

# 공공데이터 통합과 검증 시스템의 필요성 (질병예측모델을 위한 통합 데이터베이스 구축 사례 중심)

장우영\* · 손호선\*\* · 이영성\*\*\*

## I. 서론

오늘날 데이터의 의미는 정보사회를 움직이는 핵심으로 특히, 빅 데이터 활용을 통한 데이터의 가공과 분석에 따라, 상황인식, 문제해결, 미래전망 등을 예측하며, 데이터가 경제적 자산과 경쟁력의 척도로 사용되고 있다. 정부3.0이란 공공데이터의 개방을 통해 투명한 정부 실현 목적으로, 이를 통해 국민에게 맞춤형 서비스를 제공하고 일자리 창출과 창조경제를 지원하는 새로운 정부 운영 패러다임이다.

빅 데이터시대는 데이터가 가지고 있는 본래의 가치 뿐 아니라 새로운 가치를 찾고 그 가치를 활용하는 시대이다. 따라서 쉽게 데이터 수집이 가능한 공공데이터에 관심이 집중 되고, 공공데이터의 이용자 수가 크게 증가하고, 이를 이용하여 새로운 가치를 창출하는 것이 사회적으로 큰 이슈가 되고 있다. 많은 연구 분야의 이용자들이 공공데이터를 사용하기 시작하였고, 서로 다른 형태의 공공데이터 조합으로 새로운 가치 창출을 위한 도전이 시도되고 있으며, 이러한 공공데이터 통합에 대한 방법론은 다양한 형태로 제시되고 있다. 빅 데이터 활용에 효과적으로 대응하기 위해서는 데이터의 생산, 유통, 소비 등 그에 따른 많은 문제점이 제기되고 있다.

공공 데이터 활성화를 위해 정부는 선순환 생태계 조성에 노력하고 있으며, 선순환 생태계는 법제도의 개선, 정부의 적극적인 지원, 기술적인 지원 등을 통해 이루어 질 수 있다. 이처럼 공공데이터의 활용 및 통합을 통해 새로운 가치를 찾는 것이 빅 데이터 시대의 필요한 요소이다. 따라서 본 논문에서는 실제 공공데이터를 이용하여 새로운 가치 창출 사례의 하나인 질병예측모델 개발을 위한 통합 데이터베이스 구축 사례를 통해 공공데이터 통합의 필요성을 제시하고, 활성화를 위한 새로운 생태계를 제안하고자 한다.

## II. 본문

### 1. 공공데이터 통합의 필요성

우리나라 국민소득의 증가와 고령화 사회 진입으로 다양한 형태의 의료소비가 나타나고 있으며, 이것은 의료서비스산업에 큰 영향을 미치고 있다. 이러한 사회구조의 변화에 따라 기상, 지리 등의 환경변화에 따른 의료 소비가 기존의 질병 형태에서 새로운 질병으로 변화되고 있다. 이에 따라 새로운 질병에 대한 대응이 절실히 요구되고 있다. 통합 데이터베이스를 활용한 질병 예측 모델은 기존의 특정 질병뿐만 아니라 사회구조 및 기상변화에 따른 새로운 질병, 나아가 의료소비를 예측할 수 있다. 하지만 아직 국내에서는 기존의 특정 질병에 대한 예측 모델만이 연구되고, 새로운 질병과 더불어 기존의 특정 질병까지 아우르는 일반화된 예측

\* 장우영, 충북대학교 의생명과학경영융합대학원 석사과정, 043)249-1771, jangwy8838@gmail.com

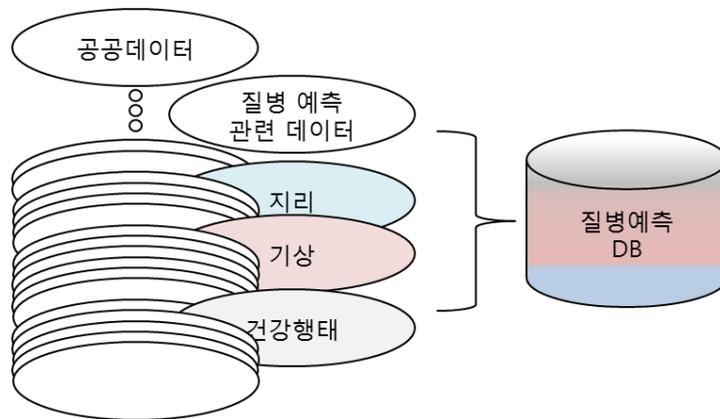
\*\* 손호선, 충북대학교 의생명과학경영융합대학원 교수, 043)249-1770, shon0621@gmail.com

\*\*\* 이영성, 충북대학교 의생명과학경영융합대학원 교수, 043-261-2869 lee.medric@gmail.com

모델이 존재 하지 않는다. 또한, 질병에 영향을 주는 요인이 다양함에도 불구하고 많은 연구에서는 오직 단일 요인에만 접근하고 있어 보다 정확한 예측이 어렵다. 따라서 새로운 질병에 대응하기 위해서는 건강행태·상태, 기상, 지리, 인구 등의 다양한 요인을 적용 시킨 일반화된 질병 예측 모델이 필요하다.

많은 공공데이터에서 원하는 가치를 가진 데이터의 탐색은 가치 창출에서 가장 중요한 부분이다. 질병예측 모델 개발에서는 정확한 질병 예측이 가장 중요한 부분이며, 질병과 관련된 건강상태·행태, 지리, 기후, 인구 등 다양한 요인을 탐색하는 것이 필요하며 이러한 데이터는 질병관리본부, 통계청, 기상청, 건강보험심사평가원 등에서 자료를 수집할 수 있다.

수집된 데이터들은 반정형화 또는 정형화 되어 있다. 예를 들어, 설문 조사를 통해 수집된 원천 데이터는 반정형화 형태로 제공되며, 지상기상관측, 인구 정보 등은 정형화된 데이터로 제공된다. 이러한 데이터를 어떻게 활용하는지에 따라 수많은 새로운 가치를 가진 데이터가 생성된다. 따라서 활용 목적에 따라서 데이터를 어떻게 가공할지를 고려해야한다.

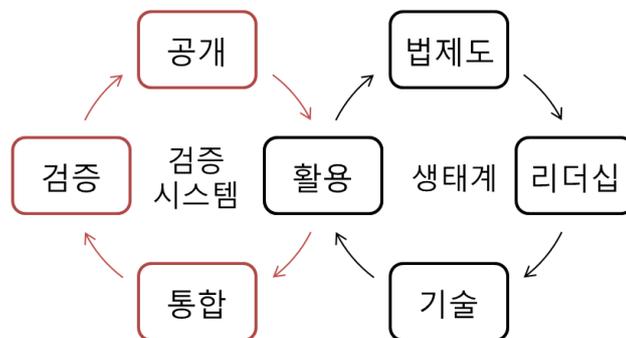


(그림 1) 공공데이터 탐색과 통합

## 2. 통합 공공데이터 검증시스템의 필요성

통합 공공데이터의 검증시스템이란 공공데이터로 창출된 새로운 가치를 갖는 공공데이터의 분류, 검증, 공개과정을 포함하며, 창조경제 실현을 위한 검증 시스템이다.

공공데이터 통합의 필요성이 표면적으로 나타나면서 공공데이터의 통합을 통해 새롭게 가치를 창출하는 과정을 각각의 연구에서 개별적으로 이루어지고 있다.



(그림 2) 확장 정부-민간 데이터 개방 선순환 생태계

질병예측모델개발을 위한 데이터베이스 구축과정에는 많은 자원이 투입되었다. 또한 기존의 질병예측모델 개발 연구에서도 이러한 과정을 선행하였다. 선행 연구의 데이터를 제공받는다 하더라도 데이터의 완결성을 다시 확인하는 과정을 반복해야하기 때문에 활발한 연구가 이루어지기 힘든 실정이다. 따라서 검증시스템의 도입을 통입 한다면, 초기에 들어가는 비용과 시간을 감소시킬 수 있다. 이는 공공데이터의 활성화를 뜻하며 지식의 축적이라 할 수 있다.

### III. 결론

의료서비스 분야뿐 아니라 여러 분야에서 새로운 가치의 필요성이 표면화되고 있다. 정부 3.0 정책으로 공공데이터가 증가하고 이를 통해 새로운 가치의 창출이 가능하게 되었다. 하지만 새로운 가치의 창출을 위해서는 데이터의 설계, 전처리, 통합 등 초기의 비용이 많이 필요하며, 이를 검증해주는 시스템의 부재로 인해 지식의 축적이 불가능하다. 질병예측모델개발이라는 새로운 가치창출을 위해 질병과 관련된 서로 다른 공공데이터를 통합하여 새로운 가치를 창출함에도 불구하고 이를 활용하기 위해서는 검증 시스템 도입이 절실하다.

기존의 선순환 생태계에서 검증 단계를 통해 공공데이터의 새로운 가치창출로 인한 창조경제의 실현이 기대된다.

### 사사표기

본 논문은 2013년도 미래창조과학부의 재원으로 과학벨트기능지구지원사업의 지원을 받아 수행된 연구임 (2013K001552).

### 참고문헌

- 관계부처합동 (2013) 「공공데이터의 제공 및 이용활성화 기본계획(13년~17년)」, 서울: 공공데이터전략위원회.
- 배인수 (2012), 「빅데이터 시대: 효과적인 공공정보 개방을 위한 데이터 플랫폼 구축방향」, 서울: 한국정보화진흥원
- 장우영 (2014), “질병 예측 모델 개발을 위한 지리정보시스템(GIS)기반 데이터베이스 구축” 「2014년 춘계학술발표대회 논문집」,21(1) : 708-711.
- 정국환 (2013) 「공공데이터 개방·활용 성과측정을 위한 평가모델 연구」, 충청북도: 정보통신정책연구원  
<https://www.gov30.go.kr/>