교통 분야 Data 관리와 공유를 위한 DCAT 기반 Data Catalogue 표준 연구: DCAT-Trans

신도겸¹ · 이상화² · 강정현³ · 박은미^{4*}

¹(주)데이터위즈 교통팀 책임연구원, ²목원대학교 산학협력단 책임연구원, ³(주)데이터위즈 선임연구원, ⁴목원대학교 도시공학과 교수, (주)데이터위즈 대표이사

Data Catalogue Standards Based on DCAT for Transportation Data: DCAT-Trans

SHIN, Doh Kyoum¹ · LEE, Sang Hwa² · KANG, Junghyun³ · PARK, Eun Mi^{4*}

¹Senior Researcher, Transportation Team, Datawiz Inc, Daejeon 35349, Korea

²Senior Researcher, Industry Academic Cooperation Foundation, Mokwon University, Daejeon 35349, Korea

³Associate Researcher, Transportation Team, Datawiz Inc, Daejeon 35349, Korea

⁴Professor, Department of Urban Engineering, Mokwon University, Daejeon 35349, Korea
CEO, Datawiz Inc, Daejeon 35349, Korea

*Corresponding author: peunmi@mokwon.ac.kr

Abstract

It is one of the core tasks of Korean Government to build a data-based industrial ecosystem to promote data trades and re-use with opening public data. The transportation sector also generates various data from existing systems such as ITS. The amount of data in the transportation sector is forecast to rapidly increase after AV and C-ITS are introduced into a real world. It is an important matter and concern for data managers and users to efficiently manage a large amount of data. EU leads to establish standards for data catalogues to manage metadata in order to manage, share and re-use data. This study suggests standards for data catalogue for transportation data based on DCAT. The study examines cases using DCAT and analyses ITS, BIS and C-ITS field test data sets to derive features of transportation data of Korea. DCAT-Trans, which this study suggests, will well represent relationships between transportation data and geographical features of them. This study is meaningful as an attempt to standardize managing and classifying transportation data.

Keywords: data catalogue, DCAT, metadata, open data, transportation data

초록

공공부문의 데이터 개방을 통해 데이터 유통과 활용을 촉진하고, 데이터 기반 산업의 발전을 위한 환경을 조성하는 것이 정부의 중점 과제 중 하나이다. 교통분야 역시 ITS 등 기존의 데이터를 생산하던 시스템에 더하여 자율주행차량과 C-ITS의 도입으로 향후 교통 데이터의 비약적으로 증가할 것으로 예상된다. 증가하는 데이터를 효율적으로 관리하는 것은 관리자와 이용자 모두에게 중요한 문제이다. 효율적인 데이터 관리와 공유, 활용을 위해 메타데이터를 체계적으로 관리하기 위해 데이터 카탈로그 표준이 EU를 중심으로 제정중이다. 본 연구는 이러한 추세에 맞추어 교통분야 데이터를 위해 DCAT에 기반한 데이터 카탈로그 작성 표준을 제안하고자

J. Korean Soc. Transp.
Vol.37, No.5, pp.430-444, October 2019
<https://doi.org/10.7470/jkst.2019.37.5.430>

pISSN : 1229-1366
eISSN : 2234-4217

ARTICLE HISTORY

Received: 30 September 2019

Revised: 18 October 2019

Accepted: 29 October 2019

Copyright ©
Korean Society of Transportation

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

한다. 본 연구는 DCAT을 사용한 해외 사례의 분석과 국내 ITS, BIS, C-ITS 시범사업의 데이터를 분석하여 적합한 표준을 도출하였다. 본 연구가 제안하는 표준인 DCAT-Trans는 교통분야의 데이터 연결관계와 공간정보 특성을 잘 표현할 수 있도록 구성하였다. 본 연구는 교통 분야 데이터의 관리와 분류 체계를 표준화할 수 있는 기초 연구로서 가치가 있다.

주요어: 데이터 카탈로그, DCAT, 메타 데이터, 데이터 공유, 교통 데이터

서론

1. 연구 배경 및 목적

현재 교통 분야에서 중요한 이슈 2가지는 자율주행차량과 Cooperative Intelligent Transportation Systems (C-ITS)이다. 자율주행차(Automated Vehicles, AV)와 C-ITS 모두 IT기술의 발전과 더불어 구현이 가능하고, 다량의 데이터를 생산할 것으로 예상된다. 실험용 자율주행차량은 단지 차량에 설치된 센서가 수집한 자료만으로도 매일 11TB-152TB의 데이터를 생산하는 것으로 보고되고 있다(Rossi, 2019). 또한 기존의 도로망을 중심으로 설치된 Intelligent Transportation Systems (ITS)를 통해 다량의 데이터를 생산하고 있다. ITS의 발전형인 C-ITS 역시 기존의 교통량 데이터와 더불어, 돌발사고, 보행자 인식, 노면 상태 정보 등 도로변에 설치된 각종 시설에 수집 자료와 교통 신호정보 등 되는 다양한 데이터를 생산할 것으로 예상된다. 따라서 데이터를 관리하고 활용하는 것이 미래 교통분야에서 중요한 한 축이라 할 수 있다.

지금까지 교통분야에서 생산되는 데이터는 데이터 생산 기관에 국한되어 그 활용이 제한적이었다. 그러나 현재는 공공데이터를 개방을 통해 데이터 기반 산업의 발전을 위해 환경을 조성하는 것이 전 세계적 추세이다(National IT Industry Promotion Agency, 2014). 우리정부 역시 교통분야를 포함한 다양한 영역에서 데이터 개방과 유통을 위한 노력을 계속하고 있다(Ministry of the Interior and Safety, 2018).

방대한 데이터를 효율적으로 관리하는 것은 데이터의 관리자와 이용자 모두에게 중요하다. 관리자는 수집 또는 생산하는 데이터의 무엇인지, 갱신 주기는 어떻게 되는지, 공간 및 시간특성이 어떻게 되는지 체계적으로 관리할 수 있어야 데이터의 품질을 유지할 수 있다. 반면 이용자 측면에서는 어떤 데이터가 제공되는지, 데이터의 주제(theme)와 속성(property)은 무엇인지, 어떤 형식으로 제공되는지 등이 중요하다. 이런 작업은 데이터베이스(Database, DB)의 데이터 카탈로그(Data Catalogue)를 체계적으로 작성함으로써 이루어질 수 있다.

데이터 카탈로그란 사용 가능한 데이터의 목록과 관련 메타데이터를 제공하여 데이터관리와 검색을 용이하게 하는 도구이다(Wells, 2019). 빅데이터 분야에서 표준화가 진행 중 인 데이터 카탈로그는 관리자 측면에서 자동화된 데이터 업데이트 관리 등 데이터관리의 자동화를 가능하게 한다(Maali et al., 2010; Neumaier et al., 2015). Figure 1은 데이터 카탈로그를 사용할 때와 아닐 때의 데이터 처리과정의 차이를 설명하고 있다(Wells, 2019). 이용자 측면에서 데이터 카탈로그가 없는 경우 이용자는 원하는 데이터를 찾기 위해 검색과 데이터 검증에 위한 단계를 반복하여야 한다. 반면 데이터 카탈로그가 있다면 데이터 카탈로그에 이미 데이터에 대한 상세한 정보가 기록되어 있으며 적합한 데이터를 찾기 위한 반복 작업 단계를 줄일 수 있는 이점이 있다. 이와 더불어 데이터 카탈로그는 데이터의 재활용과 개방성의 최상위 단계인 Linked Data 또는 Linked Open Data (LOD)를 위한 필수요소이다(Martin et al., 2013).

교통분야 역시 데이터의 개방과 활용이라는 측면과 자율주행과 협력주행(Connected and Automated Vehicles, CAV)이 활성화된 환경 하에서 방대해지는 데이터의 효율적인 관리를 위해 데이터관련 표준 작성이 중요하다. 일례로 국제표준화기구(ISO)의 TC 204에서는 교통과 자율주행 차량관련 표준을 작성하기 위해 활동 중이다. ISO TC 204의 활동은 교통과 관련한 다양한 분야를 포함하고 있으며, 특히 WG1 (Working group 1)과 WG3 (Working group 3)은 ITS분야 데이터의 정의와 데이터베이스 교환을 위한 호환 표준화 등을 연구하고 있다.

Figure 1. Process with and without a data catalog (Wells, 2019)

교통분야 데이터베이스의 효율적인 관리와 전국에서 개별 지자체 및 기관별로 수집되는 데이터를 효과적으로 공유하기 위해서는 데이터 카탈로그의 표준을 만들고 표준에 맞춰 데이터 카탈로그를 만들어야 한다. 이는 메타데이터를 관리하는데 효율성을 제공하고, 다양한 데이터를 공공에 노출하고 융합과 활용하는 측면에서 이득이다. 따라서 본 연구는 교통 데이터를 관리하기 위한 카탈로그 작성을 위해 Data Catalog Vocabulary (DCAT) 표준을 기반으로 교통데이터의 관리에 적합하도록 개선된 표준을 제시하고자 한다.

2. Data Catalog Vocabulary (DCAT)

데이터는 이용자 또는 생산자 입장에서 같은 종류의 데이터라 할지라도 생산된 기관에 따라 서로 다른 형식과 구조를 가지고 있다. 같은 주제의 데이터가 표준화된 형식과 구조를 가진다면 그 활용가치가 더 높겠으나, 현실적으로 일치하지 않는 경우가 다수이며 이를 억지로 일치 시키는 것 역시 많은 시간과 노력을 요구한다. 따라서 데이터의 형식을 억지로 일치시키기보다 데이터를 설명하는 메타데이터(metadata)를 표준화하여 기록 및 정리하는 것이 시간과 비용, 노력 측면에서 더 유리하다.

메타데이터를 표준화함으로써 데이터의 상호 호환이 가능하며, 메타데이터를 이용한 데이터 검색의 용이성으로 인해 데이터 접근성이 높아질 수 있다. 데이터 카탈로그는 이런 메타데이터를 관리하는 도구이다. 따라서 DB 전문가가 아니라도 메타데이터만으로 데이터의 특성을 파악하고 있어 보다 손쉬운 관리가 가능하고, 이용자는 데이터를 직접 확인하지 않아도 해당 데이터의 구조와 내용을 파악할 수 있다.

DCAT 표준은 데이터 카탈로그를 작성하기 위한 RDF (Resource description framework) 언어 표준이다(W3C, 2019a). RDF는 W3C (World Wide Web Consortium)에서 제정한 웹상서 사용되는 자원의 정보를 표현하기 위한 제한한 규격으로, 상이한 메타데이터를 해석하기 위한 공통적인 규칙을 제공하고 있다. 또한 웹상에서 모든 정보를 자동화한 처리가 가능하도록 하고, 어플리케이션 상호간에 효율적인 교환과 호환을 목적으로 한다(W3C, 2014a). RDF는 공통의 프레임워크를 사용함으로써 의미의 손실 없이 응용프로그램간 정보의 교환이 가능하며 특히 Linked data에 적합하다(NIA, 2014).

DCAT 표준은 원칙적으로 웹상에서 메타데이터의 상호 호환을 위해 개발되었다. DCAT표준으로 데이터 카탈로그를 작성함으로써 데이터 카탈로그의 호환이 가능하며 이는 서로 다른 곳에 존재하는 데이터라도 상호 검색 및 공

유를 가능하게 해준다. 기존 데이터 포털은 서로 다른 메타데이터 형식을 가지고 있다. 이로 인해 데이터가 파편화 되고 원하는 데이터를 찾기 어려워지는 현실적 문제에 직면하고 있다. 이를 해결하기 위해 DCAT은 서로 다른 메타 데이터 형식은 인정하면서 DCAT표준으로 데이터 카탈로그(메타데이터 카탈로그)를 작성하고 배포함으로써 상호 운용성을 높이고자 한다(Jung, 2019). DCAT은 현재 대표적인 오픈데이터 관리 프로그램인 CKAN과 유럽연합 포털에서 메타데이터 관리를 위한 표준 도구로 사용하고 있다.

DCAT 표준은 2014년 버전 1.0 (DCAT 1.0)이 등장한 이후, 2019년 3월에 버전 2.0 (DCAT 2.0)으로 개선되었다. Figure 2는 DCAT 2.0의 주요 클래스간의 관계를 나타낸 스키마(Schema)이다. DCAT 2.0은 6개의 핵심 클래스(main class)와 부가 클래스(optional class)로 구성되어 있다(W3C, 2019b). 핵심 클래스의 정의를 살펴보면, 먼저

Figure 2. Overview of DCAT model, classes and the relationship between them (W3C, 2019a)

dcat:Catalog는 Dataset에 포함되어 있는 개별 데이터의 메타데이터 기록이다. dcat:Resource는 데이터 카탈로그에 포함되어 있는 개별 데이터의 데이터명, 생성주기, 관리자 및 기간 등을 정의한다. dcat:Dataset은 생산 또는 제공할 데이터의 모음이며 수집주기, 배포방법, 공간적, 시간적 범위를 정의한다. dcat:Distribution은 데이터를 제공하는 방법과 형태 등을 정의하고 있으며, 데이터의 라이선스, 권한, 다운로드 URL, 사이즈 등을 정의한다. dcat:Distribution은 dcat:DataService를 통해 제공되며, 제공되는 데이터 목록과 엔드포인트(Endpoint, 서버, 스마트폰, 노트북 등 네트워크에 연결할 수 있는 모든 기기)의 URL (Uniform Resource Locator) 등을 정의한다. 마지막으로 dcat:CatalogRecord는 데이터 카탈로그에 등록된 데이터의 등록정보와 변경 정보를 정의한다. 현재 옵션 항목으로 구현되지 않을 수 있다.

Dataset과 Catalog, Resource 클래스를 예를 들어 설명하면 2019년 9월20일, 21일, 22일 3일간의 세종시 버스 DTG (Digital Tacho Graph) 자료가 있다고 가정하면 Dataset은 이 3일간의 개별 데이터의 모음이다. 이 DTG 데이터는 기록된 날짜를 제외하면 배포방법(Distribution), 공간범위(세종시), 시간범위(0-24시)로 모든 것이 동일하므로 하나의 dataset으로 분류할 수 있다. 반면 catalog 클래스는 Dataset에 3개의 데이터가 포함되어 있다는 것을 기록하고 있으며, 핵심 주제가 DTG이며 어디에서 제공될지를 간략하게 정의한다. Resource는 개별 데이터에 대해 상세하게 생산자, 사용권한, 종류, 작성일 등이 기록된다. 예의 경우 각각의 데이터의 경우 대부분의 항목이 동일하겠지만 생성일자는 다를 것이다. Resource에서는 주제를 DTG, 버스 등 다수의 주제로 정의할 수 있어 데이터를 상세하게 분류할 수 있다.

그 외에 부가 클래스로 Location, Period of Time, Role, Organizaion/Person, Concept 등이 존재하여 필요에 따라 추가하여 사용할 수 있다.

본론

1. 연구 방법

본 연구는 메타데이터를 관리하기 위한 데이터 카탈로그 작성 표준을 제시하는 것이 연구의 목적이다. 본 연구를 위해 사례 분석을 통해 표준의 방향을 설정하고 제시하였다. 먼저 현재 DCAT 표준을 적용하여 데이터 카탈로그를 작성하고 있는 해외 사례를 분석하였으며, DCAT 표준이 1.0에서 2.0으로 개선되면서 주요하게 변한 부분을 분석하여 메타데이터 관리의 발전 경향을 파악하였다. 마지막으로 교통분야의 데이터를 분석하여 어떤 종류의 데이터로 구성되어 있는지 파악하였다. 교통분야 데이터 분석은 도로 교통에 국한하였으며, 일반 승용차와 버스를 포함하였다. 그러나 완전한 CAV환경은 아직 구현되지 못하여 기존 ITS와 BIS (Bus Information System) 기반 데이터와 시범사업으로 테스트한 C-ITS 데이터를 중심으로 활용하였다.

2. DCAT을 사용한 데이터 카탈로그 작성 사례 분석

Table 1은 DCAT을 사용하여 데이터 카탈로그를 관리하는 데이터 포털에서 사용하는 DCAT 클래스와 속성을 보여주고 있다(W3C, 2014b). 모든 속성을 사용하는 포털은 존재하지 않으며, 포털별로 필요한 속성을 선택해서 데이터 카탈로그를 작성하고 있다. DCAT 1.0의 핵심 4개 클래스 중 Dataset 클래스를 제외하고는 포털 별로 선택하여 사용하였다. Distribution 클래스가 13개 포털 중 11개에서 사용되고 있었으며, CatalogRecord 클래스는 단 4개 포털만이 사용하고 있었다. 포털별로 데이터의 관리와 배포/공유 특성에 따라 다른 특성을 보였다.

비교 대상 중 가장 충실하게 DCAT 1.0에서 제안하는 속성을 사용한 포털은 Open Data Support (ODS), The Biological and Chemical Oceanography Data Management Office (BCO), The All-Island Research Observatory (AIRO)로 대체로 특정분야에 한정된 연구 분야의 데이터를 관리하거나, Open Data의 활성화를 지원하는 기관이었다. 순수하게 데이터를 유통하는 기관인 EU (European union), CL (Castile and Leon Open Data Catalog), SP

(Spanish Ministry of Industry, Energy and Tourism Open Data Catalog), SU (Open Data at University of Southampton), QC (The Quebec government Open Data Portal)와 같이 정부관련 포털의 경우 Dataset과 Catalog class의 속성에 중점을 두고, Distribution class는 단순하게, CatalogRecord는 경우에 따라서 없는 경우도 있다. EU 포털의 경우 Catalog 클래스 역시 없으며, Dataset 클래스의 속성만이 중요하게 관리되고 있다.

이런 사례는 부분적으로는 DCAT의 포털의 특성에 따른 확장성을 잘 보여 주고 있다. 그러나 한편으로는 오픈 소스를 활용한 데이터 카탈로그 작성의 어려움과 분야별 데이터의 상이성으로 인해 공통적으로 활용할 수 있는 속성에 한계를 보여주고 있다. 이는 특정 분야에 제한된 데이터 포털의 경우 보다 많은 수의 속성을 활용하고 있는데 반해 종합 데이터 포털의 경우 보다 간략하게 구성되어 있어 간접적으로 유추할 수 있다.

Table 1. Comparisons of properties of DCAT-based data catalogues used in data portals (W3C, 2014b)

Term	ODS	EU	CL	SP	DUB	BCO	SU	DT	SC	AIRO	PH	CKAN	QC	Count
Catalog Class	Y		Y	Y		Y	Y	Y		Y		Y	Y	9
dct:title	Y		Y	Y		Y	Y	Y		Y		Y	Y	9
dct:description	Y		Y	Y	Y		Y			Y		Y	Y	8
dct:issued	Y					Y						Y		3
dct:modified	Y					Y		Y				Y		4
dct:language	Y					Y				Y				3
dct:license	Y		Y	Y								Y		4
dct:rights	(Y)		(Y)	(Y)								(Y)		(4)
dct:spatial	Y			Y		Y				Y				4
foaf:homepage	Y		Y	Y		Y	Y	Y		Y		Y	Y	9
dcat:dataset	Y		Y	Y		Y		Y		Y		Y	Y	8
dcat:record	Y					Y				Y			Y	4
dct:publisher	Y		Y	Y		Y	Y			Y			Y	7
dcat:themeTaxonomy				Y		Y				Y				3
Dataset Class	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	13
dct:title	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	13
dct:description	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		12
dct:issued	Y	Y	Y	Y	Y	Y		Y		Y	Y	Y		10
dct:modified	Y			Y		Y		Y		Y		Y		6
dct:identifier	Y	Y	Y	Y		Y		Y	Y	Y		Y	Y	10
dcat:keyword	Y		Y	Y	Y	Y			Y	Y		Y	Y	9
dct:language	Y	Y				Y				Y				4
dcat:contactPoint	Y					Y								2
dct:temporal	Y					Y				Y				3
dct:spatial	Y	Y	Y	Y						Y			Y	6
dct:accrualPeriodicity	Y		Y	Y						Y	Y		Y	6
dcat:theme	Y	Y	Y	Y	Y	Y				Y	Y		Y	9
dcat:landingPage						Y				Y	Y			3
dct:publisher	Y	Y	Y	Y	Y	Y	Y			Y	Y	Y		10
dcat:distribution	Y	Y	Y	Y	Y	Y			Y	Y	Y	Y	Y	11
CatalogRecord Class	Y					Y				Y			Y	4
dct:title	Y					Y				Y				3
dct:description	Y					Y				Y				3
dct:issued	Y					Y							Y	3
dct:modified	Y					Y							Y	3
foaf:primaryTopic	Y					Y				Y			Y	4
Distribution Class	Y	Y	Y	Y	Y	Y			Y	Y	Y	Y	Y	11
dct:title	Y					Y			Y	Y		Y	Y	6
dct:description	Y					Y			Y	Y		Y		5
dct:issued						Y								1

Table 1. Comparisons of properties of DCAT-based data catalogues used in data portals (W3C, 2014b) (continued)

Term	ODS	EU	CL	SP	DUB	BCO	SU	DT	SC	AIRO	PH	CKAN	QC	Count
dct:modified					Y									1
dct:license	Y									Y		Y		3
dct:rights	(Y)								Y	Y				2(1)
dct:format		Y	Y	Y	Y	Y			Y	Y		Y	Y	9
dcat:byteSize		(D)			(D)						Y			1(2)
dcat:mediaType	Y					Y				Y	Y	Y		5
dcat:accessURL	Y	Y			Y	Y			Y	Y		Y	Y	8
dcat:downloadURL	Y					Y				Y	Y	Y		5

note: ODS, Open Data Support; EU, European Union Open Data Portal; CL, Castile and Leon Open Data Catalog; SP, Spanish Ministry of Industry, Energy and Tourism Open Data Catalog; DUB, Doublinked.ie; BCO, The Biological and Chemical Oceanography Data Management Office; SU, Open Data at University of Southampton; SC, semantic.ckan.net; PH, Dataset Descriptions for Open Pharmacological Space; AIRO, The All-Island Research Observatory; CKAN, Simple Data Catalog Interoperability Proposal; QC, The Quebec government Open Data Portal; DT, The DataTank.

3. DCAT 1.0과 2.0의 차이

최근 새로 제시된 DCAT 2.0과 기존 DCAT 1.0의 차이점은 크게 3가지로 정리할 수 있다. 첫째, DCAT 1.0의 Dataset, Catalog, Distribution 클래스에 속해 있던 일부 속성을 Resource 클래스를 신설하여 이전한 후 통합 관리하고 있다. 대표적으로 Dataset과 Catalog 클래스에서 반복적으로 사용하던 dct:title, dct:description 등이 Resource 클래스의 하위 속성으로 이전되었다. 이외에 버전 1.0의 Dataset 클래스에 있던 다수의 속성을 Resource 클래스로 이전하였다.

둘째, 자료의 지리적 위치정보를 관리하는 속성이 강화되었다. 기존 버전 1.0에서는 Dataset 클래스 아래 dct:spatial 하나의 속성으로 위치 정보를 관리하였다. 그러나 버전 2.0에서는 dct:spatial, dct:spatialResolutionInMeters으로 조금 더 상세하게 위치 정보를 관리하게 하고 있다. 또한 Location 클래스를 서브 클래스(sub-class)로 지원하여 위치 정보를 보다 상세하게 기록하게 하고 있다. Location 클래스 아래에 하위 속성으로 locn:geometry, dcat:bbbox, dcat:centroid, w3cgeo:lat, w3cgeo:long, w3cgeo:alt 두어 보다 정밀한 위치 정보를 수록할 수 있도록 하고 있다.

세 번째 차이점은 버전 2.0에서는 데이터 제공과 관련된 항목들이 보강되었다. 데이터를 업로드 및 다운로드 할 수 있는 URL이나 제공 포맷 등이다. 이는 기존에 Distribution 클래스 단독으로 제공관련 정보를 관리하던 것을 DataService 클래스를 추가하여 서비스 되는 data set이나 스마트폰, 데스크톱 등 이용자가 사용하는 시스템(endpoint)의 접속을 관리하도록 하고 있다.

DCAT을 적용하여 카탈로그를 관리하는 데이터 포털의 사례와 마찬가지로 교통분야 역시 데이터 관리 측면에서 DCAT에서 제시하는 모든 클래스와 속성을 사용할 필요는 없다. 그러나 DCAT 1.0에서 2.0으로 변화의 경향을 살펴볼 때 위치정보에 대한 항목의 강화는 유의미하다. 교통 데이터는 그 활용을 위해 필연적으로 다른 데이터에 비해 공간정보가 더욱 중요하다. 예를 DSRC 통신자료의 경우 장비의 위치 정보와 측정 구간에 대한 상세한 공간 정보가 없으면 유용성이 반감된다. 분석된 차량 통행 속도나 교통량이 정확히 어느 지점에서 발생한 것인지 알지 못하기 때문이다.

4. 메타 데이터 관리 측면에서의 교통 데이터 특성

미래 교통환경은 협력주행(cooperative)과 연결주행(connected), 자율주행(automated)으로 정의할 수 있으며, 자율주행 버스의 운행은 주요 시나리오 중 하나이다(EC, 2017a). 따라서 교통부문에서 데이터의 관리와 처리는 자율주행차량의 운행과 C-ITS가 활성화된 교통 환경을 고려해야한다.

지능형 대중교통 시스템(IPTS)은 ITS의 하위시스템(subsystem)으로 대중교통 네트워크의 상태를 분석하고 평가하기 위해 개발되었다(Elkosantini and Darmoul, 2013). 차세대 대중교통 시스템(C-IPTS) 역시 C-ITS의 하위시스템으로 IPTS의 역할을 동일하게 수행하지만 기능적으로 C-ITS와 같이 새로운 환경과 기술을 접목하고 있다.

C-IPTS 시스템은 자율주행 버스의 센서 데이터와 운행기록, 사고/돌발사항 등 안전, 도로면 상태 등의 관련 자료와 노드·링크 정보, 노선 정보 등으로 구성될 것이다. 이 중 버스 차량 센서 데이터는 일반 차량 센서 데이터와 기본적으로 동일하다. 반면 사고/돌발사항, 도로면 상태 등의 자료는 C-ITS에서 수집하는 자료를 공유하면 되므로 C-ITS 자료와 동일하다고 할 수 있다. 버스 운행기록은 BIS에서 기록하는 자료와 기본적으로 동일할 것으로 판단된다. 또한 분석과 결과 표출을 위한 노드·링크 정보와 노선 정보 등 역시 자료의 분석과 활용 측면에서는 기존 방식과 큰 차이가 없을 것으로 판단된다. 그러나 자율주행 차량의 운행에 필요한 정밀지도(HD Map)는 라이다(Lidar)나 이미지 영상 등 차량 센서 데이터의 추가분석을 위해 같이 제공되어야 하는 자료이다.

자율주행 차량의 센서 데이터는 기존의 자료와는 구성이 완전한 차이가 있다. 기존 ITS관련 자료는 정형 데이터베이스(Relational Database, RDB)의 정형 테이블 형태로 저장된다. 그러나 센서 데이터는 JSON, XML, 이미지 등 비정형데이터로 수집 및 저장되는 것이 일반적이다. 수집 예정 센서 데이터는 사물 영상 및 인식정보와 라이다 데이터, 레이더(Radar) 데이터, GPS 위치정보, 소나(SONAR, Ultrasonic) 데이터 등이 있다. 또한 각종 센서와 차량 장비의 상태를 기록하는 OBD2 (On-board diagnostics 2) 데이터가 수집된다. 이러한 센서 기반 데이터는 차량의 운행 알고리즘 개발 등 차량의 기계적인 분야에 초점이 맞춰져서 활용될 것으로 판단된다. 따라서 교통 관리와 운영의 측면에서는 센서기반 자료보다는 C-ITS 또는 C-IPTS 기반 수집 자료가 더 활용가치가 높다.

교통정보가 다양해지고, 데이터의 양이 증가할수록 메타데이터의 관리가 중요하다. 예를 들어 일반적으로 노변 장비를 이용해서 수집된 자율주행 차량의 위치 정보 자료의 경우 여러 차량이 혼재하고 있다. 활용 목적에 따라서는 차량의 종류별로 구분하고, 데이터를 분할하는 작업이 필요하다. 또한 연속적으로 수집되는 자료를 몇 분 단위로 분할해서 기록할 것인지 지정할 필요도 있다. 이런 다양한 전처리 또는 데이터 분할, 병합 작업을 수행하기 위해서는 어떤 데이터가 어떤 과정을 거쳐서 어떤 데이터명으로 새롭게 기록되는지 관리하는 것이 필수적이다. 이런 데이터의 수집, 가공, 저장 이력을 관리하는 것이 메타데이터이며 이 메타데이터를 관리하는 도구가 데이터 카탈로그이다. 예를 들어 연속된 자료가 분할되어 저장될 경우 데이터 카탈로그는 이 데이터를 하나의 데이터셋에 보관할 것이며, 데이터셋의 개별 데이터의 데이터 기록 시작시간과 끝시간을 기록하고, 버스, 트럭 등의 수단을 구분할 것이다. 또한 이 데이터의 생성 주체와 관리주체, 제공 데이터 유형, 수집 지역 등을 기록할 것이다.

현재 자율주행과 C-ITS 모두 아직은 테스트 단계에 머물러 있다. 이에 자율주행과 C-IPTS 환경에서 수집될 데이터의 특성 파악은 기존 ITS 데이터와 BIS 자료 등을 참조하여 추정하고자 한다. 데이터 특성 파악을 위해 대전·세종 C-ITS 시범사업을 위한 DB 테이블, 세종시 ATMS 테이블, 세종시 BIS 테이블, 자율주행차량에서 수집하는 센서 자료 목록을 활용하였다. 먼저 대전·세종 C-ITS 시범사업에 사용된 DB 테이블은 총 136개, 세종시 ATMS (Advanced traffic management system)의 DB 테이블은 290개, 세종시 BIS의 DB 테이블은 239개로 구성되어 있다. 이 중 실제 교통분야에서 분석 및 가공, 제공 등의 목적으로 활용이 가능한 테이블의 수는 각각 C-ITS DB 50개, 세종 ATMS DB 84개, 세종 BIS DB 49개이다. 나머지는 System_Connection_Status와 같이 DB 시스템 자체에 대한 정보 테이블과 임시 저장테이블 등이다.

유용성이 있다고 판단되는 DB 테이블을 분류하면 Table 2와 같이 수집자료, 기초자료, 집계자료의 3가지 유형으로 구분할 수 있다. 수집데이터는 OBU (On board unit), RSE (Road side equipment), 차량 위치정보, CCTV 영상 등 차량 및 도로면 장비에 의해 수집되는 원 자료이다. 기초데이터는 노드·링크 정보, 노선정보, 운행시간표, 지도자료 등 차량이 운행하는 도로와 해당지역, 대중교통의 경우 스케줄 등의 정보이다. 데이터 테이블의 수를 살펴보면 수집자료는 44개로 가장 비중이 적었다. 반면 기초자료는 79개로 거의 절반 수준이었으며, 집계자료 역시 68개로 다수였다.

Table 2. Classifying transportation data by data types and themes

Theme	Type	Driving records	Safety/ Incidents	Information	Facility/ Location/Map	Bus information	Vehicle sensor data	Total
C-ITS	Collected	5	8	1				14
	Basic				16			16
	Computed	15	3	2				20
Sejong ATMS	Collected	4	5	3				12
	Basic				34			34
	Computed	33	5					38
Sejong BIS	Collected	8	5					13
	Basic				7	19		26
	Computed	9	1					10
AV	Collected	1	1				4	6
	Basic							
	Computed							
Map	Collected							
	Basic				2			2
	Computed							
Total		75	28	6	59	19	4	191

집계데이터는 수집데이터와 기초데이터를 기반으로 구간 속도, 교통량 등을 단순집계, 일, 월, 년 등 특정기간 동안 집계 및 통계를 작성한 자료이다. 집계자료는 보다 세부적으로 도로 분석 구간에 따라 5분 단위 등 특정 시간단위 동안 단순 집계를 기록한 집계데이터와 1시간 교통 정보 등 집계자료를 기반으로 통계를 작성한 통계자료, C-ITS의 경우 사고 또는 돌발 상황에 따라 위험 구간과 지속시간 등을 예측한 빅데이터 분석자료로 세분할 수 있다.

수집되는 테이블을 주제별로 분류하면 차량 운행기록, 안전/사고 관련 자료, VMS (Variable message signs) 등 정보 제공, 시설정보 및 시설의 위치·지도 정보, 버스 스케줄 등 버스 운행과 관련된 자료, 차량 센서 데이터의 6가지로 구분할 수 있었다. 시설/위치/지도 데이터와 운영 자료는 Table 2의 기초자료에 해당한다. 이들 자료는 대부분 공통적으로 사용되어 별도의 주제로 분류하였다. 운영자료는 다른 DB에는 존재하지 않고 BIS 자료에만 존재하는 분류로 버스 노선, 정류장, 요금정보 등 버스 운영에 관련된 자료로 별도로 분류하였다. 이들 자료는 기본적으로 C-ITS DB에서도 큰 수정이 없이 기존 자료를 바로 사용할 수 있는 자료들이다.

주제별로 구분한 수집 테이블을 살펴보면 전반적으로 차량 운행을 기록한 자료가 다수를 차지하고 있다. 그러나 시설/위치/지도와 운영자료 등 운행 자료를 활용하기 위한 부가적인 자료가 많은 것을 알 수 있다. 이는 교통관련 자료에서 단순히 차량의 운행을 기록한 자료를 활용하기 위해서는 그에 적합한 노드와 링크, 정류장 등 위치정보와 버스 운영 자료의 역할이 중요하다는 것을 보여주고 있다.

Table 2에서 정리한 DB 테이블의 주제별 구분은 보다 세부적인 주제로 구분할 수 있다. 예를 들어 운행기록은 OBU, RSE, DTG 등으로 구분할 수 있다. 그러나 이 역시 각각의 DB가 처음 테이블을 구상하는 단계에서 구분한 분류이며, OBU는 차량에 설치된 장비이고 RSE는 도로변에 설치된 장비이므로 사실상은 동일한 자료라 할 수 있다. 이와 같이 단순히 테이블명만으로는 같은 종류인지 다른 종류인지 구분하기 어려운 경우가 있으므로 데이터 카탈로그를 활용하여 통일시킬 필요가 있다.

실질적인 데이터의 활용을 위해서는 운행데이터와 연계된 링크 자료 또는 지역 지도, 버스의 경우 운영자료 등이 동시에 제공될 수 있어야 한다. 이는 현재의 단순한 분류방법으로는 사실상 어렵다. 따라서 수집자료와 통계자료, 기초자료간의 상호 연관성을 살펴볼 수 있는 카탈로그 분류 방법이 필요하다. 이는 DCAT에서 제공하는 속성이나 Relationship 클래스를 적절히 활용함으로써 가능하다. 다음 섹션에서는 C-ITS를 활용하여 DCAT 기반 데이터 카탈로그를 작성하는 법을 제시하고자 한다.

5. 교통데이터를 위한 DCAT 기반 데이터 카탈로그 표준

교통데이터는 Table 2의 사례에서 보듯이 노드(node), 링크(link)와 같은 맵핑(mapping) 자료와 지도자료 등 기초자료와 수집자료를 가공하여 집계 및 통계를 도출한 집계 자료의 비중이 크다. 따라서 교통데이터를 분류할 때는 이러한 자료를 모두 고려하여 분류해야 하고, 적합한 카탈로그 작성 표준을 마련해야 한다. DCAT 2.0의 기본구조는 Dataset, Catalog, CatalogRecord, Distribution, DataService, Resource의 6개 핵심 클래스와 Location, Relationship 등의 부가 클래스로 구성되어 있다. 그러나, DCAT에서 제공하는 Location과 Relationship 클래스는 교통데이터를 위해서는 부족한 점이 많다. 따라서 본 표준에서는 Location 클래스는 개선하고, Relationship 클래스는 교통데이터간의 연관관계를 명확히 할 수 있도록 하위 속성을 변경하였다. 또한 교통데이터를 세세히 구분하기 위해 기존에 Dataset에 포함되어 데이터를 주제별로 분류하는 역할을 수행하던 dcat:theme을 Taxonomy 클래스로 분리하였다. 변경되거나 새롭게 제안된 클래스의 하위 속성은 모두 이미 활용되고 있는 표준을 준용하여 사용하였다. 기존 표준을 활용함으로써 DCAT 표준을 최대한 충실히 포용함으로써 이미 활성화되고 있는 표준과의 충돌을 최소화하기 위해서이다.

Figure 3은 DCAT-Trans의 구성에 대한 스키마를 표현하고 있으며, Table 3은 소속된 속성을 설명하고 있다. Dataset 클래스를 중심으로 5개의 클래스가 연결되는 구조이다. Dataset은 일반적으로 포함하는 데이터명, 작성일, 갱신일, 수집주기, 관리기관 및 관리자, 키워드, 핵심 주제(primary theme), 데이터 공간정보와 시간정보로 구성되어 있다. Dataset에 포함된 주제와 공간정보는 데이터를 대표하는 주제와 공간정보이며 세부적인 사항은 Taxonomy와 Location 클래스에서 보다 상세하게 기록된다.

Catalog 클래스는 Dataset 클래스에 포함되는 데이터들을 관리하는 역할이다. 예를 들어 Dataset이 1시간 단위 교통량 집계자료라고 하면, 하위 데이터는 년도별 또는 시간별, 요일별 등으로 존재할 수 있다. 이 데이터는 모두 동일한 내용을 포함하고 있으며 단지 년도, 시간, 요일 등이 다를 뿐이다. 따라서 동일한 데이터 셋으로 묶고, Catalog에서 년도, 시간, 요일 등으로 구분한 데이터를 관리한다.

Distribution 클래스는 데이터를 제공하는 방법을 정의하고 있다. DCAT 표준을 가장 변경 없이 수용하고 있는 클래스이다. 데이터의 사용권한은 dct:right와 dct:license로 규정하고 있다. 그리고 데이터 제공 포맷, 다운로드 URL, 제공데이터의 공간적, 시간적 범위를 규정하고 있다. 데이터 유통과 공유의 최근 추세는 하나의 데이터라도 다양한 포맷으로 제공하거나 여러 가지 방법으로 제공하는 것이 보편적이다. 따라서 동일한 Dataset이라도 제공 포맷이나 방법에 따라 각각의 Distribution 클래스가 정의 되어야 한다. 따라서 data_identifier를 사용하여 원천자료가 어디인지 구분하고 있다.

Taxonomy 클래스는 기존 DCAT의 Dataset에서 관리하던 데이터의 주제(theme)를 별도로 관리하기 위해 새로 만든 클래스이다. Taxonomy 아래에 다수의 dcat:theme 속성을 작성하여 데이터를 세부적으로 구분하여 관리할 수 있다. 본 연구에서는 데이터 구분을 위해 총 5개의 dcat:theme을 사용할 것을 제안하고자 한다. 그러나 5개의 dcat:theme을 모두 사용할 필요는 없고 필요에 따라 줄어든 수도 늘어날 수도 있다.

데이터의 주제는 크게 2가지 방식으로 진행된다. 첫 번째 구분 방식은 데이터의 활용 측면에서는 구분이다. 데이터의 활용 측면의 구분은 크게 3가지 유형으로 분류한다. 첫 째, 운행기록 등 수집된 원자료 또는 오류 수정정도의 단계를 거친 자료 등의 수집자료, 둘째, 통행 데이터의 활용을 위한 필요한 지도자료, RSU 위치 자료 등의 기초자료(맵핑자료), 셋 째, 자료의 분석결과인 집계자료와 통계자료이다. dcat:theme 1과 2가 여기에 해당하며 dcat:theme 1은 필수, 2는 옵션이다.

두 번째 구분 방식은 실제 데이터의 내용에 따라 분류하는 것이다. 현 연구단계에서는 크게 운행기록, 안전/사고, VMS 등의 정보자료, 시설/위치/지도, 차량 센서 등으로 대분류를 하고 세부적으로 버스, 승용차 등의 차량의 유형, 장비 유형 등으로 분류, 마지막은 DTG, OBU 등 보다 상세한 데이터의 기록 유형을 부여한다. dcat:theme3과 4, 5가 해당하며, dcat:theme3은 대분류, dcat:theme4는 세분류(장비 분류), dcat:theme5는 데이터의 형식이다. 예를

들어 설명하면 버스 DTG를 예로 들어 설명하면, dcat:theme 1은 기초자료, dcat:theme 3, 4, 5는 각각 운행기록, 버스, DTG 로 구분한다.

Figure 3. Schema of DCAT for transportation data

Table 3. Classes and properties of amended DCAT for transportation data

Class	Property	Definition*	Remark
Dataset	dct:title	A name give to the item	
	dct:description	A free-text account of the item	
	dct:issued	Dataset publication date	
	dct:modified	Dataset revision date	
	dct:identifier	A unique identifier of the item	
	dcat:keyword	Keyword or tag describing the dataset	
	dct:language	A language of the item	
	dcat:contactPoint	Contact information of persons or agent managing the dataset	
	dcat:theme (primary)	key theme of the dataset	
	dcat:landingPage	A web page to gain access to the dataset	
	dct:publisher	data publisher	
	dct:temporal	Temporal period that the dataset covers	
	dct:spatial	Geographical area covered by the dataset (Lv1)	
	dct:accrualPeriodicity	The frequency at which the dataset is published	
	dcat:distribution	An available distribution of the dataset	
	dcat:spatialResolutionInMeters	Minimum spatial separation resolution in the dataset (in meters)(Satellite image in 30m×30m)	
	dcat:temporalResolution	Minimum time period resolution in the dataset	
	prov:wasGeneratedBy	Activity of the creation of the dataset	
Catalog	dcat:dataset (multiple)	A collection of data that is listed in the catalog	
	dcat:theme	A main category of the dataset	
	dct:hasPart	Items that are listed in the catalog	
Taxonomy	dcat:theme (multiple)	Detailed categories of the dataset	New class, subdividing and managing themes in details
Location	locn:geometry	Geographical boundary of the dataset	
	dcat:bbox	Geographical bounding box of the dataset	
	dcat:centroid	Geographical centroid	
	w3cgeo:lat	Value of latitude	
	w3cgeo:long	Value of longitude	
	w3cgeo:alt	Value of altitude	
	locn:adminUnitL1	Administrative area at national level	Added for administrative areas
	locn:adminUnitL2	Administrative area at si, do	
	locn:adminUnitL3	Administrative area at si, gun, gu	
	locn:adminUnitL4	Administrative area at eup, myeon, dong	

Table 3. Classes and properties of amended DCAT for transportation data (continued)

Class	Property	Definition*	Remark
Relationship	dct:requires	Datasets which the dataset requires (RSU data for OBU data)	Re-structured, linking related data
	dct:isrequiredby	Datasets which the dataset is required by (Bus DTG for Bus route data)	
	foaf:primaryTopic	Theme of the related data (dcat:theme 1)	
Distribution	dct:title	A name given to the distribution.	
	dct:identifier	A unique identifier of the item	
	dct:description	A free-text account of the distribution	
	dct:issued	Date of formal issuance of the distribution.	
	dct:modified	Most recent date on which the distribution was changed, updated or modified.	
	dct:license	A legal document under which the distribution is made available	
	dct:rights	Information about rights held in and over the distribution.	
	dct:format	The file format of the distribution	
	dcat:byteSize	The size of a distribution in bytes	
	dcat:accessURL	A URL of the resource that gives access to a distribution of the dataset	
	dcat:downloadURL	The URL of the downloadable file in a given format	
	dcat:mediaType	The media type of the distribution (image, file etc)	
	dcat:spatialResolutionInMeters	Minimum spatial separation resolution in the distributed dataset (in meters)	
	dcat:temporalResolution	Minimum time period resolution in the distributed dataset	

note: Bold: Newly suggested classes and properties.

*Definitions are made based on DCAT 2.0.

Location 클래스는 데이터의 공간적 위치를 정의하는 클래스이다. DCAT에서 표준으로 제공하는 하위 속성은 다각형(polygon)이나 포인트(point)형태의 공간의 범위를 WGS 84좌표를 이용하여 표시하도록 하고 있다. 그러나 우리나라는 다수의 교통 데이터가 행정구역 단위로 수집 및 관리되고 있다. 따라서 행정구역을 표시할 수 있는 속성을 추가하였다. 추가된 locn:adminUnit은 Location Core Vocabulary로 world Wide Web Consortium (W3C)산 하의 Location and Address Community group에서 작성 중인 데이터의 위치와 주소 등을 표현하기 위한 언어 표준이다(W3C, 2019b). Location 클래스의 locn:geometry, dcat:bbox 등의 활용하여 데이터의 공간적 범위를 좌표로 명확히 규정하면 기술적으로 서로 다른 데이터를 동일한 위치를 기준으로 추출하는 방법의 기초가 된다.

Relationship 클래스의 경우 DCAT에 존재하는 Relationship 클래스는 추가하는 목적이 다르다. DCAT의 Relationship 클래스 아래의 2가지 속성인 dct:relation과 dcat:hadRole를 사용하여 연관된 데이터와 역할을 정의하고 있다(W3C, 2019a). 반면 본 연구에서 제안하는 Relationship은 하위 속성으로 dct:requires, dct:isrequiredby, foaf:primaryTopic을 사용한다. dct:requires는 A라는 데이터셋(또는 데이터)에 B라는 데이터가 필요하다는 것을 의미하고, dct:isrequiredby는 B라는 데이터는 A라는 데이터에 의해 필요하다는 것을 의미한다. 다시 말해, dct:requires는 A가 DTG 자료와 같은 수집자료일 경우 B라는 링크자료 등이 관련 자료로 필요하다는 것을 명확하게 알려준다. 반면은 dct:isrequiredby는 B라는 링크자료가 A (DTG), C (RSU)와 같은 데이터를 활용하기 위해서는 필요한 자료라는 것을 알려준다. foaf:primaryTopic은 각각의 연관자료가 어떤 유형(수집, 기초, 집계)의 자료인지 구분하여 나타낸다. 이렇게 함으로써 보다 명확하게 데이터간에 어떤 관계로 연관이 되는지 보다 명확하게 정의할 수 있다.

결론

데이터의 활용의 중요성은 점점 더 강조되고 있다. 스마트시티와 같은 도시의 정보화 과정에서 교통분야의 비중은 중요하다. 대도시의 경우 이미 설치되어 있는 ITS와 BIS 시스템을 통해 매일 많은 양의 데이터가 생성되고 있다. 또한 C-ITS와 같이 유용한 데이터를 더 많이 생산할 시스템이 시험 및 실증 사업 중이며, 2021년부터 대도시권을 중심으로 적용될 예정이다(Korea Expressway Corporation, 2019).

현재 CAPTAIN (Connected and Automated Public Transit INnovation) 연구단에서 수행중인 사업은 C-ITS 기반의 교통환경에서 자율주행 버스의 운행을 실증하는 연구이다. 본 연구 역시 자율주행차량 데이터와 C-IPTS 데이터의 생산을 고려하고 있다. 이러한 다양한 데이터가 생산 및 공유되는 환경에서 데이터의 가치를 높이기 위해서는 결국 데이터 유통 및 가치 창조를 위한 생태계 조성이 필요하다.

빅데이터의 활용과 공유는 보건분야 등 타 분야에서는 이미 많은 연구가 이루어지고 있으며, 데이터를 체계적으로 관리 및 활용, 제공하기 위해 메타데이터 관리의 측면에서 연구가 이미 수행중이다. 교통분야는 이미 ITS 사업을 통해 데이터를 오랜 동안 활용해온 경험이 있음에도 불구하고 메타데이터와 데이터 카탈로그 분야에서는 미진한 경향이 있다. 데이터 개방은 이용자가 '실제 필요로 하는 데이터를 제공할 수 있는가?'가 중요하다. 이를 위해 데이터의 발견 가능성(discoverability)을 높여야 하며, 이는 표준화된 메타 데이터의 관리를 통해 실현할 수 있다(EC, 2017b).

본 연구는 자율주행과 C-IPTS 환경 하에서 지금보다 더 다양하고 양질의 데이터가 대량의 생산될 때 교통 데이터를 보다 효율적으로 활용하기 위한 방안으로 체계적인 카탈로그 관리가 필요함을 강조하고 있다. 이를 위해 데이터 카탈로그 작성 분야의 표준 언어인 DCAT을 활용하여, 교통 데이터의 특성에 적합한 카탈로그 작성 표준을 제시하였다. 본 DCAT-Trans는 기존 DCAT에서 공간특성, 데이터간의 관계, 데이터의 분류를 중점적으로 개선하기 위해 Location, Relationship, Taxonomy 클래스와 하위 속성을 제안하였다. 이 세 개의 클래스를 통해 교통 데이터의 특징인 데이터 활용에 필수적인 기초자료를 함께 관리할 수 있으며, 지역 및 위치별 조회를 위한 방안, 상세한 데이터 구분이 가능하다. 이는 데이터 관리 측면뿐만 아니라 이용자 입장에서 연관된 자료를 함께 검색하여 다운로드 받을 수 있는 편리함을 제공하고 있다. 또한 체계적인 데이터 관리는 여러 기관 또는 지역에서 수집되는 동일 또는 유사한 데이터를 미리 표준에 의해 메타데이터를 작성함으로써 보다 쉽게 통합할 수 있는 장점이 있다.

본 연구는 교통분야 데이터를 위한 데이터 카탈로그 작성 표준을 제시하는 방법을 제시하고 있으나, 아직 많은 부분에서 한계가 있다. 먼저 본 연구가 검토한 데이터는 다양한 교통 데이터 중 일부에 불과하다. 따라서 더 다양한 데이터를 고려할 필요가 있다. 또한 본 연구는 아직 개념을 제시하는 연구이며 실제 데이터베이스나 웹포털에 적용의 측면에서 발견되는 기술적 고려가 약하다. 마지막으로 본 연구에서 제안한 Taxonomy, Location, Relationship의 세 클래스의 정의와 소속 속성에 대한 부분에서 더 연구가 필요하다.

Taxonomy 클래스는 향후 연구에서 가장 우선해야 하는 연구이다. 데이터의 주제를 어떻게 분류할 것인지에 대한 연구는 결국 데이터 분류의 상하 관계에 대한 정의가 명확히 필요하다. 본 연구에서는 theme1과 2, theme 3,4,5에 서로 다른 유형의 분류 방식을 적용할 것을 제시하고 있다. 특히 theme 3,4,5는 서로 상하관계가 정해져야 하는 문제가 있다. 예를 들어 버스 DTG 자료의 경우 본문의 예에서는 운행자료, 버스, DTG 순으로 속성값을 정했지만, 실제 가장 중요한 속성을 무엇으로 결정할지 다소 불명확하다. 그리고 현재의 분류가 과연 방대한 교통 데이터에 모두 적용될 수 있는지에 대한 검증도 필요하다. 이를 위해서는 많은 데이터베이스와 데이터 유형을 분석하여 하나의 데이터가 몇 가지 주제로 분류될 수 있으며, 여러 데이터에서 공통성을 가질 수 있는 주제를 선정하는 연구가 필요하다.

Location 클래스에서는 행정구역별 구분은 적절하지만, 하위 구분이 어느 정도까지 상세해야 하는지에 대해서는 여러 사례를 종합하여 보완해야 할 부분이다. 일반적으로 읍면동 단위의 행정구역이 보편적이긴 하지만 많은 교통 데이터 분석은 읍면동 단위보다 작은 단위로 구분할 필요가 있기도 하다. 그러나 이런 세세한 구분이 과연 필요한 것인지에 대한 연구가 필요하다. Relationship 클래스 역시 데이터간의 연결성에 따라 다소 유동적이며, 이는 데이터

를 구분하는 방법, 즉 Taxonomy 클래스와 연관이 있다. 그리고 단순히 Relationship 클래스에 2개의 속성을 추가하여 모든 연결 관계를 구성할 수 있는지에 대한 연구가 필요하다. 다시 말해 연결되는 데이터에 대한 추가적인 정보를 포함해야 보다 명확하게 연결 관계를 정의할 수 있는지 연구할 필요가 있다. 그러나 본 연구는 교통분야 데이터의 분류와 메타데이터 관리, 카탈로그 작성법의 표준화라는 앞으로 데이터 관리와 활용 측면에서 반드시 필요한 분야라는데 그 의의가 있다.

Funding

This work was supported by the Ministry of Land, Infrastructure, and Transport (MOLIT, KOREA) [Project ID: 19TLRP-B146733-02, Project Name: Connected and Automated Public Transport Innovation (National R&D Project)].

ORCID

SHIN, Doh Kyoum <http://orcid.org/0000-0003-3381-9808>

LEE, Sang Hwa <http://orcid.org/0000-0002-3477-3912>

KANG, Junghyun <http://orcid.org/0000-0002-4282-8362>

PARK, Eun Mi <http://orcid.org/0000-0002-4747-9587>

References

- EC (2017a), C-ITS Platform: Final report Phase II, European Commission, EU, 8, 58.
- EC (2017b), Re-using Open Data: A study on companies transforming Open Data into Economic & Societal Value, European Data Portal, 77.
- Elkosantini S., Darmoul S. (2013), Intelligent Public Transportation Systems: A Review of Architectures and Enabling Technologies, 2013 International Conference on Advanced Logistics and Transport, ICALT 2013, Sousse, Tunisia, DOI: <http://dx.doi.org/10.1109/ICAdLT.2013.6568465>.
- Jung U. (2019), Standards of Metadata Catalogue and Service Structure for Big Data Trades (빅데이터 유통을 위한 카탈로그 메타데이터 표준 동향 및 서비스 구조), Journal of TTA, 182, 22-27.
- KEC (2019), History of C-ITS Pilot Project (시범사업 연혁), C-ITS Pilot Project PR Hall, Korea Expressway Corporation, <https://www.c-its.kr/introduction/history.do>, 2019.9.8.
- Maali F., Cyganiak R., Peristeras V. (2010), Enabling Interoperability of Government Data Catalogues, 9th IFIP WG 8.5 International Conference on Electronic Government(EGOV), Aug 2010, Lausanne, Switzerland, 339-350. DOI: http://dx.doi.org/10.1007/978-3-642-14799-9_29.
- Martin S., Foulonneau M., Turki S. (2013), 1-5 Stars: Metadata on the Openness Level of Open Data Sets in Europe, Metadata and Semantics Research: 7th Research Conference, MTSR 2013, Thessaloniki, Greece, November 19-22, 2013 Proceedings, 234-245. DOI: http://dx.doi.org/10.1007/978-3-319-03437-9_24.
- MOIS (2018), Government Innovation with Opening Core Transport Data (교통분야 핵심데이터 개방을 통해 정부혁신 추진한다), Ministry of the Interior and Safety, Korean government press release.
- Neumaier S., Umbrich J., Polleres A. (2015), Automated Quality Assessment of Metadata across Open Data Portals, Journal of Data Information Quality, 8(1), Article No.2. DOI: <http://dx.doi.org/10.1145/2964909>.

- NIA (2014), Data Technology Guideline 1: RDF (데이터활용 신기술 가이드 1: RDF), NIA X-RER-14012, National Information Society Agency, 10-11.
- NIPA (2014), Analysis on Status of Global Strategies of Open Data (주요국의 오픈 데이터 정책 추진 현황 분석), Global ICT R&D Policy Trends, 2014-02, National IT Industry Promotion Agency, 5-29.
- Rossi T. (2019), Autonomous and ADAS Test Cars Produce over 11 TB of Data Per Day, TUXERA, <https://www.tuxera.com/blog/autonomous-and-ad-as-test-cars-produce-over-11-tb-of-data-per-day/>, 2019.09.19.
- W3C (2014a), RDF 1.1 Primer, W3C, <https://www.w3.org/TR/rdf11-primer/>, 2019.10.19.
- W3C (2014b), DCAT Implementations, ISA Programme Location Core Vocabulary, W3C wiki page, https://www.w3.org/2011/gld/wiki/DCAT_Implementations, 2019.9.20.
- W3C (2019a), Data Catalog Vocabulary(DCAT) - Version 2, W3C, <https://www.w3.org/vocab-dcat-2/>, 2019.9.8.
- W3C (2019b), ISA Programme Location Core Vocabulary, W3C, <https://www.w3.org/ns/locn>, 2019.9.8.
- Wells D. (2019), What Is a Data Catalog?, Alaton, <https://www.alation.com/blog/what-is-a-data-catalog/>, 2019.9.10.