로 중요한 비중을 차지하면서 현재의 도서관은 디지털도서관을 중심으로 한 이용자 서비스와 디지털정보자원 관리가 핵심적인 업무로 자리 잡았다. 이 중 디지털정보자원의 생애주기에 적합한 메타데이터를 개발하는 것이 도서관이 갖는 중요한 과업 중 하나였다. 최근 일반적으로 사용하고 있는 기술용 메타데이터 DC(Dublin Core), MODS(Metadata Object Description Schema), 보존을 위한 PREMIS(PREservation Metadata: Implementation Strategies) 데이터 사전들이 그 결과물이라 할

수 있다. 이제 데이터셋이 도서관, 특히 디지털도서관에서 관리하고 이용자에게 서비스해야 하는 새로운 유형의 디지털 정보자원으로 등장했다. 도서관은 이미 디지털도서관이라는 서비스 플랫폼을 가지고 있으므로 새로운 유형의 서비스를 개척하기 보다는 이제 출현하고 확산되고 있는 데이터셋에 적합한 메타데이터의 선택과 운용이 필요한 때이다. 특히 웹에 흩어져 있는 다양한 데이터셋을 수집하고 효율적으로 서비스하기 위해서는 데이터셋의 상호운용성을 보장해 줄 수 있는 표준이 필요하다.

이에 본 연구에서는 W3C(world wide web consortium)에서 제정하고 오픈 소스 형태의 다양한 데이터 관리 플랫폼과 서비스에서 채용하고 있는 DCAT을 분석하여 디지털도서관에서 데이터셋을 어떻게 관리하고 서비스에 적용할 수 있을지에 대해 다루었다. 단, 본 연구에서는 하나의 디지털도서관 솔루션이나 특정 기관의 디지털도서관 관리·서비스 시스템을 특정하지는 않았다. 즉 가시적으로 성과를 확인할 수 있는 개발보다는 DCAT의 표준 용어집과 오픈 데이터셋을 입수하여 처리하고 서비스하는 부분에 중점을 두었다는 점을 한계로 밝힌다. 또한 본 연구에서 언급하는 디지털도서관은 일반 이용자 외 연구자, 개발자, 기관과 같이 실제 데이터를 이용할 수 있는 이용자까지 포괄할 수 있는 국가 도서관 단위와 학술 데이터와 같이 구체적인 데이터셋 서비스가 가능한 학술연구 정보 검색 서비스를 온라인으로 제공하는 기관의 서비스로 한정한다. 결과는 향후 디지털도서관에서 데이터셋 검색 시스템을 구축하고자 할 때 활용될 수 있을 것이다.

2. 선행연구

DCAT은 비교적 최근의 연구주제로 대상이 데이터셋이라는 점과 형식적으로 RDF로 표현하고 용어집으로서 특정 객체를 기술하는데 사용한다는 점을 고려하여 다각적으로 선행연구를 검토하였다. 먼저 RDF로 데이터를 개방하는 대표적인 방식인 LOD(Linked Open Data)의 관점에서 보면 윤소영(2013)은 국가차원의 LOD 사업추진으로 활발하게 진행하고 있는 링크드 데이터 사업들의 결과를 분석하여 기존의 연계 방식이 아닌 새로운 형식의 데이터베이스 연계가 필요함을 언급했다. 이 연계 체계의 핵심으로 링크드 데이터 기술을 제안한 바 있다. 도서관과 관련한 링크드 데이터 활용 측면에서 조명대(2010)는 도서관 LOD를 기존의 서지모델인 FRBR, RDA와 함께 활용할 수 있는 방안을 제시하여 도서관 데이터의 시맨틱 레벨을 올릴 수 있는 모델을 제시한 바 있다. 또한 Hallo, Luján-Mora, Trujillo Mondéjar(2014)는 대학 도서관이 링크드 데이터 기술을 활용하여 도서관 데이터를 RDF 형식으로 발행함으로써 다른 도서관 데이터와 쉽게 연결 가능하여 결국 자관의 데이터 품질을 향상시킬 수 있고 이용자들은 보다 폭넓은 지식을 획득할 수 있음을 주장했다.

연구 데이터 측면에서의 연구들을 보면 노영희(2012)는 기관 저장소와 링크드 데이터가 자원을 공동으로 활용하고자한다는 점에서 유사한 점이 있음을 주장하고 국내 dCollection 데이터 구조를 분석하고 자원유형을 조사한 후 이 중 링크드 데이터로 구축한 사례를 조사하였다. 조사결과를 바탕으로 dCollection의 메타데이

터를 링크드 데이터 형식으로 구축하여 외부 데이터와 연계할 수 있는 방안을 제시한 바 있다. 김지현(2015)은 데이터 관리와 공유에 대한 대학 연구자들의 인식을 조사하는 연구에서 연구자들이 데이터의 보존과 개방의 중요성을 인식하고 있지만 연구아이디어의 도용, 표절, 논문 출판의 주도권 문제 등 부정적인 면을 우려하는 측면도 있음을 언급하면서 적절한 유인책의 필요성을 언급했다. 심원식, 안혜연, 변제연(2015) 인문학 분야 연구데이터의 수집 및 활용성 증진을 위한 전략 연구에서 정책과 인프라의 부재로 연구 데이터의 관리와 공유가 어려운데 특히 인문사회분야는 저장소 자체가 취약함을 지적했다. 그리고 이 문제의 해결을 위해 국내 인문사회 분야 연구데이터 및 자료의 대표적인 리포지토리인 한국연구재단의 기초학문자료센터를 중심으로 인문학 분야 연구데이터의 체계적인 수집, 활용 및 보존을 위한 프로세스 및 전략을 제시한 바 있다. 김지현(2012)의 연구는 대학내 연구자들의 연구데이터에 관한 것으로 연구자들이 연구데이터를 취득하는 방식이 저장소 등의 서비스가 아니라 개인적인 연락 등으로 이루어짐을 지적하고 시스템적 대책이 필요함을 지적하였다.

데이터셋에 대한 메타데이터 도구 개발의 필요성에 대한 연구들을 살펴보면 Alexander et al.(2011)은 RDF 데이터셋을 표현하기 위해 VoID(Vocabulary of Interlinked Datasets)를 사용을 제시하였다. VoID는 RDF 데이터셋에 대한 메타데이터를 표현하기 위한 RDF Schema 어휘집이다. 이 연구에서는 VoID의 사용이 데이터 게시자와 이용자 사이의 교두보 역할을 담당할 수 있음과 데이터셋에 대한 메타데이터

의 필요성을 강조하였다. Guptill(1999)은 대규모 공간데이터에 대한 관리와 이용자 제공을 위해 데이터셋의 속성을 반영한 효과적인 메타데이터 도구 개발이 필요함을 언급한 바 있다. 의학 분야에서도 데이터셋 메타데이터 개발 필요성을 지적한 사례가 있다. ZHANG, JIANG and LIU(2009)는 이기종 시스템 특히 데이터셋의 상호운용성 확보를 위한 메타데이터의 필요성을 언급한 바 있다.

도서관에서 새로운 지식정보자원을 관리하기 위한 새로운 메타데이터에 대한 연구를 살펴보면 이승민(2015)은 디지털 정보의 보존을 위한 연구에서 아카이빙의 핵심적인 역할을 수행할 수 있는 보존용 메타데이터 패키지를 4개 상위요소, 25개 하위요소로 구성하여 제시한 바 있다. 마찬가지로 보존이라는 새로운 업무영역을 위해 박옥남(2012)은 PREMIS 데이터 사전에 활용한 국립중앙도서관의 디지털정보자원을 효율적으로 관리보존할 수 있는 방안을 제시하였는데, 단순한 메타데이터 요소 도출을 넘어서 METS를 기반으로 한 목록 예까지 제시한 바 있다. 이 외에도 다양한 연구들이 다양한 정보 객체를 대상으로 한 메타데이터 연구를 진행하였다. 고영만, 배경재(2011)는 한국연구재단의 토대기초지원사업 연구성과물을 대상으로 한 연구를 진행한 바 있다. 또한 뉴스분야 적용을 위한 메타데이터 연구(송인석, 이성숙 2005), 메타데이터 자체 연구동향 파악(유사라 2010), 정책 정보 메타데이터(이명희 2019), 교육정보시스템의 학습객체를 대상(이병기 2004) 등 다양하다.

본 연구에서는 이형 데이터 구조를 가지는 데이터베이스의 통합 운영 가능성을 검토하고 사

례 연구를 통한 실제적인 방안을 설계하기 위해서 한국연구재단 토대기초지원사업의 DB 구축 연구성과물과 한국연구재단의 기초학문자료센터 DB 통합 운영 방안을 제시하였다. 토대기초연구지원사업의 성과확산을 위해서는 기초학문자료센터와의 통합 연계가 필수적으로 필요하며, 그 방안으로서 기존에 구축된 DB구축 과제의 DB는 표준 지침을 활용한 XML 데이터베이스화를 필수적으로 진행하되, 향후 구축될 DB구축 과제의 DB는 기초학문자료센터와 연계 시스템을 구성하거나 독립 시스템을 구성하는 방안을 제안하였다.

마지막으로 데이터셋 목록에 대한 연구를 살펴보면 Ding, Peristeras, Hausenblas(2012)는 지금까지의 정부 데이터를 개방실태 조사결과 표준 목록, 표준 어휘에 대한 필요성이 절실함을 지적하고 DCAT과 같은 표준을 활용하여 데이터셋을 관리하고 연계함으로써 이용자 경험을 향상시킬 필요가 있음을 주장한 바 있다.

기존 연구들을 종합해 보면 데이터셋의 개방과 공유에 대한 인식은 많이 형성되어 있지만 이를 뒷받침해 줄 기술적 인프라는 부족한 것을 확인할 수 있다. 더불어 데이터를 개방할 때 기계가독형의 표준형식을 준수하기 위한 노력은 이루어지고 있는 반면 양적으로 증가한 데이터셋을 체계적으로 관리하고 이용자가 효율적으로 검색할 수 있도록 해주는 목록화와 표준메타데이터에 대한 논의는 부족한 것 또한 확인할 수 있다. 특히 데이터셋을 대상으로 한 표준 메타데이터에 대한 연구는 사례를 국내·외적으로 찾아보기 어렵다. 따라서 표준적인 성격을 강하게 갖고 있는 DCAT을 기반으로 한 데이터셋 관리와 서비스 시스템 구성을 위한

요소분석과 구축 방안에 대한 논의가 필요한 시점이다.

3. DCAT

DCAT은 Data Catalog Vocabulary의 축약 표현으로 W3C에서 2014년 1월 16일에 권고(recommendation) 승인을 한 웹 표준이다. DCAT은 웹에 발행된 데이터 목록 간의 상호 운용성을 확보하고 활용성을 높이기 위해 설계한 RDF(Resource Description Framework) 어휘이다. 또한 여러 데이터셋 검색 사이트 간의 데이터 분산 저장과 공유 환경을 지원할 수 있도록 설계되었다.

사실 DCAT은 아일랜드의 시맨틱 기술 연구소인 DERI(Digital Enterprise Research Institute)에서 개발하였던 것을 W3C의 정부 링크드 데이터 작업반(Government Linked Data Working Group)에서 가져와 권장 표준으로 제정하였다. 애초에 유럽연합(EU, European Union)의 지원으로 진행된 결과물로 보는 것이 타당할 것이다. 이 후에도 DCAT은 유럽 집행위원회(European Commission)의 ISA 프로그램(Interoperability Solutions for European Public Administrations Programme)에서 채택하여 유럽 공공 부문 열린 데이터셋을 기술하기 위한 표준으로 활용하고 있다(DCAT Application Profile Working Group 2013).

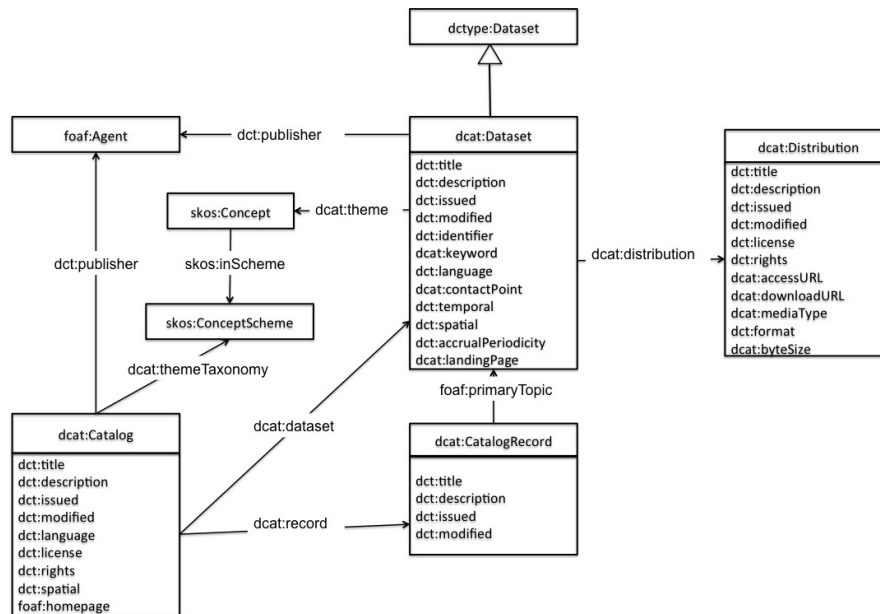
이렇듯 W3C 권고 표준과 유럽에서의 활용성을 기반으로 DCAT에 대한 활용은 점점 늘어나고 있다. DCAT을 응용하여 사용하고 있는 사례가 데이터 검색과 활용이 가능한 오픈

데이터 플랫폼들이다. 대표적인 플랫폼이 CKAN(<https://ckan.org/>), DKAN(<https://getdkan.org/>)이다. 국내의 경우도 공공데이터 포털(data.go.kr)에서 CKAN 연동을 위한 노력을 지속하고 있다(한국데이터진흥원 2017). 이런 확산은 DCAT이 갖고 있는 유연성과 확장성에 기인한다. 구체적으로 보면 DCAT은 데이터셋이라는 새로운 지식정보자원을 관리하기 위한 클래스를 새로 만들기는 했지만 대부분의 속성은 기존의 표준 용어집을 준수하고 있다.

3.1 DCAT 구조와 구성요소

DCAT을 구성하는 클래스(class)는 8개이다. <그림 1>은 DCAT의 전체 구조를 쉽게 이해할 수 있는 데이터 모델로 핵심 클래스는 Catalog, Dataset, CatalogRecord, Distribution이다. 각 속성을 구성하는 요소들은 기존의 표준 용어집으로 Dublincore, FoaF, SKOS이 주를 이루는 것을 확인할 수 있다.

DCAT 표준에서는 8개 클래스 중 Catalog, Dataset, Distribution을 주 클래스(main class)로 설정하고, CatalogRecord를 중요 클래스(important class)로 구분하여 제시하였다. 주 클래스, 중요 클래스가 데이터셋 자체에 관한 기술을 담당하고 있다면 그 외 클래스는 외부 데이터와 결합할 수 있는 다리 역할을 수행할 수 있도록 되어있다. 또한 클래스와 클래스를 연결하는데 사용하고 있는 속성들(예, dct:publisher)들은 양쪽 클래스에 모두 존재하며 각 클래스들을 연결하는 역할을 담당한다. 주 클래스, 중요 클래스, 그 외 클래스가 의미하는 바를 순차적으로 살펴보면 다음과 같다.



〈그림 1〉 DCAT 데이터 모델

(출처: Maali, Erickson, Archer 2014)

3.1.1 Catalog 클래스

Catalog 클래스는 데이터셋을 설명하는데 적합한 메타데이터의 집합으로 CatalogRecord 클래스와 Dataset클래스를 참조한다. 이 클래스에 속하는 속성에 대한 상세설명은 〈표 1〉과 같다. 아래 표를 포함하여 이 후 본 논문에서 속성명을 표시할 때는 해당 속성이 어떤 표준 용어집에서 왔는지 구별하기 쉽도록 해당 용어집의 네임스페이스(namespace)를 함께 병기하였다. 앞서 언급한 것처럼 DCAT은 RDF로 표현하여 발행하기 때문에 모든 클래스는 RDF Class, 모든 속성은 RDF Property에 속한다는 사실을 전제하고 기술한다.

3.1.2 Dataset 클래스

Dataset 클래스는 특정 행위자(발행자, 게시

자 등)가 발행하여 이용자가 직접 접근해서 다운로드 할 수 있는 데이터셋을 의미한다. 이 클래스에 속하는 속성에 대한 상세설명은 〈표 2〉와 같다.

3.1.3 Distribution 클래스

Distribution 클래스는 이용자들이 활용할 수 있는 특정 형식의 데이터셋을 의미한다. 각각의 데이터셋은 CSV, API, RSS 등 서로 다른 형식으로 제공될 수 있기 때문에 이를 함께 표현한다. 하지만 이 클래스는 활용 가능한 정보 제공이 목적이며 실제 다운로드 등 행위 등은 dcat:downloadURL 속성을 사용해서 표현한다. 이 클래스에 속하는 속성에 대한 상세설명은 〈표 3〉과 같다.

〈표 1〉 DCAT Catalog 클래스에 속하는 속성

속성명	정의	전역(domain)	치역(range)
dct:title	Catalog에 부여한 이름		rdfs:Literal
dct:description	Catalog에 관한 자유로운 문자형 설명		rdfs:Literal
dct:issued	Catalog의 공식 발급, 발행 날짜		rdfs:Literal ISO8601을 활용하여 입력한 문자열
dct:modified	Catalog를 변경, 갱신, 수정한 최근 날짜		rdfs:Literal ISO8601을 활용하여 입력한 문자열
dct:language	Catalog에서 사용한 용어로 목록의 대상 데이터셋의 제목, 설명 등 메타데이터에 사용한 언어를 의미함		dct:LinguisticSystem 미국의회도서관의 코드표 즉, ISO639-1과 ISO639-2를 활용하여 입력한 문자열
foaf:homepage	Catalog의 홈페이지		foaf:Document homepage와 Document는 역참조 (inverse functional property) 관계 로 URI를 정확하게 기술해야함
dct:publisher	온라인 상에 Catalog를 만들어서 배포한 책임있는 기관 foaf:Agent 유형의 값을 사용하는 것을 권고함		
dct:spatial	Catalog에서 다루는 지리적 영역		dct:Location
dcat:themeTaxonomy	Catalog의 데이터셋 주제를 구분		dcat:Catalog
dct:license	Catalog와 관련한 Dataset이 아니라 사용 에 관한 라이선스 문서로 연결 해당 Catalog의 라이선스가 포함하는 모 든 Dataset과 Distribution에 적용되는 경 우 각 클래스에도 복제하여 표기		dct:LicenseDocument
dct:rights	Dataset이 아니라 Catalog의 사용/재사 용에 관한 권한 설명 해당 Catalog의 권한 적용을 받는 모든 Dataset과 Distribution클래스 에도 복제 하여 표기		dct:RightsStatement
dcat:dataset	Catalog의 일부인 Dataset	dcat:Catalog	dcat:Dataset
dcat:record	Catalog의 일부인 Catalog Record	dcat:Catalog	dcat:CatalogRecord

〈표 2〉 DCAT Dataset 클래스에 속하는 속성

속성명	정의	전역(domain)	치역(range)
dct:title	Dataset에 부여한 이름		rdfs:Literal
dct:description	Dataset에 관한 자유로운 문자형 설명		rdfs:Literal
dct:issued	Dataset이 공식 발급, 발행 날짜로 반드시 최초 발행일을 입력해야함		rdfs:Literal ISO8601을 활용하여 입력한 문자열
dct:modified	Dataset을 변경, 갱신, 수정한 최근 날짜 로 Catalog Record의 변경이 아니라 실제 Dataset을 변경한 날짜를 입력해야하고 이용자는 이 정보로 Dataset이 지속적으 로 업데이트되고 있음을 확인할 수 있음		rdfs:Literal ISO8601을 활용하여 입력한 문자열

속성명	정의	전역(domain)	치역(range)
dct:language	Dataset의 언어 Catalog에서 사용한 언어와 충돌이 일어날 수 있는데 이 경우는 Catalog 언어 속성의 값을 대체하여 사용함 여러 언어로 제공하는 경우 복수로 속성을 사용할 수 있으며 Distribution의 dct:language 속성을 사용하여 각각을 설명할 수 있음.		dct:LinguisticSystem 미국의회도서관의 코드표 즉, ISO639-1과 ISO639-2를 활용하여 입력한 문자열
dct:publisher	Dataset을 사용할 수 있도록 만드는 책임을 갖는 개체 foaf:Agent 유형의 값을 사용하는 것을 권고함		
dct:accrualPeriodicity	Dataset을 발행, 게시하는 빈도		dct:Frequency
dct:identifier	Dataset의 고유 식별자 식별자는 Dataset의 URI로 사용할 수 있으며 명시적 표현도 가능함		rdfs:Literal
dct:spatial	Dataset이 다루는 지리적 영역		dct:Location
dct:temporal	Dataset이 다루는 일정 기간 시작날짜와 종료날짜로 명명		dct:PeriodOfTime
dcat:theme	Dataset의 주요 카테고리 하나의 Dataset은 다양한 주제를 가질 수 있음. dct:subject의 하위속성	dcat:Dataset	skos:Concept
dcat:keyword	Dataset을 설명하는 주제어 혹은 태그(tag)	dcat:Dataset	rdfs:Literal
dcat:contactPoint	VCard를 사용하여 관련 연락처 정보와 Dataset을 연결	dcat:Dataset	vcad:Kind
dcat:distribution	데이터셋에 연결할 수 있는 Distribution들에 연결	dcat:Dataset	dcat:Distribution
dcat:landingPage	Dataset, Distribution 혹은 추가 정보 획득이 가능한 웹 페이지 직접 Dataset을 다운로드 할 수 있는 URL을 제공할 수 없는 경우 반드시 해당 페이지에 이 URL 정보를 포함해야 함	dcat:Dataset	foaf:Document

〈표 3〉 DCAT Distribution 클래스에 속하는 속성

속성명	정의	전역(domain)	치역(range)
dct:title	Distribution에 부여한 이름		rdfs:Literal
dct:description	Distribution에 관한 자유로운 문자형 설명		rdfs:Literal
dct:issued	Distribution이 공식 발급, 발행 날짜로 반드시 최초 발행일을 입력해야함의 공식 발급 날짜		rdfs:Literal ISO8601을 활용하여 입력한 문자열
dct:modified	Distribution을 변경, 갱신, 수정한 최근 날짜		rdfs:Literal ISO8601을 활용하여 입력한 문자열
dct:license	Distribution 관련 라이선스 문서로 연결		dct:LicenseDocument

dct:rights	Distribution 관련 권리에 관한 정보		dct:RightsStatement
dcat:accessURL	Distribution에 대한 접근이 가능한 랜딩 페이지(landing page), feed, SPARQL 점점 등을 표현 직접적인 다운로드가 불가능할 경우 반드시 downloadURL이 아니라 accessURL을 사용	dcat:Distribution	rdfs:Resource
dcat:downloadURL	제공 형식 드래로 데이터셋을 배포하고 있는 파일의 위치	dcat:Distribution	rdfs:Resource
dcat:byteSize	바이트(byte) 단위로 크기를 표현	dcat:Distribution	rdfs:Literal (xsd:decimal 형식으로 표현)
dcat:mediaType	IANA에서 정의한 미디어의 유형	dcat:Distribution	dct:MediaTypeOrExtent
dct:format	Distribution의 파일 형식(format)		dct:MediaTypeOrExtent

3.1.4 CatalogRecord 클래스

Catalog Record 클래스는 하나의 데이터셋을 기술하기 위한 데이터 목록의 레코드를 의미한다. 이 클래스는 중요 클래스로 분류하지만 주 클래스는 아니며 실제 입력 시스템과 같은 환경에서 선택 사항이다. 이 클래스에 속하는 속성에 대한 상세설명은 <표 4>와 같다.

지금까지 살펴본 클래스 외에 DCAT에서 정의하고 있는 클래스는 dctype:Dataset, foaf:Agent, skos:Concept, skos:ConceptScheme 이 있다. 이들 클래스는 DCAT에서 정의한 것이 아니고 이미 표준으로 널리 활용하고 있는

dublincore, skos, foaf의 클래스를 사용하고 있다.

4. DCAT 기반 디지털도서관 데이터셋 서비스 설계

디지털도서관에서 DCAT을 활용하는 방안은 크게 두 가지를 생각해볼 수 있다. 첫 번째는 CKAN, DKAN과 같은 오픈 소스 데이터셋 관리 솔루션을 적용하여 별도의 데이터셋 관리·서비스 시스템을 구축하는 것이다. 두 번째는 ‘3. DCAT’에서 분석한 용어들을 활용하여 수

<표 4> DCAT Catalog Record 클래스에 속하는 속성

속성명	정의	전역(domain)	치역(range)
dct:title	Record에 부여한 이름		rdfs:Literal
dct:description	Record에 관한 자유로운 문자형 설명		rdfs:Literal
dct:issued	Catalog에 해당 데이터셋을 목록화한 날짜		rdfs:Literal ISO8601을 활용하여 입력한 문자열
dct:modified	Catalog 항목을 변경, 갱신, 수정한 날짜 데이터셋 자체의 날짜가 아니라 카달로그 항목에 관한 날짜		rdfs:Literal ISO8601을 활용하여 입력한 문자열
foaf:primaryTopic	해당 Catalog Record를 Dataset 자원에 연결		

집한 데이터셋을 정리하고 이용자에게 서비스 하는 방식으로 현재의 디지털 도서관 관리·서비스 인프라에 DCAT을 처리할 수 있는 기능을 구현하는 것으로 달성 가능하다.

전자의 경우는 가장 손쉽게 데이터셋 관리·서비스 인프라를 갖출 수 있는 방법으로 볼 수 있다. 이 방법이 갖는 장점으로는 이미 동일한 오픈소스를 활용하고 있는 세계의 정부, 공공 데이터, 민간 데이터 서비스에서 손쉽게 데이터 입수가 가능하다는 점이다. 반면 단점은 기존의 디지털도서관 관리시스템과 이원화된 관리체계를 유지해야한다는 점으로 도입비용은 적을 수 있지만 유지관리 비용에 상당할 것으로 예상할 수 있다. 또한 이용자들에게 도서관의 다른 정보자원들과 통합 환경에서 검색 구현이 불가능하다는 점도 단점으로 생각해볼 수 있다.

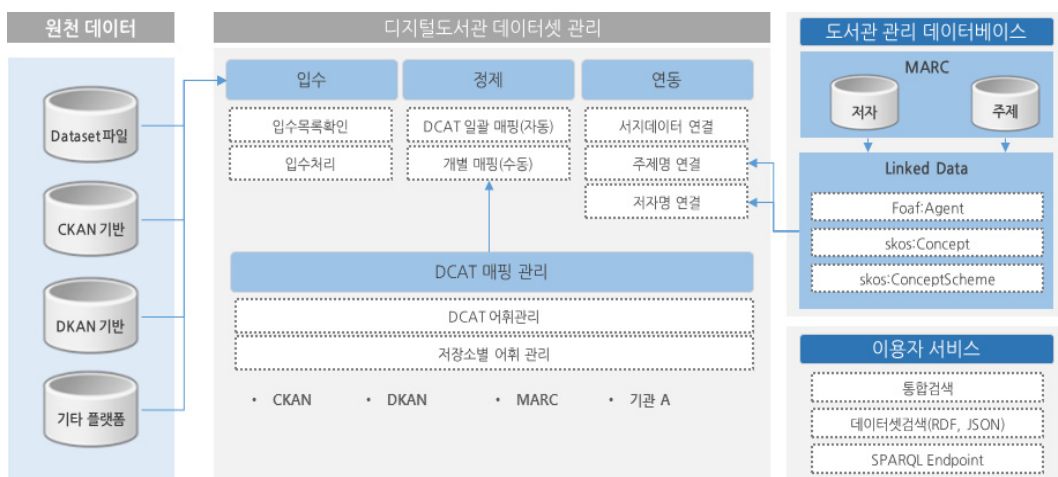
후자의 경우 전자와 반대로 초기 구축비용이 상당할 수 있다는 장점이 있지만 DCAT 표준 용어 처리가 가능한 시스템을 구축함으로써 외

부 데이터의 입수가 용이하다는 점에서 나은 선택일 수 있다. 무엇보다 데이터셋을 기존의 도서관 자원과 연계하여 관리할 수 있고 연관된 다양한 지식정보자원과 통합검색 및 연계 검색을 지원해줄 수 있다는 점을 가장 큰 장점으로 볼 수 있다. 때문에 본 논문에서는 두 번째 방식인 디지털도서관 시스템에 DCAT을 처리할 수 있는 모듈을 추가 구축할 수 있는 방법에 대해 기술하고자 하며 특히 DCAT 용어집을 활용한 외부 데이터 수집과 기존 데이터와의 연결 방식에 대해 집중한다.

4.1 시스템 구성

DCAT 기반 디지털도서관 시스템의 핵심요소는 다양한 원천데이터를 입수, 정제 후 기존 도서관 데이터베이스와 연결하는 것이다. 전체 시스템 구성도 <그림 2>와 같다.

시스템의 핵심은 디지털도서관 데이터셋 관리 시스템으로 외부적으로는 다양한 유형의 원



<그림 2> 디지털도서관의 DCAT 관리·서비스 시스템 구성도

원데이터 입수와 이용자 서비스를 포괄한다. 내부적으로는 내부적으로는 DCAT 표준 어휘에 맞게 오픈소스 데이터 플랫폼의 어휘와 입수 대상 특정 기관의 어휘를 간리하고 매핑 정보를 정제 시 활용할 수 있도록 하는 기능을 포함한다. 특히 DCAT이 RDF 형식이라는 점을 감안하여 데이터셋의 품질을 높이기 위한 링크드 데이터와의 인터링킹 기능으로 구성하였다. 이용자서비스는 기존의 디지털도서관 검색서비스와 함께 통합검색을 기본으로 지원한다. 데이터셋을 활용하고자 하는 이용자들을 위해서는 데이터셋 검색과 SPARQL Endpoint 기능까지 제공하는 것이 이용자 서비스의 기능이다. 각 구성요소별 핵심 고려사항과 기능사항은 요소는 다음과 같이 정리할 수 있다.

4.2 시스템 구성요소별 고려사항

4.2.1 원천데이터

원천데이터의 대상은 디지털도서관에서 직접 수집이 가능한 개별 데이터셋, CKAN, DKAN과 같은 널리 활용하고 있는 데이터셋 관리 프레임워크와 개별적으로 개발하여 운영중인 플랫폼 모두를 포괄하였다. 한국데이터진흥원(2017)에 따르면 전 세계 데이터 포털을 운영하는 국가 27개 중 21개는 오픈 소스 플랫폼 CKAN을 사용하고 있고 DKAN 역시 꾸준한 증가세를 보이고 있다. 한국과 노르웨이의 경우, 별도 플랫폼을 운영하고 있지만 CKAN의 일부 기능을 활용하고 있으며 특히 노르웨이의 경우 DCAT을 활용하여 데이터 연계하고 있다. 국가별 오픈 데이터 플랫폼 현황은 <표 5>와 같으며, 지역별 데이터 플랫폼을 구분하면 <표 6>과 같다.

<표 5>에서 보는 것처럼 국가 데이터 포털의 경우 CKAN의 도입 비율이 앞도적으로 높은 것을 확인할 수 있다. 또한 <표 6>에서 보는 것처럼 지역별로 국가 데이터 포털 외 다양한 민간 데이터 서비스까지 포함해도 CKAN의 채용 비율은 앞도적이다. 단, 오픈 소스 적용 외에 기타(자체개발 등)의 비중이 높은 것도 중요하게 고려해야한다.

따라서 원천데이터 입수 대상 선정에 있어서 가장 중요한 것은 CKAN, DKAN과 같은 오픈 소스 플랫폼을 반드시 고려해야한다는 점과 자체적으로 개발한 플랫폼의 데이터도 중요하게 고려해야 한다는 점이다. 자연스럽게 원천데이터의 입수 유형이 다양하다는 점은 디지털도서관의 데이터셋 관리시스템 내에 다양한 메타데이터를 매핑할 수 있는 기능의 필요성을 유도한다.

4.2.2 데이터셋 관리

데이터셋 관리는 입수, 정제, 연동, DCAT 매핑 관리를 핵심 기능으로 구성하였다. 입수는 다양한 원천데이터를 불러와 목록을 확인하고 정제 단계로 데이터를 다시 이관하는 기능을 담당한다. 정제는 가장 핵심적인 기능으로 DCAT 매핑관리와 긴밀한 관계를 갖고 작동한다. DCAT 매핑관리는 표준 용어인 DCAT 어휘관리 기능과 저장소별 어휘관리 기능으로 구성하였다. DCAT 어휘관리기는 표준 용어인 DCAT의 버전을 관리하고 해당 클래스, 속성 정보의 변화를 관리하는 역할을 담당한다. 저장소별 어휘관리는 CKAN, DKAN 등 잘 알려진 데이터 프레임워크의 어휘를 별도로 관리하여 DCAT 표준과 매핑 관리를 할 수 있는 기능

〈표 5〉 국가별 오픈 데이터 플랫폼 사용 현황

No	국가	명칭	플랫폼	No	국가	명칭	플랫폼
1	미국	Us Government Open Data	CKAN	15	벨기에	Belgium Government Open Data	Drupal 웹사이트
2	캐나다	Open data in Canada	CKAN	16	아일랜드	data.gov.ie	CKAN
3	인도	Data Portal India	OGPL	17	뉴질랜드	New Zealand Government Open Data	CKAN
4	영국	data.gov.uk	CKAN	18	브라질	dados.gov.br	CKAN
5	오스트레일리아	Australian Government Catalogue	CKAN	19	오스트리아	Open Government Data Portal Austria	OGD
6	멕시코	datos.gob.mx	CKAN	20	스위스	Opendata.Swiss	CKAN
7	프랑스	Data.gov.fr	OpenDataSoft	21	인도네시아	Indonesia Data Portal	CKAN
8	한국	공공데이터포털	자체플랫폼, (CKAN도입계획)	22	싱가포르	Singapore's Open Data Portal	CKAN
9	독일	German Government Open Data	CKAN	23	핀란드	Finland Open Data	CKAN
10	일본	Japanese Government Open Data	CKAN	24	노르웨이	Norwegian Government Open Data	자체플랫폼 (DCAT표준)
11	이탈리아	Italian Government Open Data	CKAN	25	덴마크	Denmark Opendata	CKAN
12	스페인	Spain Government Data Portal	CKAN	26	스웨덴	Sweden Open Data	CKAN
13	러시아	Russie	DKAN	27	이스라엘	Israel Government Portal Open Data	CKAN
14	네덜란드	Netherlands Public Data Catalogue	CKAN				

(출처: 한국데이터진흥원 2017)

〈표 6〉 지역별 데이터 플랫폼 사용 현황

구분	CKAN	Socrata	DKAN	junar	OpenData Soft	기타	합계
유럽	140	4	11	-	10	115	280
북아메리카	50	53	7	9	2	99	220
아시아	21	-	-	-	-	49	70
오세아니아	39	3	-	-	-	8	50
남아메리카	27	1	-	5	1	8	42
아프리카	7	1	1	-	-	25	34
Worldwide	2	-	-	-	-	-	2
합계	286	62	19	14	13	304	698

(출처: 한국데이터진흥원 2017)

이다. 이 외 전통적인 도서관 표준인 MARC, 자체 개발 데이터 플랫폼에서 제공하는 기관별 어휘 관리 기능까지를 포괄하도록 구성하였다.

CKAN, DKAN 등 오픈 소스는 공식적으로 DCAT을 지원하지만 DCAT의 어휘용어를 그대로 채용하고 있지는 않다. 〈표 7〉은 CKAN

〈표 7〉 CKAN, DKAN 데이터 요소와 DCAT 매핑 예시

DCAT class	DCAT property	CKAN field	DKAN field
dcat:Dataset	-	extra:uri	
dcat:Dataset	dct:title	title	title
dcat:Dataset	dct:description	notes	body
dcat:Dataset	dcat:keyword	tags	field_tags
dcat:Dataset	dcat:theme	extra:theme	
dcat:Dataset	dct:identifier	extra:identifier	uuid
dcat:Dataset	adms:identifier	extra:alternate_identifier	
dcat:Dataset	dct:issued	extra:issued	release_date
dcat:Dataset	dct:modified	extra:modified	modified_date
dcat:Dataset	owl:versionInfo	version	
dcat:Dataset	adms:versionNotes	extra:version_notes	
dcat:Dataset	dct:language	extra:language	
dcat:Dataset	dcat:landingPage	url	
dcat:Dataset	dct:accrualPeriodicity	extra:frequency	field_frequency
dcat:Dataset	dct:conformsTo	extra:conforms_to	
dcat:Dataset	dct:accessRights	extra:access_rights	field_public_access_level
dcat:Dataset	foaf:page	extra:documentation	
dcat:Dataset	dct:provenance	extra:provenance	
dcat:Dataset	dct:type	extra:dcat_type	
dcat:Dataset	dct:hasVersion	extra:has_version	
dcat:Dataset	dct:isVersionOf	extra:is_version_of	
dcat:Dataset	dct:source	extra:source	
dcat:Dataset	adms:sample	extra:sample	
dcat:Dataset	dct:spatial	extra:spatial_uri	field_spatial_geographical_cover
dcat:Dataset	dct:temporal	extra:temporal_start + extra:temporal_end	field_temporal_coverage
dcat:Dataset	dct:publisher	extra:publisher_uri	og_group_ref
foaf:Agent	foaf:name	extra:publisher_name	
foaf:Agent	foaf:mbox	extra:publisher_email	
foaf:Agent	foaf:homepage	extra:publisher_url	
foaf:Agent	dct:type	extra:publisher_type	
dcat:Dataset	dcat:contactPoint	extra:contact_uri	
vcard:Kind	vcard:fn	extra:contact_name	field_contact_name
vcard:Kind	vcard:hasEmail	extra:contact_email	field_contact_email
dcat:Dataset	dcat:distribution	resources	field_resources
dcat:Distribution	-	resource:uri	
dcat:Distribution	dct:title	resource:name	
dcat:Distribution	dcat:accessURL	resource:access_url	
dcat:Distribution	dcat:downloadURL	resource:download_url	
dcat:Distribution	dct:description	resource:description	
dcat:Distribution	dcat:mediaType	resource:mimetype	

DCAT class	DCAT property	CKAN field	DKAN field
dcat:Distribution	dct:format	resource:format	
dcat:Distribution	dct:license	resource:license	
dcat:Distribution	adms:status	resource:status	
dcat:Distribution	dcat:byteSize	resource:size	
dcat:Distribution	dct:issued	resource:issued	
dcat:Distribution	dct:modified	resource:modified	
dcat:Distribution	dct:rights	resource:rights	
dcat:Distribution	foaf:page	resource:documentation	
dcat:Distribution	dct:language	resource:language	
dcat:Distribution	dct:conformsTo	resource:conforms_to	

(<https://github.com/ckan/ckanext-dcat>), DKAN(<https://dkan.readthedocs.io/en/latest/apis/open-data.html>)이 제공하는 API 활용 가이드 문서에서 이들이 관리하고 있는 요소를 조사하여 DKAN 요소를 매핑한 것이다.

〈표 7〉에서 보는 것처럼 CKAN의 경우는 DCAT의 모든 클래스, 속성에 대한 대응이 가능한 데이터베이스 구조를 가지고 있다. 반면 DKAN의 경우는 일부만 지원해 주고 있다. 한 가지 더 고려해야하는 것은 CKAN이 모든 요소에 대한 대응이 가능하지만 사용하는 네임스

페이스와 필드명이 서로 상이하여 반드시 매핑 작업이 필요하다는 점이다. DCAT 매핑 관리에서 저장한 정보는 실제 입수한 데이터셋을 정제하는 데 활용하며 이전 매핑 정보를 갖고 자동으로 일괄 매핑하는 작업과 신규 데이터셋에 대한 수동 작업 기능을 구분하여 구성하였다.

데이터셋 관리의 마지막 기능은 연동이다. 주지하다시피 DCAT의 최종 형태는 RDF 형식으로 발행 가능하다. 〈그림 3〉은 RDF 형태로 발행한 예제이다.

연동에 있어서의 핵심은 주제명과 저자명 연

```
<?xml version="1.0"?>
<rdf RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dcat="http://purl.org/dc/terms/"
  xmlns:dct="http://www.w3.org/ns/dcat#"
  xmlns:ns1="http://www.w3.org/2006/time#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <rdf:Description rdf:about="http://example.com/about">
    <dct:title xml:lang="en">Jinho Park Example Dataset1</dct:title>
    <dct:description xml:lang="en">Jinho Park Example Dataset1</dct:description>
    <dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2019-03-01</dct:issued>
    <dct:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2019-03-
28</dct:modified>
    <dcat:keyword-library<dcat:keyword>
    <dcat:keyword>open</dcat:keyword>
    <dcat:keyword>data</dcat:keyword>
    <dcat:theme rdf:resource="library"/>
    <dct:language rdf:resource="http://id.loc.gov/vocabulary/iso639-1/ko"/>
    <dcat:contactPoint rdf:resource="jinho.kor@gmail.com"/>
    <dcat:landingPage rdf:resource="jinhohome.kr"/>
    <dct:spatial rdf:resource="korea"/>
    <dct:temporal rdf:nodeID="n3"/>
    <dct:accrualPeriodicity rdf:resource="http://purl.org/linked-data/sdms/2009/code#req-A"/>
    <dct:identifier="https://jinhopark.com/datasets/example1"</dct:identifier>
    <dct:publisher rdf:nodeID="n4"/>
    <dcat:distribution rdf:nodeID="n5"/>
  </rdf:Description>
  <rdf:Description rdf:nodeID="n3">
    <rdf:type rdf:resource="http://www.w3.org/2006/time#Interval"/>
    <ns1:hasBeginning rdf:nodeID="n6"/>
    <ns1:hasEnd rdf:nodeID="n7"/>
  </rdf:Description>
  <rdf:Description rdf:nodeID="n4">
    <rdf:type rdf:resource="http://www.w3.org/2006/time#Instant"/>
    <ns1:inXSDDateTime rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2019-04-
02</ns1:inXSDDateTime>
```

```
{
  "http://example.com/about": {
    "http://purl.org/dc/terms/title": {
      {
        "type": "literal",
        "value": "Jinho Park Example Dataset1",
        "lang": "en"
      }
    },
    "http://purl.org/dc/terms/description": {
      {
        "type": "literal",
        "value": "Jinho Park Example Dataset1",
        "lang": "en"
      }
    },
    "http://purl.org/dc/terms/issued": {
      {
        "type": "literal",
        "value": "2019-03-01",
        "datatype": "http://www.w3.org/2001/XMLSchema#date"
      }
    },
    "http://purl.org/dc/terms/modified": {
      {
        "type": "literal",
        "value": "2019-03-28",
        "datatype": "http://www.w3.org/2001/XMLSchema#date"
      }
    },
    "http://www.w3.org/ns/dcat#keyword": {
      {
        "type": "literal",
        "value": "data"
      }
    }
  }
}
```

〈그림 3〉 DKAN을 활용한 데이터셋 표현 예제(좌: RDF/XML, 우: RDF/JSON)

결이다. 서지 데이터 연결은 주제명, 저자명 연결 후에 관련 데이터를 확보하여 가능하다. 이유는 DCAT의 클래스 중 서지 데이터를 연결할 수 있는 클래스는 존재하지 않기 때문이다. DCAT의 클래스 중 foaf:Agent, skos:Concept, skos:ConceptScheme을 활용하면 기존에 존재하는 도서관 링크드 데이터 중 동일 발행인, 발행자, 동일 주제명에 대한 직접 연결이 가능하다. 대표적인 국내 서지데이터 LOD 발행 서비스 기관은 국립중앙도서관, 국회도서관, 한국교육학술정보원이 있다. 이 세 개 기관에서 제공하는 모델링 정보를 기초로 하면 공통적으로 foaf:Agent, skos:Concept, skos:ConceptScheme을 사용하고 있음을 알 수 있다. <표 8>에 위 3개 기관의 해당 어휘 사용 내용을 정리한 후 모델링 정보 확인 경로를 표시하였다.

데이터셋 관리 기능의 핵심은 다양한 데이터셋 출처로부터 데이터를 입수하여 처리하고 이를 이용자 서비스에 활용하는 데 있다. 또한 이미 구축하여 서비스하고 있는 다양한 링크드 데이터와 인터링킹을 통해 해당 데이터셋에 대한 주제, 저자(단체)와의 연결로 궁극적으로 관련 서지 등 다양한 정보자원과 연결이 가능하다.

4.2.3 이용자 서비스

데이터셋 이용자 서비스는 기존의 디지털도서관 이용자 서비스와 연동하여 통합적으로 데이터셋 검색이 가능하도록 구성하였다. 단, 데이터셋 자체를 직접 수집하여 연구, 개발에 활용하기 위한 이용자를 고려하여 데이터셋 검색과 SPARQL Endpoint 기능을 추가하였다. 관건은 SPARQL Endpoint 서비스로 별도의 질의어 학습이 필요한 부분을 손쉽게 접근 가능하도록 개선하는 데 있다.

5. 결론

본 논문에서는 데이터셋을 새로운 지식정보자원으로 보고 해당 자원을 디지털도서관 시스템에서 관리·서비스하기 위한 시스템을 설계하여 제시하였다. DCAT 용어집 분석에서는 DCAT을 구성하는 8개 클래스 중 주 클래스 Catalog, Dataset, Distribution와 중요 클래스 CatalogRecord 4개에 중점을 두어 해당 클래스의 정의와 이에 속하는 속성, 속성의 기술방법을 상세히 검토하였다. 용어 분석 후에는 DCAT을 가지고 디지털도서관 시스템을 설계할 수 있는

<표 8> 국내 서지 링크드 데이터에서의 DCAT 클래스 활용 현황

구분	활용현황	모델링 정보
국립중앙도서관	개인명: foaf:name, rdfs:label 단체명: foaf:name, rdfs:label 주제명 전체: SKOS	https://lod.nl.go.kr/home/about/dataset.jsp
국회도서관	서지-저자 연결: foaf:Agent 서지-주제 연결: skos:Concept	http://cloud.nanet.go.kr/lod/lodIntro.do
한국교육학술정보원	서지-저자 연결: foaf:Agent 서지-주제 연결: skos:Concept	http://data.riss.kr/ontoModel.do

방안에 대해 제시하였는데 CKAN과 같은 오픈 소스를 설치하여 별도로 서비스 시스템을 구축하지 않고 디지털도서관 인프라를 활용할 수 있는 방안을 제안했다. 단, 본 연구가 특정 기관의 특정 디지털도서관 시스템을 염두해 두고 있지 않으므로 데이터셋 관리, 서비스에 필요한 핵심 기능을 위주로 개념도 구성하고 해당 기능에 대해 기술하였다.

디지털도서관 데이터셋 관리의 핵심 요소는 외부 데이터 반입, 반입 데이터를 기반으로 한 관리 시스템, 기존의 도서관 데이터베이스를 기반으로 한 링크드 데이터와의 연계와 최종 이용자 서비스까지를 제시하였다. 원천데이터 선정 시 중요 고려사항으로 국내·외 데이터 포털 서비스 현황을 조사하여 오픈소스 활용은 물론 자체적으로 개발한 데이터 플랫폼과의 연동 중요성을 제시하였다. 데이터셋 관리에서는 DCAT 어휘집을 표준으로 두고 다양한 어휘들과 매핑할 수 있는 매핑 관리 기능을 핵심으로 제시하였다. 그러면서 DCAT을 지원한다고 하는 프레임워크들도 자체 데이터 관리에 사용하는 필드가 DCAT 어휘가 다름을 확인하면서 매핑 관리의 필요성과 중요성도 알 수 있었다. 연동에 있어서는 DCAT 자체가 RDF로 기술하여 발행하므로 foaf:Agent, skos:Concept, skos:ConceptScheme을 활용해 기존 링크드 데이터와 연결이 가능하고 정보 확장이 가능한 것을 확인하였다. 마지막으로 최종 이용자들을 위한 데이터셋 서비스로 통합검색과 함께 데이터셋 검색(RDF, JSON), SPARQL Endpoint를 제시하였다.

DCAT을 활용하여 데이터셋 목록과 이에 속하는 데이터셋을 기술하면 검색 가능성을 높일 수 있다. 검색 가능성을 높일 수 있다는 것은 검색 엔진의 접근을 높이는 효과와 일반 이용자의 데이터셋 검색 가능성을 높일 수 있음을 의미한다. 뿐만 아니라 새로운 지식정보자원인 데이터셋을 표준 메타데이터를 활용하여 관리함으로써 다른 기관, 다른 정보서비스 인프라와 상호운용성 확보가 가능하여 자체 데이터의 품질도 높일 수 있다. 이 외 디지털도서관 관점에서는 디지털정보자원 보존이 가장 핵심적인 업무 중 하나인데 DCAT 메타데이터는 디지털 보존을 용이하게 하는 매니페스트 파일(manifest file)로 사용할 수 있다. 매니페스트 파일은 특정 단위의 파일 그룹에 메타데이터가 함께 들어있는 즉, 응용 프로그램과 관련 메타데이터가 함께 기록된 파일을 말한다.

새로운 정보자원이 만들어지고 유통될 때 도서관은 새로운 메타데이터에 대한 고민과 함께 새로운 메타데이터를 만들고 보급해 왔다. 디지털 정보자원에 대응하기 위한 MARC21, MODS, PREMIS 등이 그 노력의 결과라 할 수 있다. 비록 DCAT은 도서관계의 요구가 반영되고 주도적으로 만들어진 표준은 아니지만 해당 클래스와 속성은 도서관에서 범용적으로 사용하고 있는 요소들로 구성되었다. 데이터셋은 새로운 지식정보자원 유형이자 부가가치를 창출할 수 있는 핵심 성장동력으로 주목받고 있다. 현재 도서관은 연구데이터에 대한 연구개발을 지속하고 있지만 향후 범용적인 데이터셋 서비스를 시대에 대비한 다양한 노력이 필요한 때이다.

참 고 문 헌

- [1] 고영만, 배경재. 2011. 상이한 데이터 구조의 데이터베이스간 통합 운영방안 연구. 『한국문헌정보학회지』, 45(3): 69-85.
- [2] 김지현. 2012. 대학 내 연구자들의 연구데이터 관리에 관한 연구. 『한국도서관·정보학회지』, 43(3): 433-455.
- [3] 김지현. 2015. 데이터 관리와 공유에 대한 대학 연구자들의 인식에 관한 연구. 『한국문헌정보학회지』, 49(3): 413-436.
- [4] 노영희. 2012. dCollection의 링크드 데이터 구축에 관한 연구. 『한국도서관·정보학회지』, 43(2): 247-271.
- [5] 박옥남. 2012. PREMIS 기반 보존 메타데이터 요소 개발에 관한 연구. 『한국문헌정보학회지』, 46(2): 83-113.
- [6] 심원식, 안혜연, 변제연. 2015. 인문학 분야 연구데이터의 수집 및 활용성 증진을 위한 전략 연구. 『한국문헌정보학회지』, 49(3): 155-183.
- [7] 유사라. 2010. 메타데이터 주제 국내 연구동향 분석. 『한국문헌정보학회지』, 44(2): 405-426.
- [8] 이명희. 2019. 정책정보서비스 활성화를 위한 정책정보 자료원의 특성 연구. 『한국문헌정보학회지』, 53(1): 33-55.
- [9] 이병기. 2004. 학습객체 개념을 이용한 학교도서관 정보시스템(DLS)의 메타데이터 요소 확장에 관한 연구. 『한국문헌정보학회지』, 38(4): 85-104.
- [10] 이성숙, 송인석. 2005. NewsML을 고려한 기술동향정보 메타데이터에 관한 연구. 『한국문헌정보학회지』, 39(3): 183-205.
- [11] 이승민. 2015. 디지털 아카이빙을 위한 보존 메타데이터 패키지 구축. 『정보관리학회지』, 32(3): 21-47.
- [12] 한국데이터진흥원. 2017. 『CKAN을 활용한 민간데이터 연계 활성화 방안 연구 결과 보고서』. 서울: 한국데이터진흥원, Kdata 17-002.
- [13] Alexander, K., Cyganiak, R., Hausenblas, M. and Zhao, J. 2011. Describing linked datasets with the void vocabulary. [online] [cited 2019. 5. 10.]
<http://eprints.lancs.ac.uk/id/eprint/70518>
- [14] DCAT Application Profile Working Group. 2013. DCAT application profile for data portals in Europe. [online] [cited 2019. 4. 1.]
<https://joinup.ec.europa.eu/asset/dcatapplicationprofile/assetrelease/dcat-application-profile-data-portals-europe-final>

- [15] Ding, L., Peristeras, V. and Hausenblas, M. 2012. "Linked open government data [Guest editors' introduction]." *IEEE Intelligent systems*, 27(3): 11-15.
- [16] Guptill, S. C. 1999. "Metadata and data catalogues." *Geographical information systems*, 2: 677-692.
- [17] Hallo, M., Luján-Mora, S. and Trujillo Mondéjar, J. C. 2014. "Transforming library catalogs into linked data." *In Proceedings of ICERI 2014 Conference*, 17-19 November 2014, Seville, Spain.
- [18] Maali, F., Erickson, J. and Archer, P. 2014. Data catalog vocabulary (DCAT). W3c recommendation, 16. [online] [cited 2019. 4. 1.] <<https://www.w3.org/TR/vocab-dcat/>>
- [19] Noy, N. 2018. Making it easier to discover datasets. Google blog, [online] [cited 2019. 4. 1.] <<https://www.blog.google/products/search/making-it-easier-discoverdatasets/>>
- [20] ZHANG, Y., JIANG, D. and LIU, Q. 2009. "Metadata-based integration scheme for heterogeneous datasets." *Journal of Tsinghua University (Science and Technology)*, 7.

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- [1] Ko, young-Man and Bae, Kyung-Jae. 2011. "A Study of the Integrated Operation for Databases of Different Data Structures." *Journal of the Korean Society for Library and Information Science*, 45(3): 69-85.
- [2] Kim, Ji-Hyun. 2012. "A Study on University Researchers' Data Management Practices." *Journal of Korean Library and Information Science Society*, 43(3): 433-455.
- [3] Kim, Ji-Hyun. 2015. "A Study on the Perceptions of University Researchers on Data Management and Sharing." *Journal of the Korean Society for Library and Information Science*, 49(3): 413-436.
- [4] Noh, Young-Hee. 2012. "A Study on Configuring dCollection as the Linked Data." *Journal of Korean Library and Information Science Society*, 43(2): 247-271.
- [5] Park, Ok-Nam. 2012. "A Study on Developing Preservation Metadata Based on PREMIS Focusing on Digital Data in National Library of Korea." *Journal of the Korean Society for Library and Information Science*, 46(2): 83-113.
- [6] Shim, Won-Sik, Ahn, Hye-Yeon and Byun, Jae-Yeon. 2015. "Strategies for Improving the Collection and Use of Research Data in the Humanities." *Journal of the Korean Society for Library and Information Science*, 49(3): 155-183.

- [7] Yoo, Sa-Rah. 2010. "A Diagnostic Analysis of Metadata R&D Status." *Journal of the Korean Society for Library and Information Science*, 44(2): 405-426.
- [8] Lee, Myeong-Hee. 2019. "Investigating the Characteristics of Policy Information Resources for Activating Policy Information Services." *Journal of the Korean Society for Library and Information Science*, 53(1): 33-55.
- [9] Lee, Byeong-Ki. 2004. "A Study on the Metadata Element's Expansion of DLS Based on Learning Object." *Journal of the Korean Society for Library and Information Science*, 38(4): 85-104.
- [10] Lee, Sung-Sook and Song, In-Seok. 2005. "A Study on Metadata of Technology Trends Information Based on NewsML." *Journal of the Korean Society for Library and Information Science*, 39(3): 183-205.
- [11] Lee, Seung-Min. 2015. "Construction of Preservation Metadata Package for Digital Archiving." *Journal of the Korean society for information management*, 32(3): 21-47.
- [12] KOREA Data Agency. 2017. *Report on the results of research on the activation of private data link using CKAN*. Seoul: KOREA Data Agency, Kdata 17-002.