

국토 교통 분야 데이터 공개 현황 분석 및 품질 향상 방안*

Data Opening Status Analysis and Quality Management Strategies in Land, Infrastructure and Transport Domain

윤성호 (Sung-Ho Yoon) (제1저자) | 전북대학교 기록관리학과 석사과정 | tjdgh9410@naver.com

나정호 (Jeong Ho Na) (공동저자) | 전북대학교 기록관리학과 석사과정 | jhna2012@naver.com

오호정 (Hyo-Jung Oh) (교신저자) | 전북대학교 문헌정보학과 부교수, 문화융복합이카이빙연구소 공동연구원 | ohj@jbnu.ac.kr

목 차

1. 서론
2. 국토 교통 분야 데이터 공개 현황 분석
3. 공개데이터 품질 관리
4. 공개데이터 활용방안
5. 결론

초 록

2013년 「공공데이터의 제공 및 이용 활성화에 관한 법률」이 제정되며 각급 공공기관에서 생산·수집하는 다양한 데이터가 공공데이터 포털을 중심으로 공개되었다. 적극적인 데이터 개방 정책의 결과로 현재 39,730여 건의 데이터가 공공데이터 포털을 통해 제공되고 있다. 하지만 데이터 개방에 초점을 맞춘 정책 추진으로 인해 데이터 품질을 향상시키기 위한 노력은 지지부진하다. 편향된 정책 추진으로 인해 2014년 상왕십리역 충돌사고 등 데이터 오류로 인한 사고가 발생하기도 하였다. 이에 따라 본 연구는 국민의 신체·재산에 직접적 위협을 가할 수 있는 국토 교통 분야의 데이터 공개 현황 분석을 진행하였다. 이를 통해 데이터 품질을 저해하는 3가지 요인을 도출하였으며, 각 문제점에 적합한 개선방안을 제안하였다. 본 연구가 추후 국토 교통 분야의 공개데이터 품질을 향상시키는 정책 마련 시 기초자료로서 활용되기를 기대한다.

* 키워드 : 공공데이터, 데이터 품질, 국토 교통 데이터, 공공데이터 포털

ABSTRACT

In 2013, the "Act on Promotion of The Provision and Use of Public Data." was enacted, and various data produced and collected by public institutions at various levels were disclosed mainly through public data portals. As a result of its aggressive data opening policy, more than 39,730 data are currently being provided through public data portals. However, efforts to improve data quality have been slow due to policy initiatives focused on data opening. Due to the biased policy implementation, accidents occurred due to data errors, such as the 2014 collision at Sangwangsimmni Station. Accordingly, this study conducted an analysis of the status of data disclosure in the national transportation sector, which could pose a direct threat to the people's physical and property. Through this, three factors were derived that impeded data quality, and improvement measures were proposed suitable for each problem. It is hoped that this research will be used as a basic data when preparing policies to improve the quality of open data in the national transportation sector.

* Keywords : Open Data, Data Quality, Land·Infrastructure·Transport Data, Data Portal

* 본 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2019S1A5B8099507).

• 논문접수일 : 2020년 8월 15일 • 최초심사일 : 2020년 8월 17일 • 게재확정일 : 2020년 9월 10일

1. 서론

지난 2014년 5월, 2호선 상왕십리역에서 발생한 추돌사고로 인해 388명의 부상자와 약 28억 원의 재산피해가 발생하였다. 사고 조사를 통해 사고 원인이 3일 전 발생한 신호기 데이터 오류인 것으로 확인되었다.¹⁾

이러한 사고는 데이터 품질 저하로 인해 발생한 사고로 신뢰할 수 있는 데이터 수집을 통해 체계적으로 데이터 품질을 관리함으로써 사전에 예방할 수 있는 사고이다. 이미 버스정보시스템(Bus Information System), 지능형 교통 체계(Intelligent Transportation Systems) 등 국토 교통 분야의 공공 데이터를 활용한 기술은 우리 생활에 많은 편의를 가져왔으나 품질을 보장할 수 없는 데이터는 언제든지 우리에게 물질적, 신체적 위해를 가할 수 있다. 따라서 제공되는 데이터의 오류를 확인하여 데이터 품질을 적정 수준으로 유지하고 관리하는 것이 중요하다.

한편, 데이터의 품질은 공개데이터 이용률에 영향을 미치는 가장 중요한 요소이다. 현재 OECD 국가 중 대한민국은 데이터 개방성은 1위인 반면, 이용률은 5%에 불과하다(김영현, 2016). 이는 데이터를 개방하는 것에만 초점을 맞춘 정책에 기인한 것이다. 이로 인해 높은 수준인 개방성에 비해 데이터 품질은 매우 낮은 수준으로 파악되었다. 이러한 문제를 해소하려면 현재 공개된 데이터의 현황을 파악함과 동시에 실제 공개데이터에서 파악된 데이터 품질의 문제를 확인하는 것이 선제적으로 필요하다. 또한 도출된 데이터 품질 문제를 분석하여 이를 해소할 수 있는 개선방안을 마련하는 것이 필요하다.

따라서 본 연구는 국토 교통 분야의 데이터 포털 및 플랫폼을 대상으로 데이터 공개 현황을 분석하고자 한다. 이를 통해 도출된 데이터 품질과 관련한 문제점을 제기하고 이에 적합한 개선방안을 제안하고자 한다.

2. 국토 교통 분야 데이터 공개 현황 분석

문재인 정부의 데이터 댐 구축 정책 추진으로 인해 정부, 공공기관, 지자체 등 공공영역에서 생산된 기록을 포함해 원시 데이터, 가공정보 등 다양한 유형의 데이터 개방이 권장되는 자금의 상황이다. 이러한 시대적 흐름에 따라 각급 공공기관에서 생산되는 데이터는 데이터 ‘포털’ 혹은 ‘플랫폼’ 등을 통해 일반인에게 제공되고 있다. 그 중 국토 분야 공공데이터 포털의 대표적인 예로 국토교통부의 공간정보오픈플랫폼²⁾, 국가공간정보포털³⁾, 국토지리정보원의 국토정보플랫폼⁴⁾ 등이 있다. 교통 분야 공

1) 이세현 (2020). 튼튼한 데이터 댐을 희망하며. 검색일자: 2020.09.11. <http://news.heraldcorp.com/view.php?ud=20200910000574>

2) 국토교통부 (2012). 공간정보오픈플랫폼. 검색일자: 2020.09.11. https://www.vworld.kr/v4po_main.do

3) 국토교통부 (2003). 국가공간정보포털. 검색일자: 2020.09.11. <http://www.nsd.go.kr/lxportal/?menuno=2679>

4) 국토지리정보원 (2016). 국토정보플랫폼, 검색일자: 2020.09.11. <http://map.ngii.go.kr/mn/mainPage.do>

<표 1> 국토·교통 분야 데이터 포털 현황

구분	데이터 포털	상세내용	제공 데이터 규모	
국토 분야 전문 데이터 포털	국토교통부 공간정보오픈플랫폼	· 건물, 도로와 같은 물리적인 공간 구성요소와 행정경계, 지적 등 논리적인 공간 구성요소로 구성된 공간정보를 2차원·3차원 지도를 활용해 국민에게 제공	Data 11건	OpenAPI 8건
	국토교통부 국가공간정보포털	· 지형, 지질, 토지이용 자연환경에 해당하는 국토공간정보와 도로, 토지, 가옥, 상·하수도, 가스, 전기공급시설 등으로 대표되는 도시공간정보 제공 · 이러한 공간정보를 생산, 관리, 유통, 활용하는 서비스를 제공	Data 754건	OpenAPI 43건
	국토지리정보원 국토정보플랫폼	· 수치지도, 항공사진, 기준점 등 국토지리정보원에서 생산하는 다양한 공간정보의 제공 · 민간, 공공 등 다양한 분야에서 공간정보의 융·복합, 개방, 활용을 지원 및 담당	Data 13건	OpenAPI 7건
교통 분야 전문 데이터 포털	국토교통부 교통정보공개서비스	· 교통소통정보, 공사정보, CCTV정보, 돌발상황정보 등 고속도로, 국도를 이용에 유용한 정보를 제공	Data 10건	OpenAPI 7건
	한국도로공사 고속도로 공공데이터 포털	· 한국도로공사에서 보유한 고속도로 공공데이터 정보를 제공 · 교통, 건설, 유지관리, 일반행정, 통행료, 휴게소 등 고속도로 관련 다양한 데이터를 수집 및 제공 · 수집한 공공데이터는 웹 서비스 형태로 공개하여 사용자가 원하는 공공데이터 콘텐츠 개발을 지원하는 역할을 수행	Data 223건	OpenAPI 89건
	도로교통공단 교통사고분석시스템	· 교통사고 자료를 수집·통합·분석하여 교통안전 정책 수립 등에 활용할 수 있도록 교통사고 정보를 제공 · 교통사고정보 개방 및 공유 확대를 사용자 맞춤형 교통사고정보 제공 체계 지원	Data 4건	OpenAPI 14건

공데이터 포털로는 국토교통부의 교통공개정보서비스⁵⁾, 도로교통공단의 교통사고분석시스템⁶⁾, 한국도로공사의 고속도로 공공데이터 포털⁷⁾ 등이 있다(<표 1> 참고).

국토 분야의 전문 데이터 포털은 주로 국토, 공간, 지리 등의 공간정보를 2차원 또는 3차원 지도를 활용하여 국민에게 제공하는 서비스를 수행한다. 국토교통부의 공간정보오픈플랫폼과 국토지리정보원의 국토정보플랫폼이 유사한 내용의 데이터를 제공한다. 공간정보오픈플랫폼의 경우, 영상지도로 제공되는 데이터의 범위가 한반도 뿐 아니라 전 세계 육지와 바다를 대상으로 한다. 전 세계 3차원 영상지도를 Data, Open API 형태로 제공한다. 또한 문화·체육·관광·사회복지·주요상권 등 다양한 산업 분야와 결합되어 활용될 수 있는 가치 있는 정보를 제공하고 있다. 국토지리정보원의 국토정보플랫폼은 앞서 언급한 공간정보오픈플랫폼과 동일하게 한반도 일대의 지도 정보를 제공한다는 공통점이 있으나, 국토변화 정보를 제공한다는 특징이 있다. 다시 말해 본 플랫폼은 지형·지물을 포괄한 지도 정보를 시간의 흐름에 따라 변동된 내용을 이용자들이 이용 가능하도록 관리하여 제공하고 있다. 이를 통해 국민에게 효율적으로 국토변화정보를 제공하고 다양한 활용환경을 제공 가능함과 동시에 최근 주요 정보 접근 수단으로 자리 잡은 모바일 기반의 활용서비스를 구축하여 국토변화정보를 제공 가능하도록 관련 인프라를 구축하고 있다.

앞서 언급한 국토 분야의 2가지 데이터 포털과는 다르게 국토교통부의 국가공간정보포털은 지도 정보를 포함하여 지형, 지질, 토지이용 자연환경에 해당하는 국토공간정보와 도로, 토지, 가옥, 상·하수도, 가스, 전기 공급시설 등으로 대표되는 도시공간정보를 제공한다. 이렇듯 광범위한 데이터 주제에 맞춰 전체 754건의 데이터와 43건의 OpenAPI를 제공한다. 국가공간정보포털에서 제공하는 여러 데이터 중 가장 특징적인 데이터는 공간빅데이터이다. 공간빅데이터는 국토교통부를 포함한 행정안전부, 환경부, 국세청, 기상청 등 관계 부처에서 수집한 데이터를 일컫는다. 부동산·시설물 정보·주거 유형·토양정보·물류정보 등 폭넓은 데이터를 수집하여 공간분석, 텍스트마이닝, 공간통계분석을 수행해 민관영역에서 활용할 수 있도록 정보를 제공하고자 한다. 공간빅데이터는 부처 간 연계체계를 통해 수집되어 공공부문에 도입 시 중장기적으로 약 2.1조원 ~ 4.2조원의 부가가치 유발 효과와 관련 일자리를 약 52만 개 창출할 것으로 기대되고 있다(국토교통부 2014).

교통 분야 전문 데이터 포털은 국토교통부, 한국도로공사, 도로교통공단 등 다양한 기관에서 각각 운영하고 있는데 주로 교통, CCTV, 도로, 통행량, 관련 시설, 교통사고 등의 데이터를 제공한다. 먼저 한국도로공사의 고속도로 공공데이터 포털은 앞서 언급한 주제별 데이터를 모두 제공하는 포털이다. 교통, 건설, 유지관리, 휴게소, 일반행정, 융합 데이터, 통행료 등 여러 분야의 데이터를 수집·관리·제공하며, 총 223건의 데이터와 89건의 OpenAPI를 제공한다. 전국 고속도로와 국도에 설치된 VDS⁸⁾

5) 국토교통부 (2012). 교통정보공개서비스, 검색일자: 2020.09.11. <http://openapi.its.go.kr/portal/main.do>

6) 도로교통공단 (2017). 교통사고분석시스템, 검색일자: 2020.09.11. <http://taas.koroad.or.kr/>

7) 한국도로공사 (2015). 고속도로 공공데이터 포털, 검색일자: 2020.09.11. <http://data.ex.co.kr/>

8) 한국도로공사가 운영하는 고속도로 구간은 콘존을 기준으로 분류된다. 이를 통해 전체 988개 구간으로 분류되며, 구간마다 설치된 차량 검지기를 VDS로 일컫는다(김상구 외, 2016).

를 이용하여 실시간 교통량, 통행 속도, 지정체 정보, 기상정보 등을 수집하고 원자료를 가공하고 분석한 결과값을 제공한다. VDS는 주요 구간의 경우 1~2Km마다 설치되어 있지만 아예 설치되지 않은 구간이 다수 존재한다. 따라서 해당 구간은 교통량, 통행 속도 등의 데이터가 수집되지 않을 뿐만 아니라 교통사고와 같은 돌발상황을 빠르게 인지하고 대처하기가 어렵다. 또한 이러한 데이터가 공개되는 이유 중 하나는 해당 데이터의 이용활성화이다. 하지만 VDS가 설치되지 않아 관련 데이터가 누락되기 때문에 결측값이 발생하게 된다. 이러한 결측값이 다수 존재할 경우, 해당 데이터셋을 활용해 정확한 분석을 수행하기 어려운 단점이 존재한다. 이와 관련한 사항은 다음 장에서 자세히 언급하도록 하겠다.

국토교통부에서 운영하는 교통정보공개서비스도 앞서 소개한 고속도로 공공데이터 포털과 유사한 분야의 데이터를 제공한다. 교통소통정보, 공사/사고정보, VMS⁹⁾ 표출정보 등을 제공한다. 교통소통정보의 경우, 자료 제공기간을 2019년 3월 이후로 명시하고 있다. 하지만 데이터 수집 일자를 2019년 3월 4일로 설정하여 검색한 결과, 아래의 <그림 1>과 같이 파일의 크기가 0mb인 것을 확인하였다. 데이터를 기술(記述)한 것과 실제 제공 데이터 간 차이가 존재하였다. 이는 공사/사고정보도 마찬가지였다. 이와 같이 이용자가 해당 데이터를 이해하는 도움을 주는 기술정보, 메타데이터가 부정확하고 미흡하게 기술됨으로서 데이터를 정확하게 이해하고 활용하기가 어렵다.

⊕ 교통소통정보

일자 집계주기

자료 제공기간	2019년 3월 이후	자료 제공주기	1일
기준일자	2019-03-04	파일크기	0
파일명	20190304_5Min.zip (0)		
자료 구조	집계일자, 링크ID, 속도		

<그림 1> 국토교통부 교통정보공개서비스 화면

마지막으로 도로교통공단에서 운영하는 교통사고분석시스템(Traffic Accident Analysis System, 이하 TAAS)을 살펴보고자 한다. TAAS는 앞서 살펴본 데이터 포털과는 다르게 교통사고와 관련한 통계 및 GIS 데이터를 제공한다. 교통사고와 관련한 통계자료를 사고유형별, 도로환경별, 차량·사상자 관련, 어린이·노인 교통사고 등 다양한 기준에 따라 분류하여 이용자에게 제공한다. 또한 OECD 국가의 교통사고 정보를 제공한다. 국가별 교통사고 사망자 수, 연령대별 사망자 수, 최근 5년간 교통사고 건 수 및 사망자 수 정보 등을 집계하여 제공한다. 교통사고 정보는 GIS 분석시스템을 통해 특정 조건

9) VMS(Variable Message Sign)는 가변전광표지로 전방도로상황, 기상상황 등의 교통정보를 제공하여 교통흐름관리 및 교통안전관리를 하는 ITS 중 하나이다(이상혁, 조혜진, 2015).

에 해당하는 사고 정보 지도상에 시각화하여 표현할 수 있다. 이를 통해 각종 검색조건에 따른 통계지도를 검색함과 동시에 이용자가 지정한 조건에 만족하는 시각화 정보를 활용해 교통사고 지역별 분석, 교통사고 다발지역 도출 등의 데이터 분석에 활용할 수 있다. 이와 더불어 교통사고정보와 실시간 기상정보, 집회·사고·집회 등 도로 내 돌발정보를 융합해 빅데이터 분석하여 도로 구간별 위험정보를 4단계로 제공한다. 이는 교통사고가 인적요인 외에도 복합적인 원인으로 발생하는 점에서 기인하여 사전에 도로 위험성을 평가·제공하여 운전자가 위험도를 인지하여 안전운행을 가능케 하도록 지원한다. 도로와 차종에 따라 구분하여 위험도를 안전, 주의, 위험, 심각 등 4단계로 평가한다. 하지만 위험도로 예보시스템은 웹에서만 지원되고 모바일에서는 지원되지 않는다. 운행 중 교통정보 접근 시 모바일을 이용해 관련 정보를 접근하는 경우가 대부분이기 때문에 관련 어플리케이션 개발을 통해 위험보도 예보시스템의 이용활성화를 도모하는 것이 필요하다고 판단된다.

본 절에서는 국토 교통 분야의 다양한 데이터 포털과 플랫폼, 시스템을 살펴보았다. 분석 결과, 다양한 분야의 데이터와 OpenAPI를 통해 이용자들이 데이터를 활용할 수 있도록 지원하고 있었다. 그러나 구체적인 데이터 수집 결과를 검수한 결과, 데이터 품질과 관련한 몇 가지 문제점이 도출되었다. 첫째, 데이터 내 결측값이 존재한다. 둘째, 데이터를 기술하는 정보가 미흡하고 부정확하다. 셋째, 데이터 제공 시 파일 포맷 확장자가 상이하다. 이와 같은 데이터 품질의 문제점을 다음 장에서 언급하고자 한다.

3. 공개데이터 품질 관리

3.1 공개데이터 품질의 문제점

앞서 2장에서 살펴본 6곳의 국토 교통 분야 데이터 포털에서 제공되는 이질·이종 데이터는 4차 산업혁명의 도래로 도출된 여러 IT기술과 융합되어 다양하게 활용될 수 있다. 하지만 데이터의 개방에만 초점을 맞춘 현 정책의 한계로 인해 데이터의 품질은 높은 수준인 개방성에 한참 미치지 못하는 상황이다(윤상오, 현지우 2019). 본 연구에서는 데이터 공개 현황분석을 통해 공공데이터 품질의 문제점으로 다음 3가지를 도출하였다.

첫째, 공개된 데이터에 다수의 결측값이 존재한다. 결측값은 데이터셋 내 데이터가 비어있는 값이다. 이러한 결측값은 해당 데이터셋을 활용하기에 매우 부적합하다. 예를 들어 VDS 데이터셋 중 특정 콘존 내 VDS에서 수집되는 데이터가 누락되기 때문에 해당 구간에 대한 분석을 정확히 수행하기 어렵다. 결측값이 발생했을 경우, 선행 VDS 값과 후행 VDS 값의 평균으로 해당 결측값 문제를 해소할 수 있다. 하지만 아래의 <그림 2>와 같이, 연속적으로 결측값이 존재하는 경우가 다수 존재하였다. 이

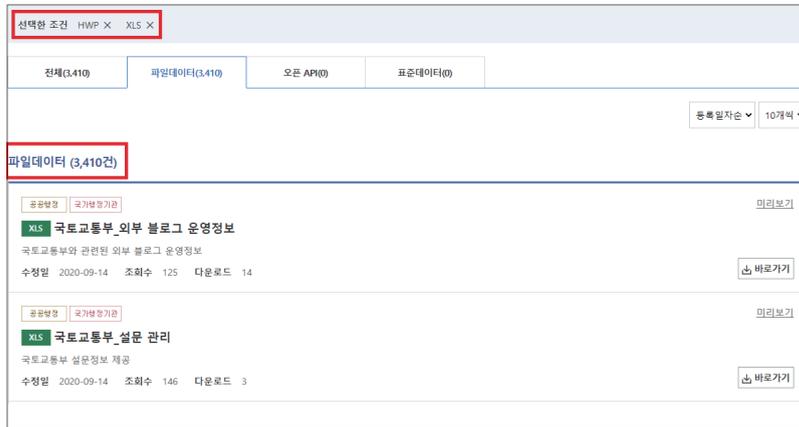
러한 경우는 해당 VDS가 설치된 구간에 대한 정보를 알 수 없기 때문에 도로 구간을 정확히 분석하는 것이 어렵다. 또한 <그림 2>의 중간에 표시된 “교통량 -1”과 같은 데이터를 확인하지 않고 이용자에게 제공하는 문제도 있다. “교통량 -1”의 의미를 파악할 수 있는 설명, 주석 등이 존재하지 않기 때문에 이용자가 해석할 수 없고 결측값으로 처리해야하는 경우가 발생한다.

date	czld	교통량	발표일시	현재일기	기상실황	지상실황	AWSID	AWS명
20190101	0010CZE010	35229						
20190101	0010CZE011	34561						
20190101	0010CZE035	30945						
20190101	0010CZE055	-1						
20190101	0010CZE080	29262						
20190101	0010CZE085	28992						
20190101	0010CZE090	35787						
20190101	0010CZE100	36534						
20190101	0010CZE105	21588						

<그림 2> VDS 데이터 예시

둘째, 데이터 기술(記述) 정보가 미흡하고 부정확하다. 앞서 국토교통부의 교통정보공개서비스 현황을 살펴본 바와 같이, 데이터 설명으로 명시된 데이터 제공기간과 실제 공개된 데이터 간 상이한 경우가 확인되었다. 데이터를 기술하는 목적은 이용자가 해당 데이터의 내부를 확인하지 않더라도 기술 정보를 토대로 해당 데이터를 이해하고 데이터의 내용을 추측할 수 있도록 지원한다. 하지만 앞서 살펴본 <그림 1>과 같이 기술이 정확하지 않은 경우가 존재한다면 이용자가 해당 데이터를 신뢰할 수 없게 된다. 이용자가 신뢰하지 못하는 데이터는 활용되지 못할 것이며, 이는 공공데이터의 이용률 저하로 이어지게 될 것이다.

셋째, 데이터 포맷의 불일치성이다. 데이터 개방 활성화에 따라 현재 공공데이터 포털에 업로드 된 데이터는 총 33,730건에 이른다. 하지만 이 중 표준데이터는 120건에 불과할 정도로 공공데이터는 다양한 파일 포맷으로 개방되어 왔다. 이렇듯 데이터 포맷이 제각각이기 때문에 이용자들은 데이터를 일괄적으로 분석하거나 활용하기에 어려움이 발생한다. 또한 CSV, HTML, XML 등 오픈 파일 포맷이 아닌 특정 제조사에 종속된 바이너리 파일 포맷(e.g. HWP, XLS 등)으로 데이터가 제공되는 문제도 발생한다. 현재 공공데이터 포털에서 제공하는 공공데이터는 15개 파일 포맷으로 제공된다. 그 중 바이너리 파일 포맷인 HWP, XLS로 제공되는 데이터는 총 3,410건으로 전체 공공데이터 중 약 10%에 이른다(<그림 3> 참고). 이러한 바이너리 파일 포맷으로 제공되는 데이터를 이용하려면 상용 SW를 설치가 선제적으로 필요하기 때문에 이용자들이 데이터에 접근하기 어려울 수 있다. 따라서 표준 데이터 규격에 적합한 파일 포맷으로 일치시켜 데이터를 개방 및 제공하는 것이 필요하다.



<그림 3> 공공데이터 포털 상용 파일 포맷 개방 현황

3.2 공개데이터 품질 개선방안



<그림 4> 공개데이터 품질 문제점 및 개선방안

본 연구에서는 선행 절에서 언급한 데이터 품질의 문제점들은 해소하기 위한 개선방향으로 <그림 4>와 같이 2가지 방안을 제안하고자 한다. 우선 공개데이터를 관리하는 인력의 전문성을 확보하는 것이 중요하다. 한국정보화진흥원의 관계자에 따르면 정보화담당관이 부재한 기관이 많이 있다고 언급하기도 하였다(임우선, 정세진, 2016). 정보화담당관이 아닌 인력이 데이터 개방 업무를 담당한다면 데이터 품질을 확보하는 것은 불가능할 수 밖에 없다. 전문 인력이 부재한 상황이라면 데이터 관련 업무를 담당할 직원에게 전문적인 교육프로그램을 지원하여 인력의 전문성을 확보하는 것이 필요하다. 또한 기관에서 생산하는 이종·이질 데이터를 효과적으로 수집, 가공할 수 있는 데이터 처리 도구의 도입도 필요하다. 현재 개방된 데이터에 대해 민간에서 제기되는 가장 큰 문제는 결측값과 같은 데이터 오류이다. 이를 해결하기 위해서는 개방 전에 이러한 오류를 식별한 뒤 보정하거나 제거하여 양질의 데이터만을 선별하는 전처리 과정이 필수적이다. 따라서 효과적으로 데이터 처리를 지원하는 도구

사용, 관리 인력의 전문성이 확보되어야 데이터에 대한 이해를 바탕으로 앞서 언급한 데이터 결측값과 데이터 기술의 미흡함에서 기인하는 문제를 해소할 수 있을 것이다.

다음으로 강화된 공공데이터 표준화 적용이 필요하다. 현재 「공공기관의 데이터베이스 표준화 지침」을 통해 공공데이터 표준화, 메타데이터 관리 등에 대한 세부 지침을 안내하고 있다(공공기관의 데이터베이스 표준화 지침(2019. 03. 26.) 제7조, 제10조). 이와 더불어 정부는 지난 7월 발표된 데이터 댐 7대 사업을 추진해 데이터 구축 공통 가이드라인을 마련하여 데이터 표준화를 추진하고 있다. 이렇듯 정부에서는 기존의 표준화 지침을 지속적으로 갱신하여 데이터 표준화를 도모하였지만 강제력이 없는 지침이기 때문에 여전히 데이터의 표준화는 더딘 상황이다. 따라서 디지털 뉴딜을 통해 개발되는 새로운 가이드라인에는 데이터 포맷 선정에 강제력을 부여해 오픈 포맷으로 데이터를 개방 및 제공할 수 있는 조항을 포함하도록 하는 지침이 필요하다.

4. 공개데이터 활용방안

앞서 살펴본 국토 교통 분야의 공개데이터는 다양하게 활용될 수 있다. 그 중 GIS 기반의 실시간 정보 서비스를 제공할 때 더욱 유용하게 이용될 것이다. 예를 들어 블랙아이스와 같은 국토, 지리, 교통, 도로, 기상 등 다양한 요인으로 발생하는 사건에 대한 예측 시스템을 구축한다면 더욱 활용도가 높을 것이다.

블랙아이스는 도로 위 얇은 결빙이 이물질과 결합하여 도로와 같은 검은색으로 나타나는 현상이다. 검은색으로 나타나는 특징으로 인해 운전자들이 운행 중 블랙아이스를 발견하여 대처하는 것이 상당히 어렵다. 특히 블랙아이스는 도로 위 포장재, 표층재, 교량·터널과 같은 도로 시설 정보, 교통량 등 이질적이고 다형적인 변수들에게 영향을 받는다. 그렇기 때문에 이질적이고 다형적인 다량의 국토 교통 분야 공개데이터와 기계학습을 활용하여 블랙아이스가 발생할 것으로 예상되는 시스템을 구축한다면 많은 이용자들이 활용할만한 유용한 시스템으로 각광받을 수 있을 것이다. 이와 더불어 이용자가 해당 정보를 이해하고 접근하기 용이하도록 위험구간 정보를 지도로 시각화하여 제공하는 방안을 마련하는 것이 중요하다고 생각된다. 운행 중 효과적으로 위험구간 정보를 확인하려면 모바일을 통해 해당 구간 접근 시 경보, 알림 등을 제공하는 것이 중요하다. 또한 위험도를 지수화로 표현하여 위험등급을 설정하여 이용자의 안전운행을 도모할 수 있다. 아래의 <그림 5>는 블랙아이스 발생구간 예측 시스템의 예시이다. <표 2>에 명시된 다양한 데이터를 가공 및 기계학습 시킨 후 네비게이션 API를 활용하여 <그림 5>와 같은 어플리케이션을 구축할 수 있다. 이용자가 이동하는 경로 내 블랙아이스 위험구간이 존재한다면 위험구간 알림과 더불어 게이지 바를 이용해 위험도를 시각화하여 운전자에게 정보를 제공하는 어플리케이션을 개발할 수 있다. 이러한 도로 위험구간을 알려주는 어플리케이션

이션을 개발한다면 관련 교통사고의 감소를 기대할 수 있을 것이다.



<그림 5> 공개데이터 활용 예시

데이터 분류	데이터명
기상	기온데이터
	강수데이터
	이슬점 온도 데이터
도로	도로 표층재데이터
	도로 포장재데이터
	교량데이터
	터널데이터
VDS	실시간 교통량 데이터

<표 2> 블랙아이스 발생구간 예측 시스템 활용 데이터(안)

5. 결론

IT 기술의 발전과 4차 산업혁명의 도래로 정부와 각급 공공기관이 보유한 공공데이터가 내포한 잠재력을 실제 가치로 실현하고자 데이터 개방 정책을 추진 중이다. 공공데이터 포털을 운영하여 적극적으로 공공데이터를 개방하는 노력을 통해 OECD가 평가하는 공공데이터 개방지수에서 2015, 2017년 세계 1위를 차지하였다. 그러나 공공데이터를 개방하는 주요 선진국에 비해 개방되는 데이터의 양이나 품질은 낮은 것으로 평가되고 있다(윤상오, 현지우 2019).

따라서 본 연구는 국토 교통 분야의 6개 데이터 포털에서 제공하는 데이터 품질 저하 원인이 되는 문제점을 도출하고자 하였다. 또한 도출된 문제점에 따른 개선방안을 제안하였다. 선제적으로 데이터 포털에서 제공되는 데이터 현황을 분석하였다. 이를 기반으로 공개데이터의 품질을 저해하는 문제점 3가지를 도출하였다. 다량의 결측값 존재, 미흡한 데이터 기술, 데이터 포맷의 불일치 등의 문제는 데이터의 신뢰성과 활용성을 저하시키는 요인이 된다. 결측값이 다수 존재하는 낮은 품질의 데이터셋은 데이터 분석 또는 시스템 구축 시 활용될 가능성이 현저히 떨어진다. 또한 데이터를 기술하는 정보와 실제 데이터 간 상이함과 공개데이터의 파일 포맷이 불일치한 경우가 다수 도출되었다. 이러한 문제점들을 개선하기 위해선 데이터 관리 인력의 전문화와 공공데이터 표준화가 도입되어야 한다. 이를 통해 데이터 품질을 재고함과 동시에 이용활성화를 촉진시킬 것으로 기대된다.

데이터는 우리 생활에 많은 영향을 주게 되었다. 일반 시민들이 필요로 하는 양질의 데이터를 수집·가공하고 지능형 교통 체계와 같이 효과적으로 활용된다면 데이터는 더욱 편리하고 운택한 생활을 영유하는데 도움을 줄 것이다. 이러한 선순환의 구조가 유지되려면 데이터의 품질이 높은 수준으로 유지하는 노력이 필요하고 민관에서 다양하게 활용할 수 있도록 지원하는 것이 중요하다.

참고문헌

공공기관 데이터베이스 표준화 지침

공공데이터의 제공 및 이용 활성화에 관한 법률

국토교통부 (2012). 공간정보오픈플랫폼. 검색일자: 2020. 09. 11.

https://www.vworld.kr/v4po_main.do

국토교통부 (2012). 교통정보공개서비스. 검색일자: 2020. 09. 11.

<http://openapi.its.go.kr/portal/main.do>

국토교통부 (2014). 국가공간정보포털 공간빅데이터. 검색일자: 2020. 09. 11.

<http://www.nsdi.go.kr/lxportal/?menuno=2684>

국토지리정보원 (2016). 국토정보플랫폼. 검색일자: 2020. 09. 11.

<http://map.ngii.go.kr/mn/mainPage.do>

기획재정부 (2020). 한국판 뉴딜. 검색일자: 2020. 09. 11. <http://www.moef.go.kr/mp/nd/newDeal.do>

김상구, 윤일수, 박재범, 박인기, 천승훈, 김경현, 안현경 (2016). VDS 자료 기반 고속도로 교통혼잡비용 산정 방법론 연구. 한국도로학회 논문집, 16(1), 99-107.

김영현 (2016). 정부3.0 기반 수요자 중심의 빅데이터 서비스 활성화 방안: 부산도시서비스분석 정보시스템·부산공공데이터 포털 중심으로. 지역정보화, 101, 16-21.

김태진 (2020). '데이터 댐', 어떤 사업들로 구성됐나. 검색일자: 2020. 09. 11. <https://zdnet.co.kr/view/?no=20200902101741>

도로교통공단 (2017). 교통사고분석시스템. 검색일자: 2020. 09. 11. <http://taas.koroad.or.kr/>

윤상오, 현지우 (2019). 공공데이터 개방정책의 실태분석 및 개선방안에 관한 연구: 공공데이터 포털의 국가중점데이터 개방 사례를 중심으로. 한국공공관리학보, 33(1), 219-247.

이상혁, 조혜진 (2015). VMS 교통정보의 교통안전효과에 관한 연구. 한국ITS학회 논문지, 14(1), 20-11.

이제현 (2020). 튼튼한 데이터 댐을 희망하며. 검색일자: 2020. 09. 11. <http://news.heraldcorp.com/view.php?ud=20200910000574>

임우선, 정세진 (2016). 공공데이터 량 적고 쓸모없는 파일 많아... 기업들 외면. 검색일자: 2020. 09. 12. <https://www.donga.com/news/article/all/20160517/78152584/1>

한국도로공사 (2015). 고속도로 공공데이터 포털. 검색일자: 2020. 09. 11. <http://data.ex.co.kr/>

국한문 참고문헌의 영문 표기

(English translation / Romanization of reference originally written in Korean)

Act on Promotion of The Provision and Use of Public Data.

Guidelines for Standardizing Public Institution Databases Article 7 and Article 10

Kim, Sang-Gu, Yun, Il-soo, Park, Jae-Beom, Park, In-Ki, Cheon, Seung-Hoon, Kim, Kyung-Hyun

-
- & Ahn, Hyun-Kyung (2016). Estimation of the Expressway Traffic Congestion Cost Using Vehicle Detection System Data. *International Journal of Highway Engineering*, 18(1), 99-107.
- Kim, Tae-jin (2020). 'Data Dam', What Kind of Businesses Are They Made Up. Retrieved September 11, 2020, from <https://zdnet.co.kr/view/?no=20200902101741>
- Kim, Young-hyun (2016). Government 3.0 based consumer oriented big data service activation plan: Busan city service analysis information system and Busan public data portal. *Local Informatization Magazine*, 101, 16-21.
- Korea Expressway Corporation (2015). Express Public Data Portal. Retrieved September 11, 2020, from <http://data.ex.co.kr/>
- Lee, Je-Hyun (2020). Hope For a Strong Data Dam. Retrieved September 11, 2020, from <http://news.heraldcorp.com/view.php?ud=20200910000574>
- Lee, Sang-Hyuk & Cho, Hye-Jin (2015). A Study on Safety Impacts for VMS Traffic Information. *The Journal of The Korea Institute of Intelligent Transport Systems*, 14(1), 22-30.
- Lim, Woo-Sun & Jung, Se-jin (2016). Open Data, Small Amount. Useless Files. Retrieved September 11, 2020, from <https://www.donga.com/news/article/all/20160517/78152584/1>
- Ministry of Economy and Finance (2020). New Deal. Retrieved September 11, 2020, from <http://www.moef.go.kr/mp/nd/newDeal.do>
- Ministry of Land, Infrastructure and Transport (2014). National Spatial Data Infrastructure Portal Spatial Bigdata. Retrieved September 11, 2020, from <http://www.nsd.go.kr/lxportal/?menuno=2684>
- Ministry of Land, Infrastructure and Transport (2012). Spatial Information Open Platform. Retrieved September 11, 2020, from https://www.vworld.kr/v4po_main.do
- Ministry of Land, Infrastructure and Transport (2012). Traffic Information Open Service. Retrieved September 11, 2020, from <http://openapi.its.go.kr/portal/main.do>
- National Geographic Information Institute (2016). National Geographic Information Platform. Retrieved September 11, 2020, from <http://map.ngii.go.kr/mn/mainPage.do>
- The Road Traffic Authority (2017). Traffic Accident Analysis System. Retrieved September 11, 2020, from <http://taas.koroad.or.kr/>
- Yun, Sang-oh & Hyun, Ji-woo (2019). An Analysis of Open Data Policy in Korea: Focused on National Core Data in Open Data Portal. *Korean Public Management Review*, 33(1), 219-247.