국가 데이터 생태계 활성화를 위한 통합 인프라 기능 요구사항

이영주[1], 김정엽[2], 유시형[3], 신신애[4]
[1]한국지능정보사회진흥원 팀장
[2,3]한국지능정보사회진흥원 수석연구원
[4]한국지능정보사회진흥원 본부장
lyj@nia.or.kr, tisphone@nia.or.kr, ysh@nia.or.kr, sashin@nia.or.kr(교신저자)

Functional requirement for integrated data platorm infrastructure in boosting nationwide data ecosystem

Young-Joo Lee[1], JeongYeop Kim[2], Si-hyeong Yoo[3,] Shinae Shin[4,]
[1,2,3,4]Dept. of AI Data, National Information society Agency
[4]Corresponding Author(sashin@nia.or.kr)

요 약

민간, 공공을 막론하고 디지털 혁신이 필수인 현재 데이터의 중요성은 갈수록 높아지고 있다. 그간 정부의 대규모 지원사업과 민간의 자체 투자로 공급 측면의 데이터는 상당량 축적되어 있으나 막상 수요자가 필요한 데이터를 적시에 구해서 활용할 수 있는 기반은 여전히 미흡한 수준이다. 본 연구에서는 민간-공공을 거쳐 데이터의 유통과 거래를 활성화하고, 수요자와 공급자의 참여를 확대할 수 있는 국가 차원의 통합 데이터 인프라를 제안하고 그 기능적 요구사항과 추진 방향을 제시한다.

1. 서론

인공지능(AI) 일상화 시대에 데이터는 기술혁신과 사회발전을 좌우하고, 나아가 디지털 경제의 원동력이 됨은 두말할 필요가 없다. 주요국은 이미 디지털 전환의 핵심 기반이 데이터 산업이라는 인식하에 국가 데이터 전략을 넘어 데이터 산업 육성과 인프라 조성에 전념을 다하고 있다.

우리나라 정부도 지난 5년여 동안 범국가 차원에서 다양한 방식으로 데이터를 공급하고 활용을 활성화하기 위해 많은 노력을 기울였다. 특히 데이터 댐 사업의 일환으로 분야별 빅데이터 플랫폼 구축을 지원하여 양질의 데이터를 공급할 수 있는 기반을 조성해 왔다. 그 결과 민간, 공공에 걸쳐 수많은 데이터 플랫폼이 우후죽순 생겨났는데, 이는 수요자 입장에서는 오히려 산재된 플랫폼들에서 원하는 데이터를 쉽게 찾기가 어렵게 되었다. 또한 공급자 입장에서도 제공한 데이터에 대한 권리와 통제권을 상실할 우려로 적극적인 데이터 제공에 주저하고 있다.

이러한 문제의식에서 출발하여 데이터의 유통과 거래를 활성화하고, 수요자와 공급자의 참여를 확대하기 위해서 공공, 민간에 흩어져 있는 데이터 플랫폼을 하나로 묶어주는 국가 차원의 통합 인프라의 필요성이 제기되었다. 국가 데이터 통합 인프라는 분야를 넘나들며 데이터가 자유롭게 흐를 수 있도록 관련된 공통기준과 표준을 제시하고, 데이터 제공자, 중개자, 수요자, 이용자를 연결하는 공통 기반을 제공한다.

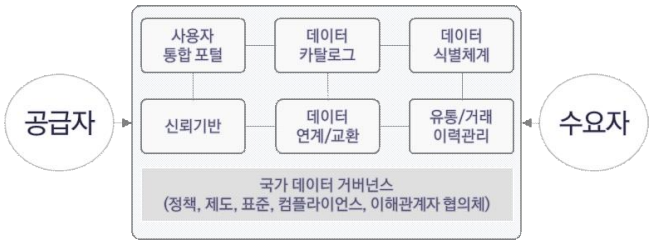
본 연구는 국가데이터 통합 인프라의 구성 개념과 요구기능, 기능별 요구사항을 정의하고 추진방향을 제시하고자자 한다. 2장에서는 국가 데이터 통합 인프라의 최상위 기능 요구사항을 개념화 및 구조화하였고, 3장에는 각 기능 요구사항 별로 국내 현황 및 문제점 분석에 따른 추진 방향을 제안하고자 한다.

2. 국가 데이터 통합 인프라

데이터 플랫폼 간 연계를 위한 공통 인프라는 시스템 측면, 운영 관리 측면, 데이터 측면으로 구분해 볼 수 있다.[10] 본 연구에서는 국내외 문헌 및 해외 국가 정책 사례 조사와, 핵심 이해관계자 인터뷰를 통해 <표 1>과 같이 국가 데이터 통합인프라

의 기능적 요구사항을 도출하였다.

<표 1> 데이터 통합인프라 기능 요구사항



통합인프라는 우선적으로 민간, 공공에 산재된 플랫폼의 데이터 정보를 한곳에서 검색하고 필요시 상호 연계 및 교환, 유통 및 거래에 필요한 기능을 제공하는 메타플랫폼(platform of platform)의 위상을 가져야 한다. 메타플랫폼에서 우선적으로 필요한 기능적 요구사항으로는 사용자 통합 포털과 데이터 카탈로그[11]의 구현이 필요하다. 다음으로는 데이터 생태계 활성화에 저해 요인으로 그간 지적되었던 참여자의 신뢰성 보장, 식별 체계, 표준 기반 연계/교환, 주권기반 유통거래 이력관리 등이 요구된다[1].

3. 기능 요구사항 별 추진 방향

3.1 사용자 통합 포털

2023년 기준 국내 공공, 민간의 데이터플랫폼은 약 190개인 것으로 파악되었다. 공공에서는 과기정통부 지원사업으로 구축된 21개 빅데이터 플랫폼 외에도 행정안전부의 공공데이터포털, 그리고 각 부처와 지자체가 자체 구축한 데이터 플랫폼이 있으며, 민간에서는 쿠콘 '쿠콘닷컴', 롯데카드 '데이터스' 등 데이터 마켓플레이스 형태의 플랫폼 들이 존재한다(NIA 내부자료). 통합 인프라의 사용자 포털에서 각 분야별 개별 포털을 연계하기 위해서는 메타데이터 기반 상향식 연계와 표준 기반 데이터 연동, 운영 인프라 확충 등이 필요하다[7]. 구체적으로는 각 플랫폼에서 유통하고 있는 데이터 상품에 대한 소재정보, 카탈로그 정보, 맞춤형 추천, 데이터 구매 협의 요청, 커뮤니티 기능 등 데이터 탐색부터 활용까지의 전 과정을 지원하는 사용자 중심 기능이 구현되어야 한다.

1) 국가 데이터거버넌스는 통합인프라의 중요한 구성요소이나 본 연구의 범위를 벗어나므로 향후 연구에서 다루고자 한다

3.2 표준 데이터 카탈로그

데이터 포털의 난립으로 인한 파편화 현상을 해결하기 위하여, 개별 메타데이터의 호환을 위한 메타-메타 데이터(Meta-metadata)의 필요성이 생겼고, 이러한 목적으로 데이터 카탈로그(Data Catalog)라는 개념이 EU와 W3C를 중심으로 제안되었다.[12]

데이터 카탈로그란 "특정 데이터 포털이 어떤 종류의 데이터를 가지고 있고, 데이터가 어떤 형태를 띄고 있는지를 표시하는, 말 그대로 카탈로그 역할을 하는 메타데이터이다. 이를 이용하면 데이터 포털의 독립성은 최대한 보장하면서, 데이터 포털이 보유한 데이터를 데이터 카탈로그 메타데이터로 제공하여, 이를 통한 데이터 검색과 유통이 가능한 상호운용 구조를 만들 수 있게 된다. EU 데이터 포털은 EU의 모든 데이터를 하나의 포털에서 유지하는 전략을 택하지 않고, 개별 데이터는 해당 데이터 포털이 유지하고 대신에 데이터에 대한 카탈로그만을 EU 데이터 포털에 유지하는 방법을 채택하였다.[12]

본 연구에서 제안하는 표준 데이터 카탈로그는 EU의 사례처럼 국가 차원의 데이터 카탈로그 구축 목적을 달성하면서도 활용 목적 별 정보제공의 수단으로서 국제표준인 DCAT(Data Catalog Vocabulary)의 하위 구성개념으로서 데이터 카탈로그의 형식도 준용하여 통합 포털에 데이터를 게시하고자 한다[13]. 나아가 기존 연계된 데이터 소재 정보와 메타데이터 등을 적극 활용하여 유통·활용에 필요한 메타데이터를 자동으로 생성하는 기반이 필요하며, 금융, 의료, 공간정보 등 분야별 특성에 맞는 표준 카탈로그의 모델을 도출하여 검증할 필요가 있다.

3.3. 데이터 식별체계

데이터 식별체계는 데이터의 생산부터 수집, 등록, 분석가공, 활용까지의 전 생애주기에 걸쳐 추적성을 확보할 수 있도록 영구식별자(PID, Persistent IDentifer) 부여하는 것으로 시간의 경과로 소재 정보의 변경과 관계없이 데이터에 접근할 수 있도록 수단을 제공하는 것이다[1].

현재 공공부문의 데이터 식별체계는 주체 별로 혼재되어 있다. 행안부의 공공데이터의 경우 행정표준코드관리시스템(code.go.kr)을 통해 업무 코드 식별은 가능하나, 범정부 차원의 표준화된 데이터 코

드 체계는 부재한 상황이다. 공공데이터포털은 개방데이터의 식별을 위해 별도 데이터 식별 체계(ID 등)가 아닌, 데이터 목록명(기관명_데이터명)을 데이터 식별자로 활용하며 동일한 목록명의 중복 등록은 불가한 상황이다.

민간에서도 각 플랫폼 또는 센터별로 자체 식별값을 적용 중이다. 의료분야의 경우 관리 주체별로 환자 식별체계를 갖추고 있으며, 이를 통합/활용하기 위해 각 병원 별로 임상 데이터 웨어하우스(CDW)를 구축 중이다. 공간정보 데이터는 공간정보 데이터베이스의 효율적인 구축·관리 및 활용을 위하여 건물, 도로, 하천, 교량 등 공간상의 주요 객체에 대하여 공간객체 등록번호를 부여하여 관리 중이다.

민간-공공을 포괄하는 데이터 생태계에서 유통·관리를 위해 데이터 객체식별자를 적용하고 거래정보, 이력정보 등을 추가할 수 있도록 확장·변환 표준 마련이 필요하다. 우선적으로 ITU-T와 ISO가 공동으로 개발한 광범위하게 적용 가능한 식별 메커니즘으로 OID(Object IDentifier)를 이용한 방법이 필요하다.

3.4 데이터 참여자 신뢰 기반

참여자 신뢰보장은 데이터 생태계에 참여하기 위해 필요한 최소한의 기준을 정의하여 검증 가능한 자격증명과 자격의 데이터 이용 환경을 연결하는 것이다[5]. 데이터 생태계에서 데이터 상품을 유통·거래하기 위해서는 데이터 제공자가 신뢰 가능한 데이터 상품을 적법하게 판매하고 있는지, 데이터 이용자가 데이터를 구매하여 도용, 도난, 대가 미지불 등 불법적으로 사용하지 않는지 데이터 유통·거래의 리스크를 최소화할 필요가 있다[5].

EU의 GAIA-X에서는 참여자(공급자, 수요자) 간 인증·접근관리 등을 지원하는 신뢰할 수 있는 데이터 인프라(GAIA-X 노드)를 제공하고 있다. 여기서는 특정 대상에게 일회성 자격을 부여하여 데이터를 선택적으로 제공하고, 연합서비스(Federation Service)를 통해 참여자 관리·신뢰메커니즘·데이터 접근 및 사용 권한 보장 등을 준수하고 있다[14].

국내 공공분야의 경우 공급자로서의 참여자가 각 부처, 지자체, 공공기관이 되기 때문에 공공데이터포털을 통해서 제공되는 데이터의 신뢰성은 국가가 보장하는 것으로 간주할 수 있다. 민간의 경우 참여자에 대한 인증체계는 마련되어 있지 않은 것으로 파악되었으며, 범 국가 차원에서 데이터가 유통, 거래되기 위해서는 데이터를 제공하는 민간의 제공자가 필요한 자격, 요건을 참여자가 준비하여 별도의 제3의 기관에게 공인받을 수 있도록 하되, 자체 인증(Self-Descripiton) 개념의 도입이 필요하다. 유사 인증체계를 도입한 EU의 GAIA-X 사례[15]를 참조하여 공급자·수요자 간 인증 및 신뢰체계와 이에 기반한 데이터 활용모델의 도입이 필요하다.

3.5 데이터 연계 및 교환

플랫폼 간, 정보시스템 간 데이터의 원활한 연계와 교환이 가능하기 위해서는 크게 소재정보의 표준화된 방식의 공유와, 연계-교환 기술의 표준화가 필요하다. 과기정통부가 지원한 21개 빅데이터 플랫폼의 소재정보를 통합 관리하는 통합데이터지도에서는 빅데이터 플랫폼에 구축·적재된 데이터의 메타데이터를 표준화된 형식(RDF/XML)으로 소재정보를 연계 API를 통해 전송하고 있다. 향후 국가 데이터 통합 인프라에서는 공공데이터 포털 등 다양한 플랫폼들과의 통합 소재 정보 제공이 가능하도록 기능 및 연계 확대와 표준 카탈로그화 추진이 필요하다.

플랫폼간 데이터 연계 및 교환을 위한 인터페이스/연계 표준은 국내에는 현재 없는 것으로 파악되었다. 통합 데이터지도는 국제표준 DCAT ver2에서 확장하여 만들어진 RDF 방식의 XML 형태를 적용중이며, 국가공간정보포털(K-geo플랫폼), 공간정보 오픈플랫폼(v-world) 등 다양한 플랫폼에서는 각자의 방식(EAI, ESB, OpenAPI)으로 연계 서비스를 제공 중이다. 금융 데이터는 신용정보법에 따른 마이데이터사업자와 정보제공자 간 금융분야 마이데이터 표준API 규격의 연계 표준이 별도로 제정되어 있다. 향후 국가데이터 표준화 과제로 빅데이터 플랫폼간 인터페이스/연계 표준화를 추진하고 공공데이터 포털을 포함한 공공·민간 공통 표준으로 확대 추진이 필요한 상황이다.

3.6. 데이터 주권 기반 이력관리

데이터 주권은 자기주권 신원증명 개념과 같이 자신의 데이터를 보호하고 공정하게 데이터의 사용 방법과 목적을 결정하는 권리[2]로 개인이 정보주체로서 자신의 데이터를 직접 확인, 통제할 수 있는 권리를 보장하는 개념으로 이해된다[4]. 즉 자신의 데

이터가 어디서 어떤 목적으로 어떻게 사용될지를 스스로 결정할 수 있는 권리이다.[5]

현재 국내 각 데이터 플랫폼에서는 이용약관 및 데이터 공급 및 거래 약관을 마련하여 서비스 이용, 개인정보보호, 데이터상품 구매 및 환불, 저작권의 귀속 및 이용제한 등 플랫폼과 이용자의 권리·의무 및 책임 사항이 규정되어 있고, 거래 약관에는 데이터센터의 판매행위, 데이터상품 품목 및 공급단가, 서비스 이용료, 지식재산권 및 사용권한 등 법률적 고려사항이 제시되어 있다. 데이터 제공자와 플랫폼 중개자의 이익 등 권리 보당 수단은 플랫폼 별 다양성과 업무 분야 특성으로 공통적 기준 마련이 쉽지 않다. 따라서 분야별로 추진하되 공통 기준에 대해서는 데이터 거래 수수료와 정산, 환불 등에 대한 참조 가이드라인 제공이 필요하다.

주권 기반 이력관리 관점에서는 현재 국내에는 데이터 유통 이후에도 원천 데이터의 소재 파악 및 이동 경로를 추적을 할 수 있는 이력 관리가 아닌, 데이터 거래정보에 대한 일반적인 관리만 수행되고 있다. 다양한 플랫폼에서 유사한 데이터나 중복된 데이터가 서로 다른 기관에서 수집, 생성될 수 있으므로, 데이터 개방·유통 과정에서 데이터 간 구분이 필요하며, 앞에서 언급한 데이터 식별체계에 기반하여 데이터의 유통·거래 이력을 관리할 수 있는 공통시스템이 필요하다. 다만 빅데이터의 특성, 물리적 관리 범위의 불가항력으로 인해 중앙집중형 관리체계는 불가하므로 블록체인 기술을 기반으로 한 분산 관리 체계를 도입해야 한다.

4. 결론

그동안 행정정보 공유, 공공데이터 개방, 빅데이터 플랫폼 구축 등 정부 주도로 국가 전반에 데이터를 공급하고 개방, 활용하기 위한 정책적 노력들이 추진되어 왔다. 이제 본격적으로 데이터 생태계와 데이터 경제가 실현되기 위해서는 공급자 관점에서 수요자 관점으로의 전환이 필요한 시점이다. 본 연구에서 제안한 국가 데이터 통합인프라는 모든 참여자가 자유롭게 데이터를 유통, 거래할 수 있는 새로운 장(場)이 되어, 정부가 공급자 역할에서 조정자 역할로 전환할 수 있는 출발점을 제공한다는 점에서 의의를 가진다.

참고문헌

[1] Altman, Micah, and Mercè Crosas. 2013. “The Evolution of Data Citation: From Principles to Implementation.” IAssist Quarterly, 37: 62–70.

[2] European Commission, (2017), European Data Economy

[3] Gartner Research, Data Catalogs Are the New Black in Data Management and Analytics, 2017.12.

[4] IITP, (2020), Weekly Technology Trend, Availability and Development
Status of Blockchain Technology in the Data Sovereignty Era, no. 1956,

[5] 김근형. 신뢰가 보장된 데이터 경제를 위한 자기주권 데이터 유통플랫폼 설계. 디지털콘텐츠학회논문지, 22(3), 483-490, 10.9728/dcs.2021.22.3.483, 2021

[6] 김병선, 안종욱, & 신동빈. . 서비스 중심의 국가공간정보 플랫폼 연계 방안에 관한 연구. 한국공간정보학회지, 22(2), 11-18. (2014)

[7] 박진호 & 한상우. 공공데이터 포털과 개별 포털 간의 데이터 연계방안 연구. 한국비블리아학회지, 33(4), 249-269. 2022.

[8] 유순덕, 최광돈, 신선영. 빅데이터 산업 활성화 전략 연구. 디지털융복합연구, 12(4), 1-9. 2014

[9] 이상은, 김인숙, 이규엽, GAIA-X 현황 및 도전과제 분석, 디지털플랫폼정부 시리즈. 2022.

[10] 임태훈. 데이터 생태계의 시작, 데이터 플랫폼. 한국통신학회지(정보와통신), 39(11), 3-8. 2022.

[11] 임철홍. 공공·민간 데이터 플랫폼의 데이터 유통과 활용성 향상을 위한 데이터 카탈로그 구축 방안 연구. 정보기술아키텍처 연구, 19(3), 217-228. 2022.

[12] 정의현, 데이터 유통 및 데이터 품질 기술동향, AI Network Lab 인사이트, 6-8. 2021.

[13] https://www.w3.org/TR/vocab-dcat-2/#introduction

[14] https://gaia-x.gitlab.io/policy-rules-committee/trust- frame work/gaia-x_trust_framework/

[15] https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html