

오픈 데이터 유통 참조모델과 데이터 유통 플랫폼 기술



원희선_한국전자통신연구원(ETRI) 책임연구원

1. 머리말

데이터가 가치를 창출하는 데이터 경제의 부상과 함께 모든 산업 분야에서 데이터에 대한 관심과 수요가 급속히 증가하고 있다. 이에 국내외 공공기관뿐만 아니라 민간 기업에서도 데이터에 대한 개방을 확대하고 있으며, 데이터 기반의 새로운 비즈니스 모델을 기획하고 실험할 수 있는 데이터 유통 플랫폼을 구축하고 있다.

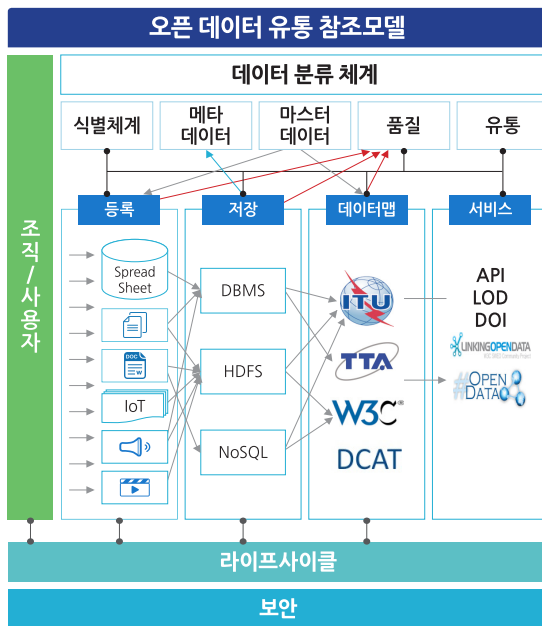
그러나 데이터 활용도는 개방 속도에 비해 저조한 편인데, 주요 원인 중에는 대부분의 기관들이 독자적으로 데이터를 수집·관리하고 있어서 상호 간 데이터 소재 파악 및 품질과 신뢰성 확보가 어려운 점 등을 꼽을 수 있다. 따라서 양질의 데이터를 개방하고 공유하기 위한 표준화된 관리체계와 상호운용 지침이 필요하다.

CKAN(Comprehensive Knowledge Archive Network)은 데이터 개방을 지원하기 위해 영국 비영리 단체인 OKF(Open Knowledge Foundation)에 의해 개발된 오픈소스 기반의 개방형 데이터 플랫폼이다. CKAN은 W3C(World Wide Web

Consortium) DCAT(Data Catalog Vocabulary) 표준을 기반으로 데이터 카탈로그를 배포하고 수집하는 하베스팅 기능을 제공하며, 영국, 미국, 캐나다, 호주 등 146개 정부 및 공공기관에서 CKAN을 기반으로 데이터 포털을 구축하여 각국의 데이터세트 정보를 상호 공유하고 있다[1][2].

그러나 IoT 확산과 함께 데이터의 종류와 형태가 더욱 복잡하고 다양해지며, 품질과 신뢰성에 대한 이슈가 커짐에 따라, 2017년부터 W3C DXWG(Data Exchange Working Group)에서 DCAT 규격을 확장하는 작업을 진행하고 있다. 국내에서는 정부과제를 통해 ITU-T SG 13/WP 2[3]에서 빅데이터 유통을 위한 메타데이터 프레임워크 및 개념 모델의 표준화 작업과 이와 연계하여 DCAT 등 국제표준과의 호환성을 고려한 오픈 데이터 유통 참조모델의 국내 표준안을 개발하고 있다.

본고에서는 오픈 데이터 유통 참조모델의 프레임 워크 개요와 참조모델의 핵심 구성요소인 오픈 데이터의 분류체계 및 메타데이터 관리에 대해 소개한다. 아울러 이를 기반으로 개발 중인 오픈 데이터 유통 플랫폼 기술을 소개한다.



[그림 1] 오픈 데이터 유통 참조모델 프레임워크

구분	내용 및 지침
데이터 분류체계	분류체계 및 카테고리 운영 관리
	국제표준 호환성 지원
조직/사용자	오픈 데이터 거버넌스 및 참조모델 관리
	오픈 데이터 커뮤니티 운영관리
	조직/사용자 관리, 역할 관리
식별체계	데이터 식별체계 운영 관리
	국제표준 식별체계 호환성
메타 데이터	메타데이터 관리
	국제표준 호환성
마스터 데이터	마스터데이터 관리
	국제/국내 마스터데이터 상호운용
데이터 품질	품질기준, 품질지표 및 품질평가 체계 관리
	데이터, 메타/마스터데이터 품질관리
보안	개인정보 보호, 비식별화
유통	라이선스, 가치산정, 유통경로/이력관리
라이프 사이클	등록~저장~활용~변경~폐기 관리
데이터맵	데이터맵 형식, 배포, 하베스팅 관리

2. 오픈 데이터 유통 참조모델

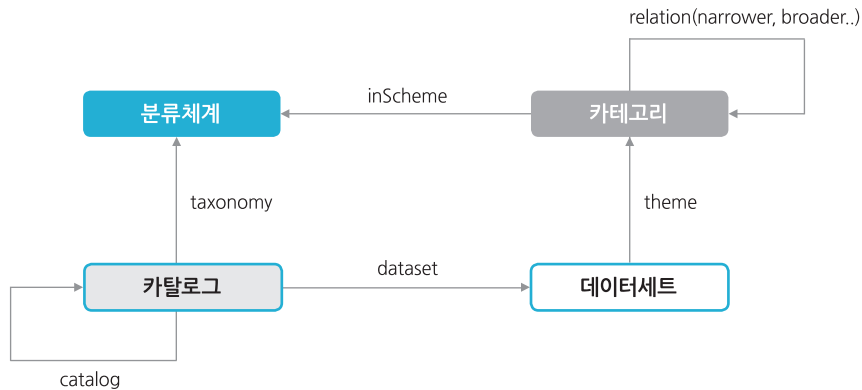
오픈 데이터 유통 참조모델은 공공, 민간에서 생성되는 방대한 데이터를 포괄하는 범용적 관리 체계와 데이터 연계·공유·활용성을 높이기 위한 기능 및 정책적 요소들로 구성된다. 또한 새로운 데이터 유형의 출현 증가를 고려한 참조모델의 확장과 운영관리 지침을 포함한다.

2.1 오픈 데이터 유통 참조모델 구성요소 및 운영 방안

오픈 데이터 유통 참조모델은 전사 아키텍처(EA)에서의 데이터 참조모델(DRM)과 유사성이 있으나 적용 대상의 특성과 관점에서 차이가 크다. 데이터 참조모델은 조직의 업무 지원을 위한 데이터 계층 표준화, 교환, 재사용 지원 등이 주 목적인 반면, 오픈 데이터 유통 참조모델은 외부의 불특정 다수가 대상인 개방된 환경에서 다양한 유형의 데이터를 체계

적으로 관리하여 검색·유통·활용을 효과적으로 지원하는 데 목적이 있다. 따라서 데이터의 유통 구조, 특징 분석 등을 통해 구성요소들이 도출되어야 하며, 시장 요구사항 변화에 따라 유통 참조모델의 진화도 필요할 것이다. [그림 1]은 주요 구성요소들을 포함하는 오픈 데이터 유통 참조모델의 프레임워크를 나타낸다.

참조모델의 각 구성요소는 확장이 용이한 유연한 구조로써 운영관리 지침과 함께 정의된다. 구성요소 중에서 데이터 공유·활용의 기초를 제공하는 분류체계와 식별체계, 이력관리 등은 관련된 상·하위 국제표준과의 호환성을 유지하여 국내외 데이터의 연계, 공유를 지원해야 한다. 또한 분류체계, 메타데이터, 품질, 보안 등의 각 구성요소의 실제 콘텐츠는 산업 및 응용분야에 따라 매우 다양하므로, 오픈 데이터 커뮤니티와 해당 분야 데이터 전문가 그룹 등을 통한 확장과 공유를 지원하는 시스템 및 운영환경이



[그림 2] 오픈 데이터 관리 모델

필요하다. 또한 이를 통해 공공 및 민간 기관에서 데이터 유통 플랫폼을 구축·운영하기 위한 실용적인 지침으로 활용될 수 있을 것이다.

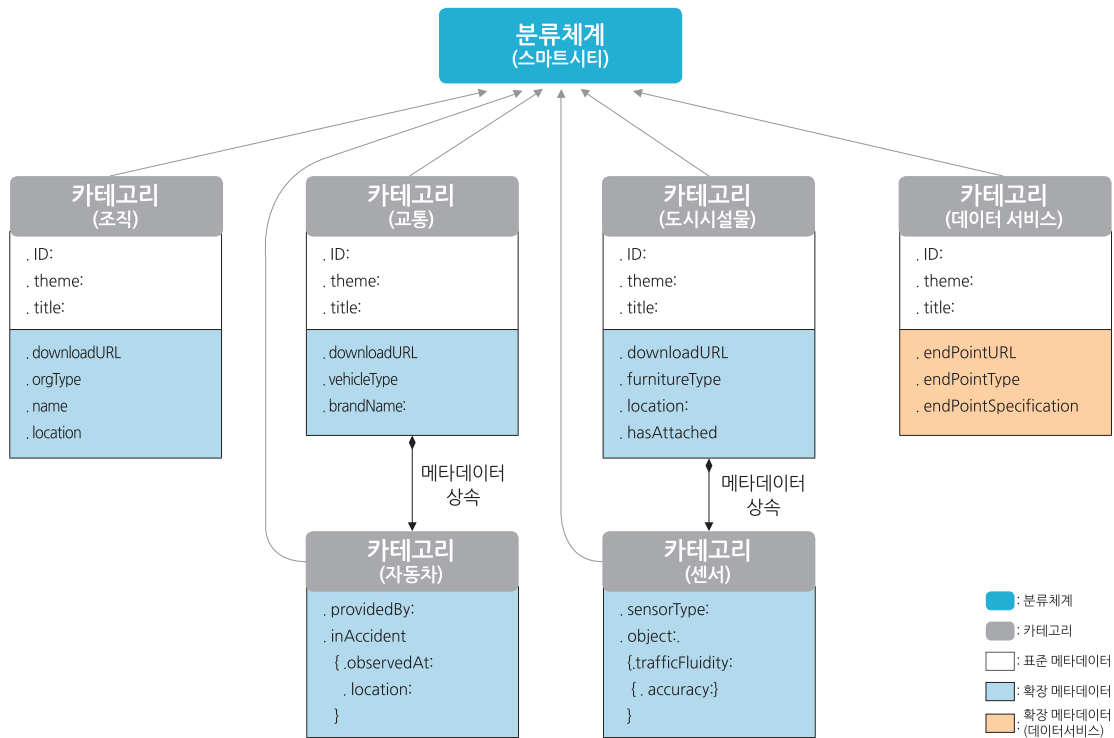
2.2 오픈 데이터 분류체계와 메타데이터 관리

오픈 데이터 유통 참조모델에서 데이터 분류체계는 산업 또는 응용 범위를 기준으로 유사하거나 연관된 속성을 지닌 카테고리들을 계층적 구조로 체계화한 것이다. [그림 2]는 W3C의 DCAT 모델에서 분류체계 중심의 데이터 관리 모델을 요약하여 나타낸 것으로, 데이터세트의 카탈로그는 분류체계와 매핑이 되고, 데이터세트는 분류체계에 속한 카테고리에 매핑이 되어 관리된다[2].

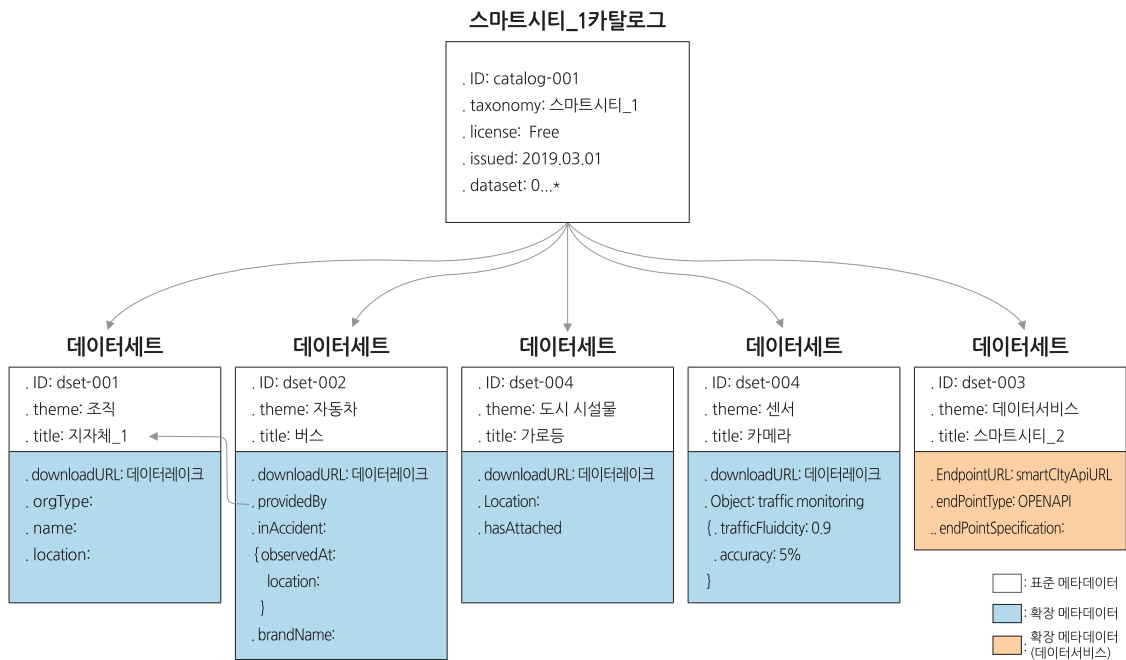
메타데이터는 데이터세트의 등록부터 검색, 유통, 활용, 품질관리 등의 전체 라이프사이클에 걸쳐 필요한 속성 정보를 제공하며, 오픈 데이터 유통 참조모델에서는 분류체계의 카테고리 단위로 메타데이터 스키마를 정의하도록 한다. 카테고리의 메타데이터 스키마는 DCAT에서 정의한 데이터세트의 표준 메타데이터 항목들과 상위 카테고리로부터 상속받은 메타데이터 항목들에 필요한 메타데이터 항목을 추가하여 정의된다.

[그림 3]은 스마트시티 분야를 대상으로 분류체계의 카테고리 및 각 카테고리의 메타데이터 스키마 구성 과정을 나타낸다[4]. 데이터세트 접근은 데이터 원본에 대한 URL 또는 Open API, LOD(Linked Open Data) 등과 같은 데이터서비스 URL을 통해 가능하며, 각 방식에 따라 메타데이터 항목이 구성된다. [그림 3]에서 ‘데이터서비스’ 카테고리는 다양한 데이터 서비스 Endpoint의 메타데이터 항목을 포함하고 있으며, [그림 4]는 [그림 3]의 분류체계와 카테고리를 기반으로 구축된 스마트시티 카탈로그와 데이터세트의 예시로서, ‘데이터서비스’ 카테고리에 외부 시스템에서 제공하는 Open API가 등록된 예를 보여준다.

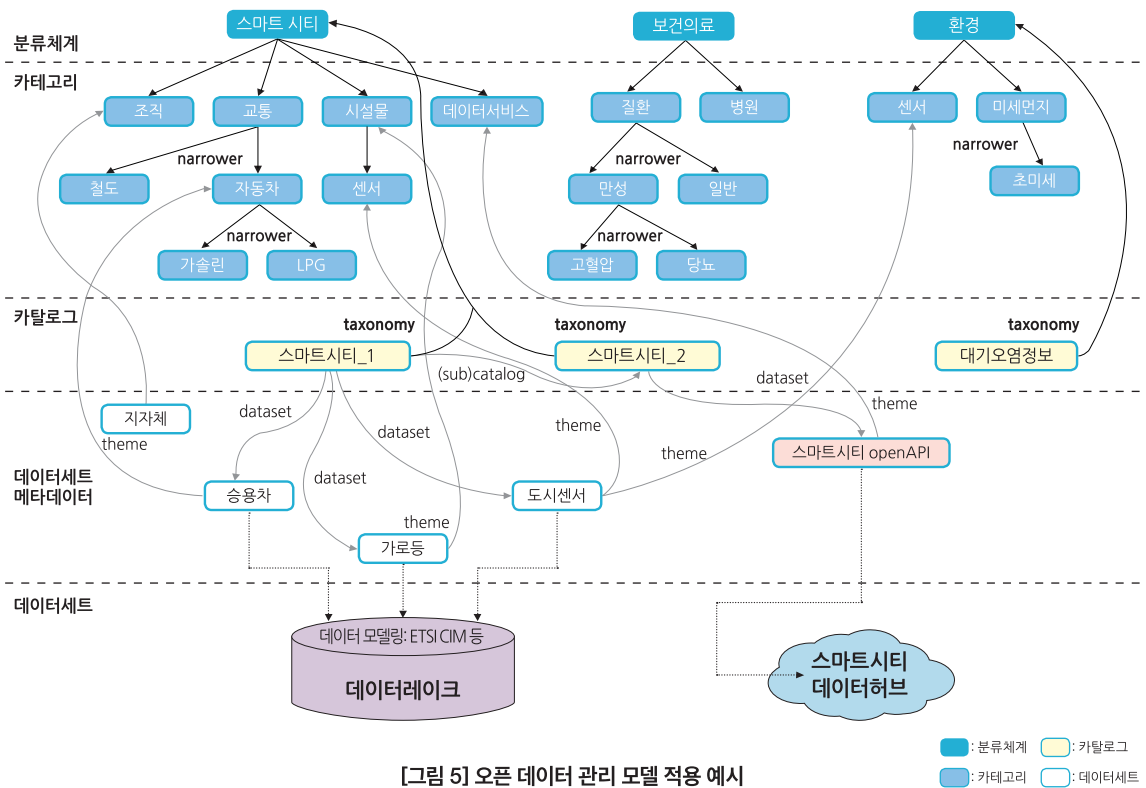
오픈 데이터 유통 참조모델은 다수의 분류체계를 가질 수 있으며, 분류체계 및 각 분류체계의 카테고리, 메타데이터 항목 등은 오픈 데이터 커뮤니티와 전문가 그룹 등을 통한 제안과 별도의 협의 및 승인 과정을 통해 확장되고 새로운 버전이 공유되도록 한다. 카탈로그는 데이터세트에 대한 정보를 제공하고 데이터 검색에도 사용되므로, 하나의 데이터세트가 복수의 분류체계 및 카테고리에 매핑이 가능하도록 하여 여러 응용 분야에서 활용될 수 있도록 한다.



[그림 3] 분류체계와 카테고리의 메타데이터 스키마 예시



[그림 4] 스마트시티 카탈로그 구성 예시



[그림 5]는 상기 내용에 따라 구성된 여러 분야의 분류체계, 카테고리, 카탈로그, 데이터세트의 메타데이터 간의 관계를 도식화하였으며, 데이터세트는 각 응용 분야의 특성과 표준에 따라 모델링되어 저장된 데이터레이크의 접근 URL을 제공하거나, 독립 시스템으로 구축되어 서비스 Endpoint를 제공할 수 있다.

3. 오픈 데이터 유통 플랫폼

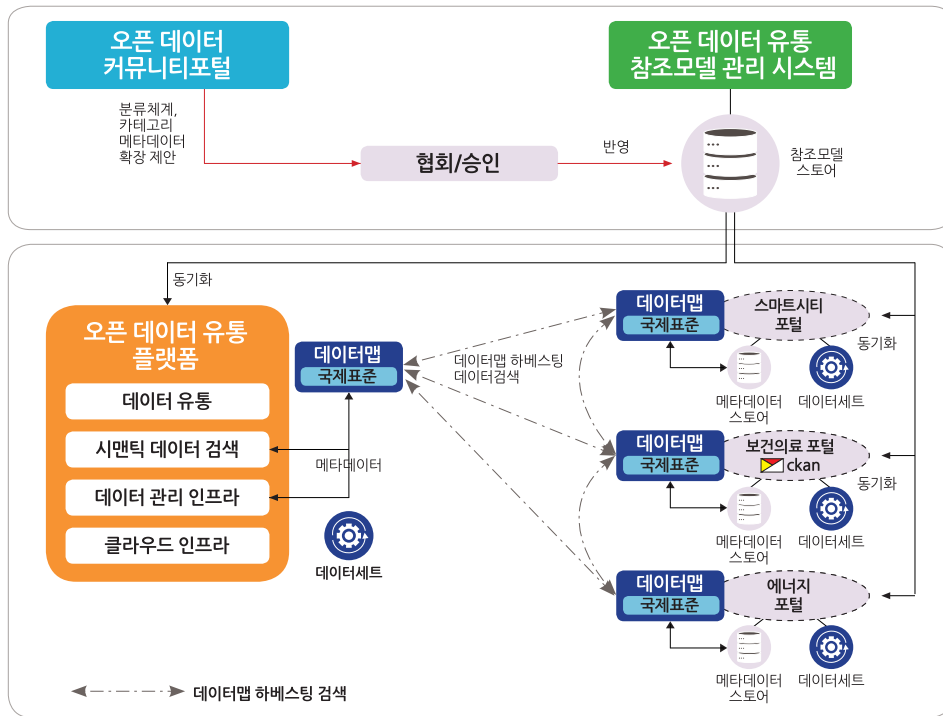
오픈 데이터 유통 플랫폼은 데이터 관리 인프라, 플랫폼 상호연동, 데이터 서비스 등의 세부 기술로 구성되며, 오픈 데이터 유통 참조모델은 각 세부 기술을 구현하는 중요한 요소로서 플랫폼의 실효성과 활용성을 좌우한다. 예를 들어 데이터세트의 메타데이터 관리 및 플랫폼 상호연동을 위한 데이터맵 생

성, 공유, 검색엔진 등은 분류체계와 카테고리의 메타데이터 스키마를 참조하여 구현된다. [그림 6]은 오픈 데이터 유통 참조모델을 기반으로 플랫폼에서 데이터맵을 생성하여 공유하는 개념도이다.

참조모델에서 제시하는 보안, 품질, 유통 등에 대한 기준 및 운영 지침의 반영을 위해서는 빅데이터, 클라우드, 블록체인 등의 최신 기술을 융합한 확장된 개념의 아키텍처가 필요하며, 이를 기반으로 구현된 플랫폼은 오픈 데이터 유통을 위한 거버넌스의 중심적 매개 역할을 담당하게 될 것이다[5].

4. 맺음말

오픈 데이터 유통 참조모델은 공공기관 및 민간기업에서 오픈 데이터 유통 플랫폼을 구축·운영하기



[그림 6] 오픈 데이터 유통 플랫폼의 데이터맵 상호운용 개념도

위한 실용적인 지침과 적용 기준을 제공한다. 이러한 역할의 실효성을 높이기 위해서는 플랫폼 구현에 반영되고, 잘 통제된 관리 하에 요구사항을 반영하고 확장되어야 참조모델의 의미를 유지하며, 오픈 데이터 유통 플랫폼의 활용 가치를 지속적으로 제고할 수 있을 것이다.

데이터 개방 추세에 따라, W3C DCAT, ITU-T SG 13 등의 국제표준 워킹그룹에서 데이터 관리 모델의 고도화와 표준화를 진행 중이며, 이와 병행하여 국내에서는 국제표준 모델과 연관되는 참조모델의 분류체계와 메타데이터 등의 콘텐츠 생성, 확장 및 운영관리 지침의 표준화를 진행 중에 있다. 참조모델의 상기 구성요소들은 데이터맵 생성에 필요한 정보를 제공하므로, 이를 플랫폼에 우선 적용하면 데이터 연계, 활용을 촉진하여 새로운 용

복합 서비스 개발에 기여할 것으로 기대된다. TTA

※ 본 연구는 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(2017-0-00253, 국제표준 기반 오픈 데이터 유통 플랫폼 확장 기술 개발).

※ 본 연구는 산업통상자원부와 한국산업기술진흥원의 '국제공동기술개발 사업'의 지원을 받아 수행된 연구결과임.

[참고문헌]

- [1] CKAN(Comprehensive Knowledge Archive Network) <https://ckan.org/>
- [2] DCAT(Data Catalog Vocabulary) <https://www.w3.org/TR/vocab-dcat>
- [3] ITU-T, Draft Recommendation ITU-T Y.bdm-sch: Big data - Metadata framework and conceptual model 2018.
- [4] 정승명, '스마트시티 데이터 모델 표준화 동향', TTA Journal Vol.176, 2018.03.
- [5] 'Implementing the Data Hub: Architecture and Technology Choices', Gartner, 2018.08.