

ISSUE REPORT

데이터 경제의
새물결,
혁신 사례



VOL

03

헬스케어 시장의

핵심 역량이 된 '데이터'

헬스케어 시장의 핵심 역량이 된 ‘데이터’

글 | 이지현 IT전문기자 j.lee.reporter@gmail.com / 우창완 선임연구원 woo@nia.or.kr

의료업계에서 데이터는 복잡한 요소다. 공공부터 민간까지 의료 데이터를 생성하는 곳은 많지만 막상 이를 쓰려면 기술적인 요소부터 법적인 요소까지 넘겨야 할 관문들이 많다. 그럼에도 의료계에 기술혁신을 불어넣고 있는 일명 ‘헬스케어’ 시장에서는 데이터를 활용하려는 노력이 끊임없이 이뤄지고 있다. 최근 코로나 바이러스 사태가 더해지면서 그 활용 사례는 더 많아지는 추세다.

의료 학계 협업의 중심 ‘오픈엑세스’

먼저 대학 내 풍경을 보자. 요즘 코로나와 연계된 생물정보학, 생명공학, 의료 학계에서는 전례 없는 협업이 진행되고 있다. 학계 전체가 코로나와 관련된 연구결과나 데이터를 확보 할때마다 외부에 공개하고 이에 대한 피드백을 활발히 주고 받고 있는 것이다. 그 핵심에는 ‘오픈엑세스(Open Access)’라는 운동이 있다.¹ 오픈엑세스란 비용과 기술 제약없이 누구나 연구결과를 열람하고 동시에 공유하는 문화를 말한다. 이에 반대되는 개념은 유명 학술지에 논문을 기고하고 공유하는 방식이다. 과거 학계에서는 네이처, 사이언스같은 유명 학술지에 연구결과를 공개했는데, 이 과정에서 일부 연구자들은 비용 부담으로 연구결과를 보지 못하는 경우도 있었다. 이런 현상을 극복하고자 시작된 것이 오픈엑세스이며, 최근은 논문 내용뿐만 아니라 연구에서 활용된 데이터도 함께 공유되는 추세다. 이를 따로 ‘오픈 데이터’라고 부르기도 한다.

1 코로나, 오픈 사이언스 운동에 불을 지피다. <https://blog.lgcns.com/2234>



최근 몇 년간 오픈액세스가 성장하고 있긴 했으나 그렇다고 주류 문화라 할 수 있는 수준은 아니었다. 2018년 기준 오픈액세스 형태로 공개된 자료는 전체 논문의 36% 정도². 불과 2020년 2월까지만 해도 코로나 바이러스와 관련된 논문 수는 약 500개였는데 이 중 약 160개³가 오픈액세스 형태였고, 나머지 논문은 사실상 열람하기 위해 돈을 지불하는 상황이었다. 하지만 코로나 사태가 전세계로 확산되면서, 글로벌 주류 학술 출판지 대다수는 코로나바이러스와 관련된 모든 논문은 모두 공개하는 정책을 시행했다.⁴ 데이터도 무료로 개방하고 다른 연구를 위해선 라이선스 제약 없이 재사용하도록 열어두고 있다. 사스(SARS)나 메르스(MERS) 발병 시기에도 볼 수 없었던 풍경이다. 출판사 및 대학들은 각자 별도의 웹페이지를 만들어 코로나 연구와 관련된 논문과 데이터를 공유하고 있으며, 정부나 교육기관이 중심으로 만든 오픈액세스 플랫폼인 바이오

2 https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/trends-open-access-publications_en#open-access-to-publications

3 <https://theconversation.com/the-hunt-for-a-coronavirus-cure-is-showing-how-science-can-change-for-the-better-132130>

4 Publishers make coronavirus (COVID-19) content freely available and reusable <https://wellcome.ac.uk/press-release/publishers-make-coronavirus-covid-19-content-freely-available-and-reusable>

아카이브(bioRxiv), 메드아카이브 (medRxiv), 펍메드(PubMed) 등에서도 관련 논문이 활발히 공유되고 있다. 중국의 경우 국가과학기술 관련 학술지 플랫폼 CQVIP⁵를 통해 1만 4천 여개의 코로나 관련 논문을 무료로 공개하기도 했다. 이번 사건을 계기로 앞으로 대학계에선 연구성과를 공유하는 문화가 더 확장될 것으로 보인다.

코로나 사태로 데이터 공유 플랫폼에 대한 인기도 높아지고 있다. 이때 기존 공공데이터 포털 사이트와는 별도로 정부와 외부 기관이 협업해 만든 데이터 공유 플랫폼들이 특히 성장하고 있다. 먼저 미 백악관, 미국 국립보건원은 AI2⁶라는 연구기관과 협업해 '코로나19 오픈리서치 데이터세트 (COVID-19 Open Research Dataset)'를 공개했다. 코로나 관련 논문 5만 7천여 개와 메타데이터를 검색하고 다운로드를 할 수 있는 곳이다. AI2는 이 코로나19 오픈리서치 데이터세트를 활용한 공개 아이디어 경연대회⁷를 열고 있는데 MS, 페이스북, 구글 등의 지원을 받아 상금도 제공하고 있다. IBM의 경우 코로나19 오픈리서치 데이터세트를 자체 인공지능 트레이닝에 활용했고, 코로나 관련 데이터를 보다 정교하게 검색할 수 있는 플랫폼⁸을 무료로 개방했다.

독일 비영리 단체⁹가 운영하는 'GISAID'¹⁰도 최근 주목받는 데이터 공유 플랫폼이다. 이곳에선 주로 인플루엔자 바이러스와 관련된 염기서열 정보, 임상실험 및 유행병 정보를 찾을 수 있다. 미국 국립과학재단도 국립 인간유전체 연구소, 존스 홉킨스 대학교, 펜실베이니아 주립대학교와 함께 GISAID와 비슷한 '갤럭시'¹¹를 플랫폼을 운영 중이다. 여기에선 의학공학 분야 데이터를 업로드하거나 특정 데이터를 추출하고 정렬하는 등 데이터 관리도구도 제공하고 있다. 중국

5 <http://www.cqvip.com/>

6 마이크로소프트(MS) 설립자 중 한명인 폴 알렌(Paul Allen)이 설립한 AI 연구소다. <https://allenai.org/>

7 <https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge>

8 <https://s2s-omxware.us-south.containers.appdomain.cloud/>

9 독일 정부 지원을 받아 운영되는 비영리단체다.

10 <https://www.gisaid.org/>

11 <https://usegalaxy.eu/>

정부가 운영하는 '국립 유전체학 데이터 센터'¹², 프레드 허친슨 암연구센터가 만든 '넥스트 스트레인'¹³ 에서도 다양한 염기서열 정보를 무료로 공개 중이다.

한국 코로나 관련 통계 정보도 인기다. 대신 데이터는 정부 공식 홈페이지가 아니라 '캐글'(kaggle)이라는 데이터 사이언스 경진 대회 플랫폼에서 주목받고 있다. 이곳에서 한국 개발자들은 해외 연구자들이 국내 자료를 쉽게 이해하고 분석할 수 있도록 국내 질병관리 본부에서 공개한 데이터를 영문화하고 정제해 공유하고 있다.¹⁴

헬스케어 최신 흐름 ① - 모든 데이터가 의료데이터가 될 수 있는 시대

이번엔 코로나 바이러스 이전에 보였던 헬스케어 시장 흐름을 살펴보자. 국내 헬스케어 전문 엑셀레이터를 설립하고 헬스케어 전문가로 활동하는 최윤섭 소장은 헬스케어 분야를 크게 세가지로 분류한다. 데이터를 측정하는 곳, 통합하는 곳, 해석하는 곳이다.¹⁵ 첫 번째 데이터 측정에 집중한 기업들은 의료기록, 유전정보, 스마트폰, 웨어러블 기기 등으로 데이터를 획득하고 있다. 두 번째 통합은 플랫폼 역할을 하는 기업들이다. 종류가 다양한 의료 데이터를 통합해 저장하고 데이터 포맷을 정리해주거나 호환성을 높여주는 기업이 여기에 속한다. 마지막 해석은 데이터를 최종 분석을 해주는 곳이다. 이런 기업은 인공지능이나 데이터 분석 기술을 활용해 맞춤형 의료 서비스, 조기 진단 기능을 제공한다.

이 중 최근 눈에 띄는 흐름은 다양한 데이터를 활용하는 시도다. 의료 데이터라 함은 일반적으로 키, 몸무게, 혈압, 심박수나 엑스레이 사진, 진료기록 등 병원기록을 떠오른다. 하지만

12 <https://bigd.big.ac.cn/>

13 <https://nextstrain.org/>

14 <https://www.kaggle.com/kimjihoo/coronavirusdataset>

15 최윤섭, 「디지털 헬스케어 의료의 미래」, 클라우드나인, 2020

[전세계 주목받고 있는 헬스케어 기업들]



🏠 cbinsights.com/research/digital-health-startups-redefining-healthcare/#startups

최근 인공지능 기술이 발전되면서 최근엔 전통적인 의료데이터 외에 새로운 데이터를 의료서비스에 결합하려는 시도를 종종 볼 수 있다.

예를 들어, 마인드스트롱 헬스(Mindstrong Health)¹⁶라는 스타트업은 정신 건강을 진단할 때 스마트폰 사용패턴을 활용한다. 타이핑 하는 속도나 자주 오타가 나는 단어들, 전화번호부 목록을 스크롤하는 속도 등을 데이터로 얻어와 이를 인공지능과 결합해 우울증을 파악해준다.¹⁷

16 <https://mindstrong.com/>

17 The smartphone app that can tell you're depressed before you know it yourself <https://www.technologyreview.com/2018/10/15/66443/the-smartphone-app-that-can-tell-youre-depressed-before-you-know-it-yourself/>

미 일리노이 대학도 비슷하게 '삭제' 버튼 이용 횟수, 휴대폰이 흔들리는 횟수같이 정보를 기반으로 조증 및 우울증을 알아내고 있다.¹⁸ 이를 위해 '바이이펙트'¹⁹라는 iOS 앱을 운영중인데 여기서 일반 시민과 우울증 환자들의 스마트폰 사용 패턴을 획득하고 있다.

소셜네트워크에 올라온 자료들도 최근 의료계에서 활용하고 있다. 하버드 대학과 버몬트 대학은 인스타그램에 업로드한 사진 필터 색감이나 사진 속 얼굴로 우울증 여부를 알 수 있다는 연구 결과를 발표했다. 우울증이 있는 사람들은 파란색, 회색 및 어두운 색깔의 필터를 자주 쓰거나 사진에 얼굴을 노출하는 경우가 더 적었다는 정보를 활용한 것이다.²⁰

치료 서비스 외에 의료계 정책을 정할 때도 새로운 형태의 데이터를 활용하기도 한다. 최근 애플은 사회적 거리 두기 흐름을 확인할 수 있도록 모빌리티 관련 데이터를 공개하고 누구나 다운로드 받을 수 있도록 만들었다. 이때 데이터는 애플 지도앱에서 생성된 검색어, 길찾기 경로, 교통정보 등을 활용했다고 한다. 이를 통해 정부기관이 사회적 거리두기 현황을 그래픽으로 파악할 수 있게 만들었다.²¹

18 <https://www.psych.uic.edu/in-the-news/can-typos-give-insight-into-your-mental-health>

19 <https://www.biaffect.com/>

20 인스타그램으로 우울증 감별 가능해 : 연구 <https://research-paper.co.kr/news/view/581>
Your Instagram Posts May Hold Clues to Your Mental Health <https://www.nytimes.com/2017/08/10/science/instagram-mental-health-depression.html>

21 Apple makes mobility data available to aid COVID-19 efforts <https://www.apple.com/newsroom/2020/04/apple-makes-mobility-data-available-to-aid-covid-19-efforts/>

헬스케어 최신 흐름 ② - “데이터 공유할 분을 찾습니다”

헬스케어 시장에서 데이터는 기업의 핵심 역량을 만들 수 있는 중요한 요소다. 그러다보니 데이터를 수집하고 획득하는 과정에서 많은 비용과 시간이 필요하다. 여기에 암호화나 개인정보 보호 기술도 요구돼 데이터를 공개하는 쪽이나 활용하는 쪽이나 기술적 장벽에 부딪힌다. 이런 환경이 반영돼 근래 의료계에서는 자발적으로 데이터를 공유하는 환자들을 찾고 있다. 사용자들은 자발적으로 데이터를 공유하면서 의료 연구에 도움을 주거나 맞춤형 서비스를 받는 혜택을 얻을 수 있고, 의료계는 데이터 획득을 보다 용이 하게 하는 것이다.

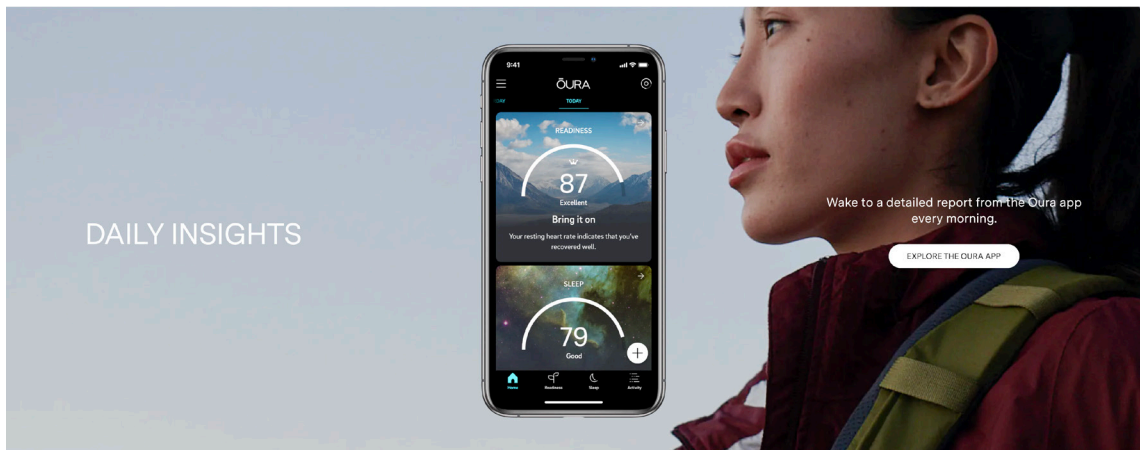
예를 들어 싱가포르정부는 코로나 사태 이후 ‘트레이스 투게더’²²라는 앱을 만들어 시민들의 동선을 파악하고 있다. 블루투스를 이용한 이 앱은 확진자의 이동경로를 추적할 수 있도록 기획됐고, 동시에 근처 확진자가 지나가면 알림을 주는 기능도 제공한다. 싱가포르 국민들은 이 앱을 사용하면서 모바일 앱에 담긴 자신의 정보를 정부와 자발적으로 공유하고 있으며, 싱가포르 정부는 해당 정보를 익명화하는 과정을 지원하고 있다. 관련된 전체 기술은 오픈 소스로 공개하기도 했다.²³

페이션츠라이크미(Patients Like Me)는 환자들을 위한 소셜네트워크 서비스다. 의학 정보 공유부터 친목 활동까지 이뤄지는 온라인 커뮤니티라고 볼 수 있다. 주요 사용자가 환자 이다 보니 이곳에선 특정 약, 병원, 건강정보 등 다양한 데이터를 다른 사용자와 공유하는 것을 자주 볼 수 있다. 2019년 기준 75만 명 회원이 2900개가 넘는 질환에 대해 이야기 하고 있으며 4300만개가 넘는 데이터가 축적돼있다고 한다.²⁴ 그러다보니 대학 연구소나

22 <https://www.tracetgether.gov.sg/>

23 <https://bluetrace.io/>

24 <https://www.patientslikeme.com/about>



제약회사들은 이곳에 데이터를 기반으로 약의 부작용을 검토하기도 하고, 미 FDA도²⁵ 내부 연구에서 이 소셜 데이터를 활용하고 있다. 페이션츠라이크미는 아예 오픈 리서치 익스체인지²⁶라는 데이터 플랫폼도 운영하는데 여기서 연구자와 환자들은 연결해 건강 정보를 검증하고 개발할 수 있도록 도와주고 있다.

스마트 반지 기업 오라²⁷는 최근 캘리포니아 대학교와 협업해 코로나 바이러스 증상 연구를 같이 하고 있다.²⁸ 이를 위해 데이터를 공유할 참여자를 공개 모집하고 있다. 오라 반지가 없는 사람에게는 반지도 직접 대여해준다. 이번 프로젝트에 참여한 사용자는 오라 반지로 측정하는 건강 데이터를 연구자와 공유하고, 매일 설문조사에 참여면서 추가 상황을 전달하고 있다. 오라와 캘리포니아 대학은 이렇게 모인 데이터를 기반으로 현장 의료진에 대한 건강을 추적하고, 궁극적으로 코로나19와 관련된 조기 증상을 더 정밀하게 밝혀낼 계획이다.

25 <https://www.mobihealthnews.com/content/patientslikeme-fda-explore-how-patient-generated-data-could-help-event-reporting>

26 <https://www.openresearchexchange.com/>

27 <https://ouraring.com/>

28 https://ucsf.co1.qualtrics.com/WRQualtricsControlPanel_rel/File.php?F=F_b9gHUu4nFjdMHmR

미국 정부는 의료 데이터를 공유할 국민들을 직접 모으는 중이다. 이를 ‘올 오브 어스’²⁹라는 리서치 플랫폼이라고 부른다. 2015년 미국 국립보건원(NIH)가 주축이 돼서 시작한 이 프로젝트는 자발적으로 의료 데이터를 공유할 100만 명 시민을 찾고 이들의 정보를 학계 및 민간 기업이 활용할 수 있는 기반을 만들고 있다.³⁰ 아직 100만 명을 달성하진 않았지만 지원자 수는 꾸준히 늘어나고 있다.

나이트스카우트³¹는 데이터를 외부에 공유하는 것은 아니지만 환자가 원하는 방향으로 데이터를 추출할 수 있도록 지원하고 있다. 비영리단체인 나이트스카우트는 혈당 측정 기기의 데이터를 클라우드 저장소에 업로드하고 스마트 기기에 연동할 수 있는 기술을 만들었다. 이를 통해 소아당뇨를 가진 환자 부모들이 자녀의 혈당 정보를 실시간으로 알 수 있게 돕고 있다. 이 기술은 현재 오픈소스로 공개됐고, 당뇨병 환자들의 커뮤니티 역할도 함께 하고 있다.

헬스케어 최신 흐름 ③ - 의료 시장에 확장되는 웨어러블 기기

웨어러블 기기는 과거 헬스케어의 부흥을 일으킬만큼 중요한 데이터 수집 매체였다. 초기에는 스마트 시계를 중심으로 이 시장이 성장했지만 최근엔 의료 장비 역할을 하는 웨어러블 기기들도 늘어나고 있다.

29 <https://allofus.nih.gov/>

30 All-of-Us: 모든 사람의 모든 데이터를 모으겠다!, <https://www.yoonsupchoi.com/2019/08/19/all-of-us/>

31 <http://www.nightscout.info/>

리프헬스케어³²는 가슴에 붙이는 센서를 개발한 곳으로 환자 자세를 모니터링할 수 있도록 도와 준다. 이 서비스는 욕창 위험이 있는 환자들을 관리하는 간호사들에게 특히 유용하다. 간호사들이 직접 환자 자세를 언제 변경해줬는지 기억할 필요가 없기 때문이다. 이 덕에 간호사들의 업무 효율을 올라갔으며, 동시에 환자들이 제때 치료를 받을 수 있는 환경이 조성됐다고 한다.

미국 스타트업인 아이리듬³³은 가슴에 붙이는 심전도측정기를 개발했는데 이를 통해 얻은 데이터로 부정맥을 진단할 수 있는 딥러닝 알고리즘을 개발했다. 이를 통해 병원 방문이 어려운 환자들을 치료할 수 있는 환경을 만들고 있다.

한국에서 스마트 벨트를 생산하고 있는 '웰트'는 센서 데이터를 기반으로 사용자의 낙상 위험을 조기에 알려주는 서비스를 개발했다. 웰트는 사업 초기 허리 둘레, 걸음수, 앉은 시간 등을 이용해 건강과 체형을 관리해주는 서비스를 제공했는데 2019년부터 노인 사용자를 공략하여 낙상 감지 기능을 추가했다. 분당서울대병원과 협업해 만든 이 기능은 걸음 속도와 보행 유형, 좌우 대칭 등을 데이터로 축적하고 이를 기반으로 비정상적인 보행을 찾아내는 원리로 구현했다.³⁴

참고문헌

- 최윤섭, 「디지털 헬스케어 의료의 미래」, 클라우드나인, 2020
- 홍윤철, 「팬데믹」, 포르체, 2020

32 <http://www.leafhealthcare.com/>

33 <https://www.irhythmtech.com/>

34 [CES 2020] 웰트, 애플워치에 없는 '낙상 감지 기능' 스마트 벨트 공개 https://biz.chosun.com/site/data/html_dir/2020/01/07/2020010701904.html

VOL

03

헬스케어 시장의 핵심 역량이 된 '데이터'

작성 IT전문기자 **이지현**

한국정보화진흥원 정책본부 정책기획팀 **우창완** 선임

기획 한국정보화진흥원 정책본부 **박원재** 본부장

한국정보화진흥원 정책본부 정책기획팀 **이정아** 팀장

문의 j.lee.reporter@gmail.com / woo@nia.or.kr

주소 대구 광역시 동구 첨단로 53(우 41068) 한국정보화진흥원

T. 053 230 1114 F. 053 230 1907 www.nia.or.kr

- 이 보고서는 방송통신발전기금으로 수행한 과학기술정보통신부 정보통신·방송연구개발사업 (ICT진흥 및 혁신기반 조성-지식정보사회의 국가발전전략연구 사업)의 결과입니다.
- 보고서 내용의 무단전재를 금하며, 가공·인용할 때는 반드시 출처를 밝혀 주시기 바랍니다.
- 이 보고서의 내용은 한국정보화진흥원(NIA)의 공식 견해와 다를 수 있습니다.