

Abstract Submission No. : IL-9116

## CDM 기반 바이오-헬스 빅데이터 임상활용 최신지견

Rae Woong Park  
Ajou University School of Medicine, Korea, Republic of

### CDM 기반 바이오-헬스 빅데이터 임상활용 최신지견

환자의 임상정보를 활용한 빅데이터 분석은 새로운 건강서비스 기획과 맞춤 신약, 융합 의료기기 개발을 함에 있어 4 차 산업의 핵심이라고 봐도 무방하다. 그러나 의료 데이터는 데이터 구조, 형식의 이질성, 데이터의 질과 양 등 기술적인 어려움과 기관의 허락, 개인정보보호문제 등 법적 문제 그리고 타인에게 제공하는 데이터가 자신에게 불리하게 사용될지 모른다는 두려움 등 자신의 자료를 타인과 공유하고 싶지 않는 문제를 들 수 있다.

이런 제약을 극복하는 방법 중 하나가 분산연구망이다. 이는 의료데이터 유통의 제약 요인 극복을 위해 병원과 기업 등 수요자 간에 원본데이터의 공유 없이 분산된 형태로 데이터를 관리하면서 기관 내에서 분석한 결과만 공유하는 방식이다. 이를 위해서 각 기관별로 서로 다른 데이터 구조와 의미를 동일한 하나의 구조와 의미를 갖도록 변환하는 방법이 공통데이터모델 (Common Data Model, CDM)이다.

분산연구망은 각 병원의 환자정보를 표준화 및 익명화한 CDM 으로 변환한 후 데이터보유 폐쇄망 안에 두고, 사용자의 요청에 따라서 분석된 요약 집합정보(평균, 합, 표준편차, 오즈비, 위험도 등)만 사용자에게 회신하는 방식이다. 사용자는 폐쇄망 안에 있는 환자의 개별 정보를 보거나 취득할 수 없음에도 불구하고, 궁극적으로 원하는 결과값을 얻을 수 있는 방식으로 현실적으로 법개정없이도 의료 빅데이터를 공유할 수 있는 방식이다.

전세계적으로 20 개국 200 개 이상의 기관이 참여하고 있으며 국내에서는 60 개 이상의 기관이 참여 중이다. 한국오딧세이, 유럽오딧세이, 중국오딧세이가 결성되어 있으며 유럽에서는 유럽연합이 지원하는 에덴 (EHDEN)프로젝트가 2018년 11월부터 출발하여 5년간 진행간 중이며 유럽 12 개국이 참여하고 있다. 한국에서는 산자부 지원으로 분산형바이오헬스 빅데이터 사업단이 출범하여 2018년부터 5년간 진행 중에 있다.

최근 Covid-19 사태를 맞이하여 전세계 30 개국 360 여명의 과학자들이 온라인상으로 모여서 3일간 온라인 연구마라톤을 개최하였고 이 때 한국을 포함하여 8 개의 기관에서 Covid-19 환자의 임상데이터를 CDM 형태로 분산연구망에 공개하였고, 그외 30 여개의 CDM 데이터가 공유되어 많은 연구 성과를 이루었다. 국제위기 상황에서 CDM 기반 분산연구망이 위기 극복을 위한 좋은 수단임이 증명된 것이다. 최근 CDM 을 이용한 다기관 연구도 급격히 증가하고 있다. 그 결실로 고혈압 치료제나 antiplatelet agent 간 효과 및 부작용 비교 연구가 Lancet, JAMA 등 주요 임상논문지에 연구결과가 실리는 성과를 내고 있다.

분산연구망은 후향적 자료에 기반하고 있다. 만약 실시간으로 타 기관의 자료를 수집해 활용할 경우 병원정보시스템 이용에 큰 도움이 될 것이다. 다양한 활용을 위해 데이터 교환 표준인 FHIR 와 데이터 저장 표준인 CDM 을 결합할 필요성도 대두되고 있다. 임상정보에 국한된 CDM 모델을 확장하여 생체신호, 라이프로그, 유전체 정보, 레지스트리, 영상정보를 포괄할 수 있는 실시간 데이터저장소 형태로 확장시킬 경우 진정한 의료 빅데이터 시대로 한걸음 더 다가설 수 있게 될 것이다.