개인 정보를 보호하는 보건의료 빅데이터 연계 및 표준화 프로세스

Healthcare bigdata linkage and standardization process with privacy protection

김 현 준, 정 승 현, 이 경 희, 조 완 섭*
충북대학교

Kim hyun-joon, Jung seung-hyun, Lee kyung-hee, Cho wan-sup*
Chungbuk National Univ

요약

데이터의 다양성은 빅데이터를 이용해 새로운 가치를 창출하는데 있어 매우 중요하다. 데이터의 다양성을 위해서 다양한 데이터의 연계는 필수적이며, 여러 활용영역 중에서도 보건의료분야에서의 데이터 연계는 그 요구가 특히 증가하고 있다. 또한 활용성에 있어서도 높은 기대전망이 있는 분야이다. 그러나 보건의료 데이터의 연계는 개인정보 중에서도 많은 민감 정보를 포함하고 있기 때문에, 이에 관한 개인정보 보호에 대한 이슈 해결이 선행되어야하며, 데이터 연계에 관련 있는 주체간의 합의 역시 선행되어야 한다.

I. 서론

1. 연구의 배경과 목적

최근 빅데이터에 대한 전 세계적인 관심이 집중되고 있는 가운데, 빅데이터를 활용하여 새로운 가치를 찾음에 있어 데이터의 다양성은 매우 중요한 부분이다. 데이터의 다양성이 확보되었을 때, 그동안 발견하지 못했던 인사이트를 발견할 수 있고, 좀 더 정확한 미래에 대한 예측이 가능해 질 것이다. 이러한 측면에서 다양한 기관이 가지고 있는 데이터의 연계에 대한 필요와 요구가 커지고 있다.

그러나 데이터 연계를 논할 때 필수적으로 등장하는 이슈 중 하나는 '개인정보 보호'문제로, 데이터 연계가 이루어지기 위해서 가장 먼저 선행되어 해결해야하는 이슈이다. 개인정보를 보호하기 위해 무조건적으로 이용을 제한하기 보다는 '안전하고 투명한 정보의 분산'을 통해 '보호'와 '활용'의 두 가지 가치를 적절히 고려해야할 필요가 있다.

빅데이터에서 여러 분야의 활용성 중에서도 빅데이터의 도입, 활용이 매우 전망되는 분야 중 하나로 보건의료분야가 꼽히고 있다. 다양한 보건의료데이터를 연계함으로써 건강과 질병의 패턴 발견 등에 크게 기여할 것으로 기대되고 있다.

이에 개인 정보를 보호하면서 보건의료 데이터를 연계하여 분석하고 활용할 수 있는 데이터 연계 프로세스를 제안하고자 한다.

Ⅱ. 본론

2.1 참여 기관

본 데이터 연계 프로세스에 있어 참여 기관은 데이터를 보유하고 있는 기관을 포함하여 크게 다섯 분류로 나눌 수 있다. 첫째, 데이터 연계를 요청하고 의뢰하는 기관(병원), 둘째, 연계 요청을 심의하고, 개인 식별자(주민등록번호)를 암호화하는데 필요한 암호키를 할당하며, 연계에 필요한 데이터를 보유하고 있는 각 기관에 연계 요청을 하는 기관(심의 및 암호화담당기관), 셋째, 연계를 위한 데이터를 보유 및 제공하는 기관(보건의료 데이터 보유 기관), 넷째, 데이터 연계를 위해 각 암호화된 개인 식별자에 임의의 연계번호를 부여하는 기관(연계담당기관), 마지막으로 개인 식별자가 제거된 데이터를 연결번호를 통해 연계 후 안전한 저장소에 저장하는 기관(연계플랫폼)이다.

2.2 개인 식별자에 대한 암호화

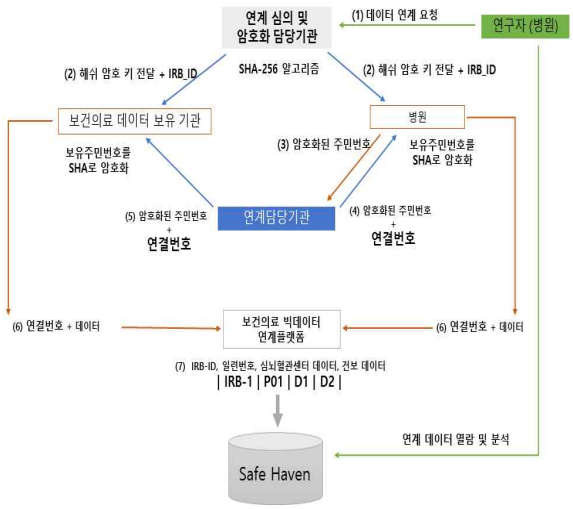
개인 정보에 있어 가장 중요한 개인 식별자(주민등록번호)는 원문 그대로 주고받을 수 없기 때문에 반드시 암호화를 거쳐야 한다. 따라서 개인 식별자 암호화는 권장 보안강도인 112bit 이상의 단방향 해쉬 암호화 알고리즘인 SHA-256 알고리즘을 사용하여 암호화한다.

2.3 Open API 표준화

다양한 기관이 가진 데이터를 연계하기 위해서는 약속되어진 신뢰성 있는 데이터 통신이 필요하다. 이를 위해서는 표준화된 Open API 설계가 필요하며, 각 기관은 연계 프로세스에 따라 데이터 전송 시에 다음의 API를 사용하여 데이터를 전송해주어야만 연계 플랫폼에서 표준화된 형식의 연계 데이터 셋을 생성할 수 있다.

2.4 연계 프로세스

개인정보를 보호하기 위해서 가장 중요한 것은 한 기관이 지나치게 많은 정보를 갖지 않아야 한다는 것이다. 각각의 정보만으로는 연계 데이터에서 개인을 유추할 수 없지만, 정보의 퍼즐이 맞혀지게 된다면 연계 데이터로부터 개인을 유추할 수 있게 되기 때문이다.



▶▶ 그림 2. 연계 프로세스 과정

따라서 그림 2와 같은 프로세스를 통해 심의 및 암호화담당기관, 보건의료 데이터를 보유하고 있는 기관, 연계담당 기관, 연계 플랫폼이 정보를 나눠 갖으며 최종적으로 Safe Haven이라 불리는 안전한 저장소에 연계된 데이터를 저장하게 된다.

이렇게 저장된 데이터는 처음 연계를 요청한 연구자만이 접근할 수 있도록 하며, 데이터의 열람 및 분석 역시 Safe Haven에서 벗어나지 않도록 제한한다. 또한 별도의 요청이 없다면 일정 기간 경과 후 접근하지 못하도록 제한한다.

연계 저장된 데이터 역시 일반적인 상황에서는 Safe Haven에 영구 저장하는 것이 아닌, 일정 기간 저장 후 폐기하여 외부 유출의 가능성을 최소화 한다.

Ⅲ. 결론

본 논문은 보건의료분야의 데이터를 연계함에 있어 개인정보를 보호하기 위한 프로세스를 설계 하였다. 이와 같은 프로세스를 통해 개인수준의 연계 데이터가 연구자들에게 제공된다면, 개인정보의 노출은 최소화 하면서 제한적인 데이터로 인해 발견하지 못했던 건강과 질병의 패턴 등 새로운 가치를 발견하고 창출하는데 기여할 수 있을 것이라 기대된다. 하지만 이와 같은 프로세스가 실질적으로 각 분야에서 사용되기 위해서는 각 기관이 사용하는 용어 및 코드에 대한 표준화가 필요할 것이며, 더욱 중요한 것은 개인정보 이용에 대해 폐쇄적인 기관들의 인식변화와 법적 규제의 적절한 변화가 선행되어야 할 것이다.

본 연구는 산업통상자원부 국가기술표준원에서 지원하는 국가표준기술력향상사업의 일환으로 수행되었음(과제번호: 10053645)

▌참 고 문 헌▐

[1] “SHIP : A Blueprint for Health Records Research in Scotland”, http://www.scot-ship.ac.uk, No. 10, 2012.
[2] 정영철 “의료분야 빅데이터 활용을 위한 개인정보 비식별화 규정 현황과 과제”, 제227호, pp.50-60, 2015.
[3] 오미애 “보건복지분야 데이터 연계 필요성 및 활용방안”, 제227호, pp.17-28, 2015.