국가데이터맵의 개념 및 모델



김학래 _ 중앙대학교 교수

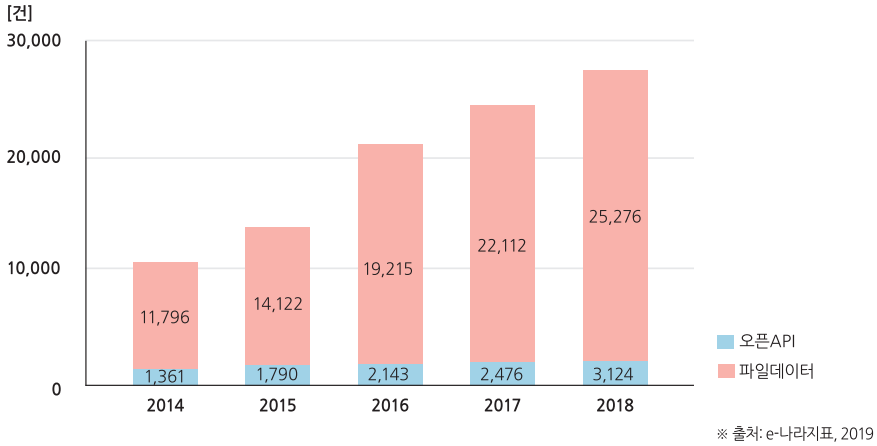
1. 머리말

대한민국 정부는 공공기관이 보유하고 있는 공공정보를 '공공데이터'라는 이름으로 민간에 개방 및 제공하고 있다. 공공데이터는 공공기관이 직무상 전자적 방식으로 처리·작성·취득하여 관리하고 있는 전자적 방식으로 처리된 자료이다. 정부는 「공공데이터의 제공 및 이용활성화에 관한 법률」(약칭: 공공데이터법)을 제정(2013. 7월) 및 시행(2013. 10월)하고[1], 공공데이터 정책을 적극적으로 추진하고 있다. 실제 이 법률에 근거하여 중앙부처, 지자체, 산하기관 등 780여 개 각 공공기관은 공공데이터제공책임관 및 실무자를 지정하고 있다. 또한, 민간에서 공공데이터의 활용을 지원하기 위해 공공데이터활용지원센터와 공공데이터 관련 정책을 심의·조정 및 추진사항을 점검하기 위해 공공데이터전략위원회가 각각 설립되었다. 공공데이터제공분쟁조정위원회는 공공데이터 제공거부나 중단 문제의 해결을 지원하고 있다.

공공데이터의 개방은 대한민국 정부, 지방자치단체 및 공공기관이 보유하고 있는 공공데이터를 이용자가 자유롭게 활용할 수 있도록 다양한 형태로 개방·제공하는 것을 의미한다[1]. 이때 개인정보보호, 제3자의 권리가 포함된 정보 및 국가 안보와 관련된 정보를 제외한 모든 데이터는 상업적·비영리적으로 이용할 권한이 부여된다. 공공데이터 개방은 범정부 차원에서 추진되고 있으며 2013년 이후 지속적으로 증가하고 있다. [그림 1]에서 보듯이 공공데이터 개방 현황[1]은 오픈 API와 파일 데이터가 각각 3,124건, 25,276건이고(2018년 기준), 2019년에도 지속적으로 데이터 개방이 진행되고 있다.

공공데이터법 시행 이후, 데이터 개방은 폭발적으로 증가하고 있고 데이터 활용을 위한 요구도 확대되고 있다. 경제협력개발기구(OECD) 개방지수 평가에 의하면 우리나라는 2회 연속 1위로 평가되었다. 그러나 개방된 데이터의 활용이 제한적이고 요구하는 데이터를 적시에 제공하지 못하는 한계가 존재하는 것이 사실이다. 이를 해결하기 위해 공공기관이 보유하거나 개방하고 있는 데이터 현황을 파악하기

1) 공공기관이 보유하고 있는 공공데이터를 개인 혹은 기업이 서비스 개발 등에 활용하기 위하여 범정부 제공창구인 공공데이터포털(data.go.kr)을 통해 개방하고 있는 수치



[그림 1] 공동데이터 제공통계

위해 공공데이터 보유현황 전수조사가 실시되었다[2]. 전수조사는 공공기관에서 생성, 취득해 운영중인 모든 정보시스템의 데이터베이스를 대상으로 정보시스템별 기본 조사(조직, 예산 등), 데이터베이스의 물리적 정보(데이터베이스명, 테이블명, 칼럼명)를 포함한다. 전수조사를 통해 도출된 결과는 공공데이터 개방 확대를 위한 기초자료로 활용될 예정이다[2]. 더불어 범정부적 데이터의 용어, 형식 등 표준화, 품질관리 등의 기반 인프라 구축을 위한 범정부 데이터 플랫폼 구축사업을 추진하고 있다. 본고에서는 범정부 데이터 플랫폼의 핵심인 데이터맵에 대한 개념을 소개하고 향후 개선 방안에 대해 논의한다.

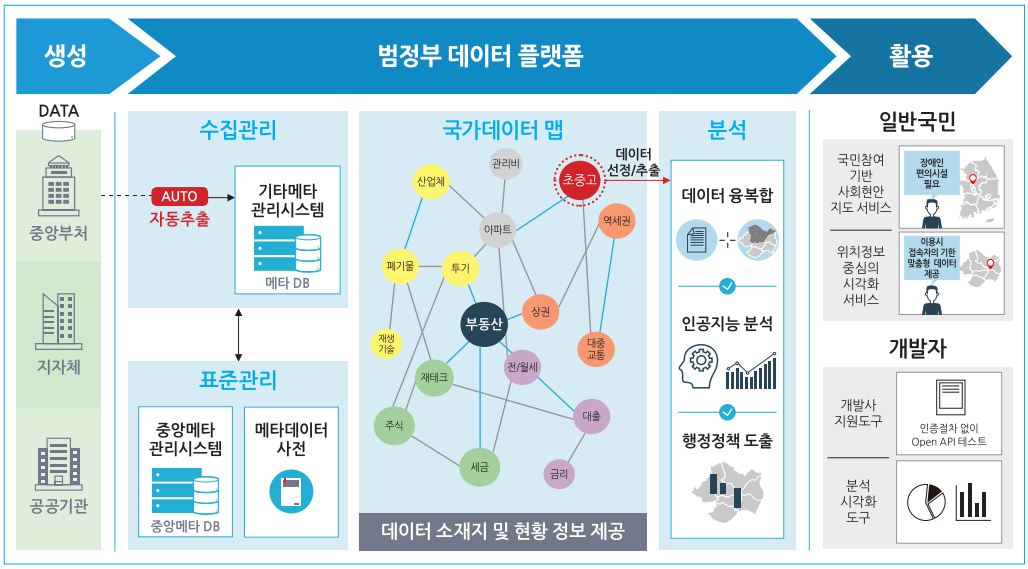
2. 범정부 데이터 플랫폼의 개요

범정부 데이터 플랫폼은 공공기관이 보유한 공공데이터의 소재와 메타데이터 정보를 통합·관리하기 위한 목적으로 추진 중인 범정부 사업이다. 공공기관이 활용하는 모든 메타데이터는 기관메타관리시스템으로 통합되어 관리되고, 범정부적 활용을 위해 메타데이터 정보는 용어, 형식을 표준화하여 메타데이터 사전으로 관리한다. 범정부 데이터 플랫폼에 수집된 모든 공공데이터는 국가데이터맵을 통해 데이터의 소재 정보 및 연관관계를 시각화된 형식으로 제공된다.

범정부 데이터 플랫폼의 주요 구축 목표는 다음과 같다[3].

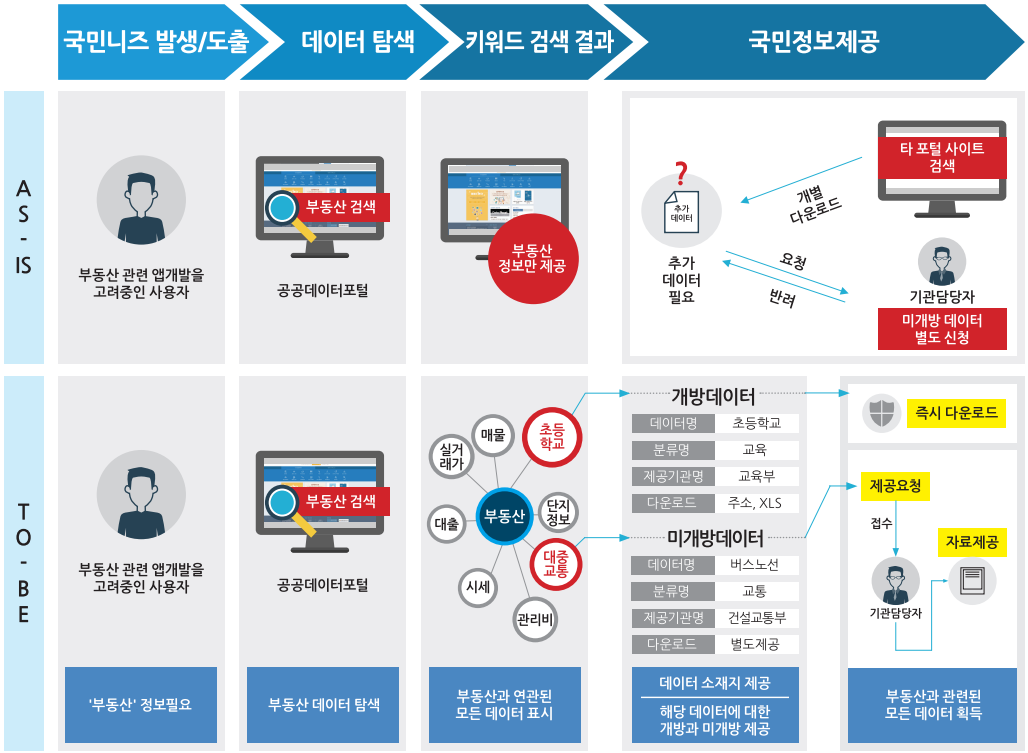
① **데이터 플랫폼 구축:** 메타데이터 통합관리 및 국가데이터맵 제공, 공공데이터 제공 체계 기반 마련
② **메타데이터 수집 체계 구축:** 개별 기관이 보유한 데이터베이스에서 메타데이터를 자동 추출·전송하기 위한 에이전트 개발 및 메타데이터 수집 에이전트 설치
③ **메타데이터 관리체계 구축:** 범정부 차원의 지능화된 메타데이터 관리를 위한 중앙메타데이터관리 시스템 및 기관메타데이터관리시스템 구축 및 적용
④ **국가데이터맵 구축:** 중앙메타데이터관리시스템의 보유 데이터, 공공데이터포털(data.go.kr)의 개방 데이터 등 데이터 연관 관계 및 연관도 파악을 위한 데이터맵 엔진 및 국가데이터맵 서비스 구축

범정부 데이터 플랫폼은 공공기관이 보유한 공공데이터의 소재 정보와 데이터 사이의 연관성 정보를 통해 데이터 보유 여부를 파악할 수 있다. [그림 3]을 살펴보면, 범정부 데이터 플랫폼은 기존의 단순 검색을 넘어 데이터의 연관 관계를 통해 데이터 활용 여



※ 출처: http://www.itdaily.kr/news/photo/201807/90256_98727_2414.jpg

[그림 2] 범정부 데이터 플랫폼 개념도



※ 출처: http://www.itdaily.kr/news/photo/201807/90256_98729_263.jpg

[그림 3] 범정부 데이터 플랫폼을 통한 변화

부를 검토하고 데이터 소재 정보에서 개방·미개방 여부를 확인할 수 있다. 기관담당자는 사용자의 개방 요청을 신속히 할 수 있으며, 범정부적 차원의 공공데이터 관리를 통해 데이터 기반 과학적 행정 지원이 가능해진다.

3. 국가데이터맵의 개념

범정부적 차원에서 공공데이터의 개방이 정책적으로 추진되고 있지만 정부가 보유하고 있는 데이터의 보유 현황을 파악하는 데는 한계가 있다. 개별 공공기관이 보유하고 있는 데이터 현황에 대한 자료가 존재하지 않았고, 기관별로 서로 다른 관리체계를 적용하고 있기 때문이다. 국가데이터맵은 공공기관이 보유하고 있는 공공데이터의 보유 현황과 소재 정보를 바탕으로 데이터 사이의 연관 정보를 제공하는 것이 목표이다. 데이터기반행정 활성화에 관한 법률은 데이터관리체계 마련을 위해 제4장(기반구축) 제12조에 다음과 같은 세부 조항을 정의하고 있다[4].

① 공공기관의 장은 생성·수집하여 보유하는 데이터에 대한 메타데이터 및 해당 데이터간의 연관 관계를 시각적으로 표현한 데이터관계도를 체계적으로 관리하여야 하며, 소관 데이터의 최신성·정확성 및 상호연계성이 유지되도록 하여야 한다.

② 행정안전부장관은 각 공공기관의 메타데이터 및 데이터관계도를 통합·연계한 시스템(이하 "중앙메타데이터관리시스템"이라 한다)을 구축·운영하여야 한다. 이를 통해 각 공공기관의 주요 정책 수립·추진에 필요한 데이터의 연관관계를 시각화 방법(이하 "데이터맵"이라 한다)으로 제공하여야 한다.

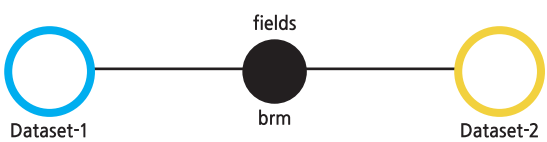
데이터맵은 데이터의 소재정보와 데이터 간 연관관계를 시각적으로 표현한 것이며, 데이터 사이의 관계성을 분석하는 데 활용할 수 있다[5]. 그러나 「데이터기반행정 활성화에 관한 법률」은 '데이터 관계'를 시각적으로 표현한 도구로 데이터맵의 개념을 한정하고 있다. 실제 데이터맵을 구성하는 대상에 대한 범위와 항목은 구체적이지 않다. 데이터맵이 데이터를 보유하고 있는 개별 공공기관에 적용되고, 국가 수준의 데이터맵으로 연결 및 확산되기 위해 데이터맵의 개념과 구성요소를 상세하게 정의할 필요가 있다.

4. 국가데이터맵의 개념 확장

국가데이터맵은 공공데이터의 소재 정보와 데이터 사이의 연결을 위한 데이터 구조, 데이터 구성요소 및 시각화를 포함하는 범정부 거버넌스 체계이다. 먼저 데이터맵의 논리 모델을 정의하면 다음과 같다.

$$DM = \langle D, A, R \rangle \quad \text{---------- (1)}$$

데이터맵 DM은 D, A, R의 관계로 정의한다. 이때, D는 데이터세트의 집합, A는 D가 보유한 속성의 집합이다. 데이터세트의 속성은 개별 데이터세트에 따라 다를 수 있지만, 구성항목은 공통적으로 정의할 수 있다. 예컨대 공공데이터의 보유기관, 데이터베이스명, 필드명은 데이터세트를 이해하기 위한 주요 요소이며 속성 집합에 포함되어야 한다. R은 D와 A의 이진 관계이다. 개별 데이터세트는 속성 정보를 통해 다른 데이터세트와 연관성을 만들 수 있다. [그림 4]와 같이 'Dataset-1'과 'Dataset-2'는 공통적으로 보유하고 있는 속성명 ('fields')과 정부기능분류체계에 대한 정보 ('brm')로 연결 관계를 만들 수 있다.



[그림 4] 속성정보를 이용한 데이터세트의 연결

이와 유사하게 개별 데이터맵은 상위 수준의 데이터맵과 상호 연결될 수 있다. 즉, 데이터맵은 다음과 같이 확장될 수 있다.

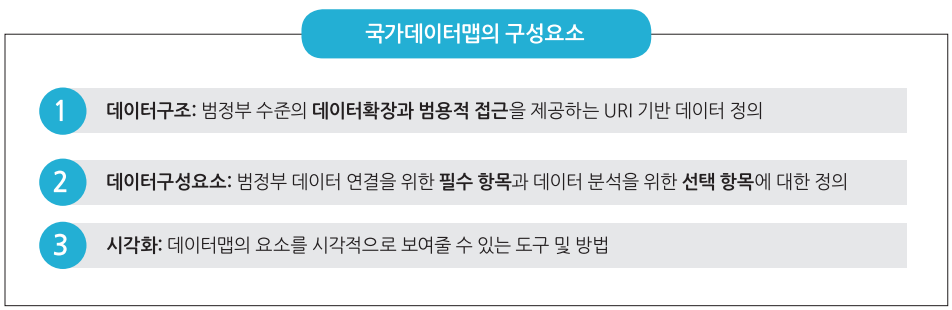
$$DM_G = \langle DM_1, DM_2, \ldots DM_n \rangle \text{---------- (2)}$$

개별 도메인의 데이터맵(DM_1, DM_2)은 서로 연결되어 메타 수준의 데이터맵(DM_G)을 구성할 수 있다. 즉, 데이터맵을 구성하는 모든 정보는 범정부 차원의 표준화된 관리체계로 관리되고, 이를 기준으로 생성된 모든 데이터맵은 국가데이터맵과 논리적인 연결관계를 함의한다고 정의한다.

아래 [그림 5]는 국가데이터맵의 구성요소와 기능을 요약하고 있다. 먼저 데이터맵의 모든 구성요소는 유일한 식별자를 부여할 수 있도록 URI(Uniform Resource Identifier) 체계를 적용하고, 데이터 사이의 연결을 위해 그래프(graph) 구조로 표현한다. 데이터 구성요소는 필수항목과 선택항목으로 구분할 수 있다. 먼저 필수항목은 데이터맵을 구성하는 핵심요소로써 보유기관명, 데이터베이스명, 테이블명, 필드(칼럼)명이 포함될 수 있다. 특히, 필드명은 데이터의 연결 관계를 파악하는 데 중요한 객체이며, 모든 항목이 정제되어 통제어휘(controlled vocabularies)로 정의되어야 한다. 선택항목은 메타데이터관리시스템에서 수집 및 관리하는 항목을 대상으로 한다. 마지막으로, 시각화는 데이터맵의 개별 요소를 시각적으로 보여주기 위한 방법으로 시각화 방법에 따라 필요한 데이터를 선별적으로 적용할 수 있다.

5. 맺음말

인공지능, 사물인터넷에 대한 폭발적인 관심과 함께 데이터 부족에 대한 우려도 증가하고 있다. 빅데이터 현상이 보편화되었지만 특정한 문제 해결을 위해 양질의 데이터를 확보하는 것은 여전히 수월하지 않다. 공공데이터는 사회 전반에 걸쳐 정부 및 공공기관이 수집·생성했기 때문에 향후 활용성이 매우 높다. 이런 측면에서 데이터의 소재정보를 파악하고 접근하는 것은 데이터 활용성을 제고하는 데 큰 역할을 할 수 있다. 현재 국가데이터맵은 범정부 데이터 플랫폼을 통해 구현되고 있으나, 범국가적 수준의 데이터맵 적용을 위해 데이터 구조, 구성요소에 대한 개념적 정의가 필요하고, 서로 다른 수준의 데이터맵을 연결하기 위한 프레임워크 연구가 필요하다.



[그림 5] 국가데이터맵의 구성요소

[참고문헌]

[1] 행정안전부(2013), 공공데이터의 제공 및 이용 활성화에 관한 법률.
[2] 공공데이터전략위원회(2018), 공공데이터 현황 전수조사 추진 계획(공공데이터전략위원회 2차 회의).
[3] 한국정보화진흥원(2018), 범정부 데이터 플랫폼 구축 사업(1단계).
[4] 행정안전부(2018), 데이터기반행정 활성화에 관한 법률안.
[5] 4차산업혁명위원회(2018), 데이터 산업 활성화 전략: I-Korea 4.0 데이터 분야 계획, I-DATA.

[주요 용어 풀이]

- 메타데이터: 데이터의 체계적인 관리와 편리한 검색·활용을 위하여 데이터의 구조, 속성, 특성, 이력 및 용어 등이 표현된 자료(공공기관의 데이터베이스 표준화 지침 참고)
- 국가데이터맵: 데이터 소재지 및 데이터 간 연관관계를 그래프 형태로 표현한 서비스
- 데이터: 전자적 방식으로 처리되어 부호, 문자, 음성, 음향, 영상 등으로 표현된 모든 종류의 자료
- 데이터기반행정: 공공기관이 업무를 수행하는 과정에서 공공 및 민간의 대용량 데이터, 정형·비정형 데이터를 수집·가공·분석·시각화하고 이를 정책수립·집행·평가 등 행정업무에 반영하는 것
- 공공기관: 「국가기관, 지방자치단체, 공공기관의 운영에 관한 법률」제4조에 따른 「공공기관, 지방공기업법」에 따른 지방공사 및 지방공단, 그 밖에 대통령령으로 정하는 법인·기관 및 단체