

출장보고서
2017.10.19.-10.24.
영국

개인정보 규제완화 연구 영국 사례조사

2017. 11.

1 출장 개요

- 과제명: 개인정보보호 규제에 관한 해외사례 연구
- 출장자: 이경애 전문위원, 윤혜민 연구원
- 주요출장장소: 영국 런던 및 맨체스터
 - 런던: 영국통계청, Farr Institute
 - 맨체스터: UK Anonymisation Network
- 출장기간: 2017년 10월 19일(목) ~ 2017년 10월 24일(화)

2 출장 목적

- 본 출장은 영국의 개인정보 비식별화 규제 완화 사례를 중심으로 비식별화 전문가, 통계전문가, 산업전문가 등의 방문을 통하여 데이터 기반의 신산업 창출과 관련한 정책이슈를 조사하고, 한국의 신산업 활성화를 위한 정책시사점 도출하는데 목표가 있음.
 - 데이터 추적 및 활용 시스템에 관한 해외 제도 조사
 - 기업 간 데이터 협력 방안에서의 세부 실행과제 조사
 - 비식별화 가이드라인 제공을 위한 정부 역할 조사
- 출장 결과는 ‘개인정보보호 규제에 관한 해외사례 연구’ 보고서에 정리하여 수록할 예정(2017.12)

3 출장 배경 및 필요성

- 4차산업혁명의 촉발로 인해 데이터의 보호 및 활용이라는 상반되는 목표를 조화시키는 것이 매우 중요
- 4차산업혁명이 지향하는 디지털 경제 시대는 빅데이터 처리와 관련한 법률적인 문제들을 둘러싸고 이해관계를 조정하기 위한 장기간 논의가 진행
 - 데이터 활용이 부각되는 신산업(IoT, IT보안, 빅데이터 분석 등)분야의 등장으로 데이터 생성 및 활용의 중요성이 증대함에 따라 데이터 보호·정보보안의 방법에 관한 사회적 관심이 증가
 - * 한국: 「개인정보보호법」을 기본으로 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」, 「위치정보의 보호 및 이용에 관한 법률」에 따라 정보 보호 관련 사항을 규제
 - * 영국: 법무부 산하 정보위원회에서 2012년 익명화 행동강령 (Anonymisation: managing data protection risk code of practice)을 출간하고 개인 식별 가능성을 최소화하는 방향으로 개인정보 비식별화 범위를 업계가 자율적으로 판단하도록 함
- 영국은 개인정보의 비식별화를 통해 핀테크 및 헬스케어 분야에서 신사업을 창출하고 선도적 입지를 구축
- 따라서 영국 현지에서 전문가들의 심층인터뷰를 통해 향후 우리나라 데이터 규제 및 IT 보안 관련 정책 수립에 기초자료로 활용하고자 함.
- 더불어 한국의 데이터 관련 산업의 해외 진출을 대비해 현지 시장 및 정책흐름을 조사할 필요

4 세부 조사 내용

가. Farr Institute

- Farr Institute는 21개 연구기관의 연합으로, 영국의 보건의료 빅데이터를 활용하여 국민의 건강 수준을 향상시킬 수 있는 혁신 연구를 진행함.
 - Swansea University, University of Manchester, University College London, University of Dundee가 주축이 되어 4개의 센터 (Farr Institute CIPHER, HeRC, London, and Scotland)를 운영, 각 지역 내 연구기관과 협력
- 설립 당시 영국의학연구회(Medical Research Council)이 주도하여 구성된 컨소시엄(10개 기관)에서 총 3,950만 파운드(한화 약 600억 원)을 출연
- 전자의무기록과 보건의료 행정데이터(1차의료, 2차의료, 질병레지스트리)를 연계한 ‘CALIBER’이라는 데이터 플랫폼을 구축하여 연구 목적으로 활용하고 있음.
- 빅데이터와 보건의학 두 분야에 대한 이해도가 높은 ‘보건의료 데이터 과학자’라는 새로운 커뮤니티 형성을 목표로 함.
- 보건의료와 관련한 정보는 각국이 ‘민감한 개인정보’로 구성하고 있는바, 빅데이터 활용과 연구가 연구현장에서 어떻게 발생하는지를 조사
 - 연합연구소의 빅데이터 성공사례 청취
 - 연구결과 활용 방안 및 협업 독려 방안 조사
 - 보건의료 관련 데이터의 비식별화 애로사항 조사
 - 다양한 형태의 보건데이터를 산업에 접목하는 방안 조사
 - 빅데이터 활용과 관련하여 Farr Institute의 비전, 역할, 경영방안 등

- 데이터 플랫폼(CALIBER) 구축 및 연구 사례
- 빅데이터의 보안 위험 대응 방안 조사
- 대학과 산업연구소의 협업 강화 방안 조사

나. 영국 통계청

- 영국 통계청은 표본조사에 기반한 전통적인 통계자료와 공공기관에서 작성하는 각종 행정 빅데이터를 결합하여 새로운 빅데이터를 창출하여 활용하고 데이터의 질을 제고하는 역할을 담당
- 행정데이터 연구 네트워크(ADRN)와 협력하여 행정데이터를 활용한 경제사회연구 확산에 기여
 - 영국 통계청이 보유한 데이터 사용을 위한“5 Safes”원칙 준수
 - * 허가를 받은 연구자(safe people)만이 검증된 프로젝트(safe project) 연구에 한하여 비식별화된 데이터(safe data)를 보안된 연구실(safe setting)에서만 사용할 수 있으며 결과물(safe outputs)은 통계청 직원 2명의 검증을 거친 후에 반출될 수 있음.
- 행정 및 사회과학 데이터를 활용하여 영국의 경제·사회에 대한 이해도를 높이고자 지난 3월에‘데이터과학 캠퍼스’를 개원하였고, 네덜란드 통계청 과도 양해각서를 맺는 등 공익을 위한 빅데이터 활용에 노력을 기울이고 있음.
- 데이터 전문기관으로서, 빅데이터 활용 확산을 위한 영국 통계청의 노력에 관하여 조사하고, 이를 통해 한국 빅데이터 정책에의 시사점을 확보
 - 빅데이터 활용과 관련하여 ONS의 역할과 비전 청취
 - 빅데이터 활용을 위한 데이터 전문인재 확보 방안 조사

- 공공부문의 행정데이터를 기존의 표본조사와 연계하는데 있어 영국 통계청의 중장기 전략 청취
- 빅데이터 활용을 위한 공공기관 및 민간부문과의 협업 방안 조사
- 비식별화 데이터를 안전하게 관리할 수 있는 보안 강화 방안 조사

다. UK Anonymisation Network

- 영국 익명화 네트워크(UK Anonymisation Network, UKAN)는 개인정보를 취급하고 이를 공유하기 원하는 대상에게 실용적인 조언·정보를 제공
- 2016년에 <익명화 의사결정 프레임워크 (Anonymisation Decision-Making Framework)>를 출간하여 정보위원회의 <익명화 행동강령> 보다는 실용적이고 활용이 쉬운 데이터 비식별 가이드라인을 제시
- Manchester 대학이 주축이 되어 Southampton 대학, Open Data Institute, 영국 통계청을 포함한 전문기관의 컨소시엄을 통해 운영되며 관련 기관 및 분야의 고위급 인사들이 포함된 'Core Network'와 기타 전문가들로 이루어진 'Extended Network'로 구성
- 독립적인 비식별 전문 네트워크로서 민관협력을 통해 융통성 있는 신산업 가이드라인을 제시하고 있는 사례로 주목
 - 빅데이터 활용 가이드라인 작성시의 정책 기준 청취
 - 컨소시엄 형태인 UKAN의 구체적인 운영방안(재원조달방법, 조직의 거버넌스, 사업진행 절차, 성과물 확산 등) 조사
 - UKAN의 활동 사례 청취
 - 재식별화 위험 및 해킹 위험에의 대응 방안 조사
 - 전문인재 확보 및 글로벌 이슈 대응 방안 조사

5 세부 일정 및 인터뷰 명단

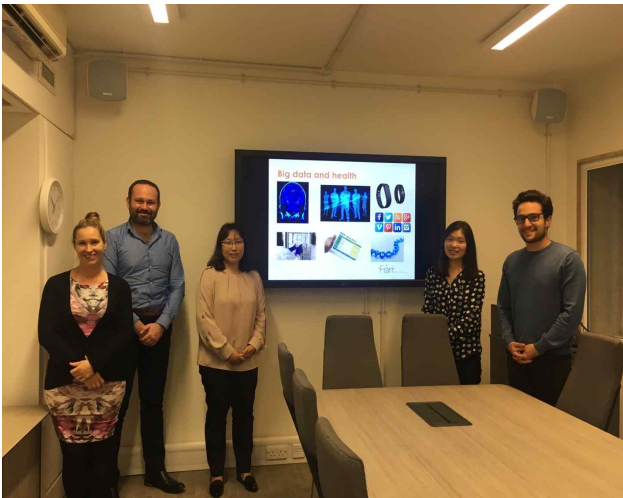
일자	시간	일정 및 주요 면담자
10.19	13:25 ~	인천공항 출발 (한국 시간)
	~17:20	런던 도착 (영국 현지시간)
10.20	11:00~14:00	Farr Institute 방문 <ul style="list-style-type: none"> - Natalie Fitzpatrick, Data Science Facilitator and Public Engagement Coordinator - Nathan Lea, Senior Research Associate - Kenan Direk, Research Associate
	16:00~18:00	영국 통계청 방문 <ul style="list-style-type: none"> - Peter Stokes, Head of Microdata Access and Exploitation - Tony Chapple, Head of Unpublished Microdata Access - Julie Mills, Senior Research Officer - Clara Boyd, Statistical Manager
10.22	15:00~18:00	이동(런던 → 맨체스터)
10.23	11:00~13:00	UKAN 전문가 인터뷰 <ul style="list-style-type: none"> - Mark Elliot, Professor of Data Science
	15:45~	맨체스터 공항 출발(런던현지시간)
10.24	~14:35	인천 공항 도착(한국 시각)

* 현지시간을 기준으로 작성

6 주요 면담 내용

가. Farr Institute

- 일시: 2017년 10월 20일(금)
- 장소: Farr Institute(영국 런던)
- 참석자: (Farr 연구소) Natalie Fitzpatrick, Nathan Lea, Kenan Direk,
(KDI) 이경애 전문위원, 윤혜민 연구원



① 영국의 헬스케어분야 빅데이터 활용

- NHS(National Healthcare Service, 영국 국민보건서비스)는 의료분야에서 정보보호와 정보공개라는 두 가지 목표를 충족시키기 위해 가이드라인¹⁾을 제정
 - (SUS(Secondary Uses Service) 구축) 의료정보의 보고와 분석을 위한 데이터 저장 및 활용 서비스로, 보건복지정보센터(HSCIC, 現 NHS Digital)가 재구성한 데이터를 수요자에게 제공

1) 건강 및 소셜 케어 데이터의 공표를 위한 익명화 표준(Anonymization Standard for Publishing Health and Social Care Data, 2013.2)

- NHS Digital은 헬스케어와 관련한 국가통계를 총괄하며 환자가 자신의 정보를 반대하지 않는 한, 환자의 정보가 NHS Digital로 업로드되는 것이 디폴트로 설정(Opt-out)
- 수집한 헬스케어 데이터는 개인식별정보가 제거된 새로운 형태로 변형되어 연구진, 의료진, 교육기관 등에서 활용이 가능
 - * 생년월일, 주소, NHS고유번호 등은 비식별처리되어 다른 빅데이터와 결합사용이 가능
 - * 제공하는 데이터의 종류: ①모두가 이용할 수 있는 익명화된 데이터, ②정보주체의 동의하에 제공되는 식별가능한 데이터, ③정보제공자의 동의는 필요 없으나 법률규제는 반드시 준수해야만 하는 데이터
 - * 상업적 목적으로 이용하는 것은 불가능

② Farr 연구소의 설립 배경 및 과정

- 의료정보학의 발전 및 다양한 유관 분야의 협업을 촉진시키고자 설립
 - 데이터 공유가 용이하도록 기반시설을 구축하고 공통의 표준모델 설정
 - 규모 있는 데이터 분석을 위한 데이터 거버넌스 확립
 - 보건의료 빅데이터를 활용하여 영국 국민의 건강 수준을 향상시킬 수 있는 혁신적인 연구 진행
- 2012년 영국 영국의학연구회(Medical Research Council)가 주도하여 구성한 컨소시엄(10개 기관)에서 총 2,000만 파운드(한화 약 300억 원)을 출연하였음.
 - 설립 당시에 출연 받은 금액 외에 더 이상 받는 지원금은 없음.

③ 조직 및 운영

- (구성) 4개의 센터(런던, 맨체스터, 스코틀랜드, 웨일스)로 구성
 - Farr Institute @ CIPHER
 - Farr Institute @ HeRC N8
 - Farr Institute @ Scotland
 - Farr Institute @ UCL Partners
- (책무성) 12명의 의료 분야 전문가가 속해 있는 국제 자문위원회 (International Advisory Board)와 1년에 1번 만나 연차보고서를 검토
 - 설립 당시 받은 지원금으로 진행하는 연구 및 활동에 대해서는 영국의 학연구회를 비롯한 컨소시엄 기관들에게 소명할 수 있어야 함.
- (향후 계획) Farr 연구소의 지원금이 올해로 종료됨에 따라 영국의학연구 회의 주도로 영국 의료데이터 연구소가 새로 조직될 예정 → 영국 의료데이터 연구소(HDR UK)
 - * 5년간 5,000만 파운드 (한화 약 700억 원)를 지원
 - * 현재 HDR UK에 소속될 파트너 기관을 모집하는 중

④ 보건의료 데이터 공유

- (기술적인 문제) 개별병원이 EMR 시스템을 별도로 도입시켰기 때문에 원천적으로 시스템 호환이 되지 않은 기술 이슈가 존재
 - 국제 표준(HL7, EN ISO, open-EHR 등)에 기반하여 영국 내에서도 기준을 정비하려는 시도가 계속 되고 있음.
- (제도적인 문제) 정부에게 데이터 활용 연구에 대한 장점을 더욱 피력하고 데이터 공유 및 활용에 대한 정책이 수립될 수 있도록 해야 함.

- 지난 2014년 콜디콧 리뷰(Caldicott Review)라는 보고서는 영국 내에서 의료데이터가 적절하게 공유되지 않고 있다는 점을 강조하며 개인정보 보호만큼 중요한 것이 적절한 정보 공유라는 것을 주장 → 법과 비슷한 효력을 지니는 원칙을 수립하여 의료데이터가 원활하게 공유될 수 있도록 할 필요성 제기

⑤ 비식별 데이터 활용

- Farr 연구소에서는 데이터를 보유하거나 관리하지 않고(법적인 권리가 없음) 분석할 뿐임.
- 데이터가 필요한 사유(연구 목적)를 명확하게 밝히고 연구가 윤리적임을 입증해야 함.
 - * 영국 NHS에서 위임한 연구윤리위원회로부터 허가를 받고 데이터 제공자(개인)으로부터 동의를 받아야 하는 경우도 있음.
 - * 각 개인에게 동의를 받는 것이 (실질적으로) 불가능한 경우에는 연구가 공익을 위한 것이라고 입증하거나 데이터를 비식별처리 필요
- 이 경우에는 데이터셋 내에 특정 정보를 활용할 수 없기 때문에 연구에 영향을 미칠 수 있음.
- 의료데이터를 활용함에 있어서 여러 이해관계자가 있음.
 - * 영국 보건부와 NHS가 데이터를 보관하는 서버를 운영
 - * 의료 전문가(GP, 의사 등)는 정보를 수집한 입장에서 데이터에 대한 책임이 있음.
 - * NHS Trust 또한 데이터에 대한 법적인 책임을 가지고 있음.
 - * 연구자는 데이터를 소유하거나 보유할 수는 없지만 법적인 절차에 따라 데이터를 제공받아 활용

- 유럽 내에서는 의료데이터 활용에 대하여 점차 관대해지고 있음.
 - 최근 독일에서 대학병원 혹은 의료연구기관 환자들은 자신의 정보가 연구 목적으로 활용될 수 있음을 주지하고 있어야 한다고 판결을 내림.
 - 영국은 법·제도적인 문제와 대중의 인식 때문에 자유롭게 의료데이터를 활용하지 못하는 문제가 있음.
 - * 연구자 입장에서는 데이터를 활용한 연구가 실보다는 득이 많다는 것이 명확할 수 있지만 대중에게 이를 설득시키는 것은 어려움.
 - * 데이터 대상자의 동의를 받지 않아도 되는 경우 엄격한 규제 적용
- 영국은 NHS Digital에서 의료데이터의 수집·보유·관리 등을 담당하고 있는데, 최근 2단계의 옵트아웃 제도를 실시
 - Type 1: 1차 의료시설 밖에서는 본인의 정보가 공유 및 활용되지 못하도록 하는 것
 - Type 2: NHS Digital이 관장하는 의료시설 및 기관 외에는 본인의 정보가 공유 및 활용되는 것을 거부하는 것
 - 옵트아웃 제도는 일반대중 및 환자들이 이해하기 쉽도록 수립되어야 하며, 옵트아웃을 실시하기 전에 필요한 정보를 정확하게 전달하여 현명한 결정을 내릴 수 있도록 해야 함.
 - * 본인에게 유용할 수 있음에도 정보를 전달받지 못해 옵트아웃 하는 경우가 생길 수 있음.
- ‘CALIBER’은 전자의무기록과 보건의료 행정데이터(1차 의료, 2차의료, 질병레지스트리, 통계청 기록 등)를 연계한 데이터 플랫폼
 - 여러 데이터를 통합하여 일괄적으로 조직·관리함으로써 추후(미래에) 연구를 반복하여도 같은 결과가 도출될 수 있도록 함.
 - 현재 CALIBER은 활용하여 40-50개의 연구가 출간되었음.

□ 데이터 보안 및 해킹 대비책

- 해킹에 대한 위험 제거 및 사이버보안을 완벽히 보장할 수는 없으나, 그 위험을 최소한으로 줄이는 것이 목표
 - * UCL의 정보보안(서비스)국 내에 Farr 연구소를 담당하는 팀이 별도로 있음.
 - * 효과적인 데이터 보안 및 해킹위험 제거를 위해서는 고위급 관리자의 책임이 수반되어야 함.
- 데이터 안전보관소(Data Safe Haven)
 - * Farr 연구소가 개발한 ISO 인증된 정보보안 시스템으로, VPN과 비슷한 형식이며 컴퓨터 내 가상서버에 접속하여 데이터와 통계 프로그램 등을 활용할 수는 있지만 데이터 반출(다운로드, 복사 등)은 일체 불가
 - * 암호화된 사이트를 통하여 데이터가 업로드 됨.
 - * UCL 내 일반 방화벽보다 더욱 보안된 방화벽을 구축함으로써 데이터 안전보관소를 해킹의 위험으로부터 보호함.

⑥ 대외협력 및 협업

□ Farr 연구소는 대외협력에서 다른 기관보다 강점을 가지고 있음.

- 비슷한 연구 및 활동을 하고 있는 기관들이 서로 소통을 하지 않는 것은 비효율적. 타 연구기관들과 협업을 통해 연구 성과를 내고 대중에게 알리는 것이 목표
- 새로 설립될 HDR UK도 이러한 맥락에서 의료정보학의 발전을 위해 동분야의 협업을 제고시키려고 하는 것

- (협력 대상) 일반대중, 환자, 의료분야 종사자, 보건의료 분야 기업(제약회사 등) 및 산업 관계자 등
 - 데이터 보유기관: Genomics England, 영국 Biobank, NHS Digital, 영국 보건부 등

나. 영국 통계청

- 일시: 2017년 10월 20일(금)
- 장소: 영국통계청 런던 사무소
- 참석자: (영국 통계청) Carolyn Watson
(KDI) 이경애 전문위원, 윤혜민 연구원



① 영국의 오픈데이터 정책과 개인정보보호 규제

- 영국은 2010년부터 내각사무국(Cabinet office)을 비롯한 범부처 전략으로 공공분야 정보를 공개하는 오픈데이터 정책에 주력
 - 영국 수상이 건강, 교육, 범죄 및 정의, 운송을 포함하여 주요 공공서비스 데이터 공개를 선언(2011.7)
 - (효율적이고 강력한 추진체계) 내각사무국이 투명성과 오픈데이터 아젠다를 직접 선도하며, 데이터의 수요와 공급기능을 분리
 - (생태계전반을 포괄) 데이터의 생산-유통-활용 및 이를 위한 기반을 포함하여 생태계 전반을 아우르는 전략
 - (주요정책) 정보자유법 제정(Freedom of Information Act, 2000), 오픈데이터 연구소 설립(Open Data Institute, 2012), 데이터역량강화전

략 발표(A strategy for UK data capability, 2013), 오픈데이터전략 발표(Open Data Strategy 2014), 디지털경제법 제정(Digital Economy Act 2017)

- 익명화행동강령(2012)과 디지털경제법(2017)은 개인정보를 비식별화하여 활용하는 제도 기반
 - 디지털경제법(2017)은 디지털기술 및 인프라와 연관된 경제활동에 관한 법으로, ‘디지털정부’ 규정을 통해 개인정보의 활용을 위한 법적 근거를 마련한 것으로 평가
 - * (디지털 정부) 디지털 및 데이터 기술 전문가를 정부가 직접 고용하여 온라인 공공서비스를 구현
 - * (오픈데이터 전략의 지속) 혁신바우처, Connected Digital Economy Catapult 등
 - * (데이터 공유) SOC와 연관된 공공기관은 다른 기관과 데이터 공유가 가능사기방지, 채권회수, 연구목적을 위한 개인정보의 교환이 가능공공기관은 비식별화된 개인정보를 활용 가능하나 재식별위험은 근절특수한 경우(범죄 또는 반사회적 행위 예방 등) 개인정보의 공개가 가능
 - 익명화 행동강령(Anonymisation: managing data protection risk code of practice)은 법무부 산하 정보위원회(Information Commissioner’s Office, ICO)에서 제정한 것으로 법적 강제력은 없으나 준수여부가 재판에서 증거로 채택 가능
 - * (개념을 정의) 개인정보, 익명화, 가명처리, 재식별화 등
 - * (구체적인 가이드라인 제시) 데이터 마스킹, 가명처리 등 개인정보 비식별화 조치 방법을 제시

* (재식별위험 대응) 데이터의 민감도에 따라 각각 공개 옵션을 고려하고 민감한 데이터 공개 시 사용제한 및 접근통제의 안전장치 마련을 권고

(1) 관리주체의 측면: 정부가 공개한 정보 vs. 연구 또는 상업목적으로 공개한 정보

(2) 정보유형의 측면: 어떤 종류의 정보는 다른 종류의 정보보다 잠재적인 공격자에게 더 큰 관심을 받게 되어 집중적인 공격의 대상이 될 수 있고 더 큰 해악의 결과를 가져올 수 있음.

(3) 잠재적 공격자에 대한 검증(motivated intruder test): 상대적으로 비전문가가 재식별에 성공할 수 있는지 판별하여 식별 위험의 최소한의 기준을 정하는 것

② 빅데이터 활용

□ 영국 통계청 내에 빅데이터를 활용하여 영국 경제, 사회를 분석하는 부서가 있음.

○ Zoopla라는 온라인 부동산 거래 웹사이트에 있는 정보를 활용하여 부동산 시세뿐만 아니라 각 지역 경제, 주거환경 등을 분석

○ 겨울철마다 영국에 유행하는 위장염이 있어 트위터 등 SNS를 분석하여 얼마나 많은 사람들이 이 위장염에 걸리는지 파악

* SNS를 데이터원으로 활용하는 것에 대한 신뢰도 문제가 있음.

○ 주로 젊은층이 SNS를 사용하므로 나이에 대한 편향(bias)이 있을 수 있음.

○ SNS에서 사용하는 언어는 일상생활에서 사용하는 언어와 다를 수 있기 때문에 단어만으로 데이터를 취사선택한다면 잘못된 정보를 분석할 수

있음.

예) 영어로 'sick'는 원래 '아픈'이라는 뜻을 가지고 있지만 은어로는 '매우 좋음'을 뜻할 수도 있음.

- 최근 개소한 데이터과학 캠퍼스(Data Science Campus)에서는 해상 운송 빅데이터를 수집
 - 항구에 드나드는 선박 수량, 물류 등 정보를 통하여 무역 흐름을 파악할 수 있음.
- 국가 통계에 빅데이터를 포함하여 활용하기까지는 아직 그에 대한 이해도가 부족하지만, 현재로는 아주 좋은 참고 자료가 됨.

③ 행정데이터(개인정보) 익명화 방안

- 개인정보는 많은 데이터셋(데이터베이스)을 활용하는 것이 활용자의 입장에서 유리하지만 재식별의 위험이 커짐.
 - 각 데이터베이스마다 개인식별자(personal identifier)를 암호화(one-way encryption)하여 암호화된 식별자로 다른 데이터베이스에 있는 정보와 매칭한 후 사용
 - * 이 방법의 한 가지 문제점은 개인식별자에 오류가 있을 때(이름을 잘못 기입하거나 주민등록번호를 잘못 적는 등) 이미 암호화한 후에는 그 오류를 알아낼 수 없다는 점
 - 이 문제점을 극복하기 위해 개인식별자를 암호화하는 대신 통계청이 생성한 새로운 식별자를 부여하여 여러 데이터베이스를 연계하여 활용할 수 있음.
 - * 이 또한 간접식별자(indirect identifier)는 드러나기 때문에 재식별 위험이 없지는 않음.

* 간접식별자는 더 큰 단위의 데이터로 대체

예) 우편번호 대신 지역(동, 구)데이터를 활용

- 최소 250명은 속해 있는 단위의 데이터가 기준
- 희귀 질병 등 개인을 식별할 수 있는 가능성이 있다면 (지역) 단위를 높여야 함.

* 많은 내용의 세부 데이터를 원하면 지역이 커질 수밖에 없으며, 지역을 좁혀야 한다면 세부 데이터를 잃게 됨.

④ 정부부처 간 협업

- 영국은 잉글랜드, 웨일즈, 스코트랜드, 북아일랜드가 각각 소속 지역의 행정데이터에 대한 권한을 가지고 있어서 데이터를 공유하는 것이 까다로웠으며, 부처마다 공유에 대한 제도가 다르기도 했음.
- 최근 제정된 디지털경제법(Digital Economy Act 2017)으로 인해 정부부처 간 데이터 공유가 용이해짐.
 - 이로써 통계청이 원하는 데이터를 어느 부서에나 수집해올 수 있는 법적 근거는 있지만, 내년까지는 타 부처가 데이터를 제공해야 하는 의무는 없음.
 - 정부부처 간 데이터를 제공할 때에는 데이터보호법에 의거하여 타 부처에서 필요한(요구한) 데이터만 제공 가능하며 데이터를 제공받는 부처에서 올바르게 관리할 수 있도록 보장해야 함. 데이터에 대한 책임은 데이터를 제공하는 부처에 있음.

⑤ 데이터 공유 네트워크

- 2011년에 영국 내에서는 최초로 행정데이터를 결합하여 활용할 수 있는 보안 네트워크를 구축하여 현재까지 통계청에서 사용 중
 - 네트워크를 통해 데이터를 활용하여 얻을 수 있는 이익을 증명하고 데이터 보안에 대한 우려를 잠식시켰음.
- 지난 6년간의 경험과 신기술을 기반으로 보안을 최우선적으로 고려하여 데이터 공유를 위한 새로운 클라우드 네트워크를 개발 중에 있음.
 - 데이터 활용자에게 제한된 데이터를 공유할지, 많은 데이터를 공유한 후에 데이터 활용자에 대한 교육 및 보안을 엄격하게 할지는 아직 고민 중

⑥ 인구조사와 설문을 대체하는 행정데이터 활용 방안

- 최근 많은 국가들은 설문조사 응답률이 떨어지고 설문을 시행하는데 많은 비용과 시간이 드는 것에 대한 대안을 찾고 있음.
 - 영국 통계청은 이미 존재하는 행정데이터를 활용하여 필요한 정보에 대한 질문은 설문 문항에서 제외하고 조금 더 짧은 설문지를 제작함으로써 온라인으로 설문을 배포하고 응답률을 높이고자 함.
 - * 데이터 특성에 따라 설문에 반드시 포함하여야 하는 것도 있음.
 - 예) 자신의 민족성에 대한 사람들의 인식은 시간이 흐르면서 변할 수 있기 때문에 예전 인구조사나 행정데이터에서 얻는 정보가 틀릴 수 있음.
 - 북유럽 국가들이 관련 작업을 선도적으로 시행하고 있음.
 - * 설문조사 전에 행정데이터로 답변을 미리 채운 후에 설문 대상자에게 답변이 여전히 유효한 지 확인을 받음.

⑦ 민간부문 협업

- 사우스햄튼 대학교, 카디프 대학교 등 대학 및 연구기관과 협업하고 있음
 - 가상현실 마이크로데이터 랩을 통한 교수, 박사 등 연구자 지원
- 사기업과 직접적인 계약을 맺지는 않지만 기업이 보유한 데이터를 공유 받는 경우가 많음.
 - 금융 분야 데이터를 분석하여 영국 경제를 파악하고자 함.

⑧ 전문인력 확보

- 정부통계서비스(Government Statistical Service)는 통계 분야에 종사하는 영국 공무원의 커뮤니티
 - 타 부처에서 근무하고 있는 전문인력이 통계청으로 이직하거나 통계청에서 직접 사람을 고용하기도 함.
- 데이터과학캠퍼스에 있는 데이터마이닝 전문가 활용
- 통계청과 협력하고 있는 사우스햄튼 대학교, 카디프 대학교 등의 통계학과 졸업생들을 통계청으로 스카우트하기도 함.
- 통계청 내에 있는 전문인력의 역량강화를 위한 방안
 - 직원들이 한 부서에 있기 보다는 여러 부서를 경험하고 다양한 분야의 경력을 쌓도록 함.
 - 교육 및 훈련 프로그램 제공
 - 사우스햄튼 대학과 공동으로 석사 프로그램 운영

⑨ 데이터 보안 및 해킹 대응 방안

- 국가 사이버보안센터(NCSC)에서 각 정부부처가 안전하게 데이터를 보유할 수 있도록 조언
 - 통계청과 NCSC는 긴밀하게 협력하며 안전장치 마련을 위해 노력
- 통계청 내부에도 데이터 보안 관련 전문가가 있으며 데이터 활용과 안전의 균형을 유지하는 것을 중요하게 생각함.
- 정보 공개 및 활용에 대한 대중의 인식
 - 정보 공개 및 데이터 보안 인식은 정보의 특성과 개인마다 차이가 있음.
 - * 연령이 젊을수록 개인정보에 대한 공유가 더 자유롭고 편할 수 있음.
 - * 정보의 종류(범죄기록 vs. 애완동물 보유 여부)에 따라 정보 공개에 대한 선호도가 달라짐.
 - 통계청에서는 개인정보보호와 관련된 이익집단, 기관 등과 항상 소통하며 행정데이터를 활용하는 것은 통계 분석을 위한 자료 활용이 목적이며 개인정보 보호에 힘쓰고 있음을 전달
 - 데이터를 활용함으로써 사회에 주는 이익을 강조
 - * 통계청은 ‘문제가정 프로그램’²⁾ 수혜 가정을 대상으로 데이터를 수집하여 프로그램 전후를 비교한 후 어떠한 정책·제도가 이 가족들에게 가장 많은 영향을 주었는지에 대한 평가를 진행하였는데, 통계를 활용하여 국민 세금으로 시행한 프로그램의 효과성을 증명할 수 있기 때문에 국민들에게 데이터 활용에 대한 긍정적인 생각을 심어줄 수 있음.

2) 영국 정부에서 진행하는 프로그램으로 복지수당을 받거나 영국 보건부, 사법부 등 특정 정부부처와의 교류가 많은 가족들을 대상으로 자립할 수 있도록 지원

⑩ 실시간 데이터 활용 여부

- 데이터의 목적과 관리하는 방법에 따라 실시간 데이터를 수집하고 활용함의 여부가 달라짐.
 - 정보의 트렌드가 필요하다면 매달, 매주, 혹은 매일 동일한 데이터를 수집하는 것이 유용할 수 있지만 변화가 거의 없는 정보라면 무의미함.
 - 또한, 실시간으로 모든 데이터를 송부 받게 된다면 그 중에서 취사선택을 하고 통계분석이 가능한 형식으로 구성하는 것도 어려울 가능성이 있음.

다. 영국 익명화 네트워크(UKAN)

- 일시: 2017년 10월 23일(월)
- 장소: 111 Piccadilly(영국 맨체스터)
- 참석자: (맨체스터 대학교) Mark Elliot 교수, Elaine Mackey 박사
(KDI) 이경애 전문위원, 윤혜민 연구원



① 영국의 익명화 가이드라인

- 법무후 산하 정보위원회(Information Commissioner's Office, ICO)에서 2012년 익명화 행동강령(Anonymisation: managing data protection risk code of practice)을 출간
- 개념 정의
 - 개인정보: 살아있는 개인을 식별할 수 있는 정보 혹은 그 개인을 지칭하거나 개인에 관한 의견을 포함한 정보로 특정 개인에 대해 다른 제3자 또는 데이터 처리자가 특정인을 지칭하는 정보를 포함
 - 익명화된 데이터: 다른 데이터와 조합하여도 개인을 식별할 수 없는 데이터로 정의
 - 가명처리: 현실에서 쓰이지 않는 독특한 식별자를 이용하여 데이터 세트

에서 개인을 구분하는 것

- 재식별화(비익명화): 비식별 처리된 데이터를 다른 데이터와 조합하여 개인을 식별해내는 과정
- ICO의 강령은 법적 강제력이 없으며, 개인 식별 가능성을 최소화하는 방향으로 개인정보 비식별화 범위를 업계가 자율적으로 판단하도록 함.
- 「ICO 익명화 행동 강령」에 따른 데이터 분류
 - 녹색 데이터: 집단의 평균값과 같은 비식별 처리된 정보로 대중에게 공개 가능
 - 적색 데이터: 식별 가능한 개인정보로 위급한 보건 상황이 발생했을 경우에만 활용
 - 황색 데이터: 가명처리된 개인정보를 NHS 보유 정보 및 보건기록과 결합하여 연구소, 학교, 민간기업 등에게 제공(유료)\
- 익명화에 대한 우수 사례를 구현하는 것을 목표로 하여 개인정보를 취급하고 이를 공유하기 원하는 대상에게 실용적인 조언과 정보를 제공하기 위한 협업체를 구성
 - 영국 익명화 네트워크(UK Anonymisation Network, UKAN): 2012년 설립되어 2년 동안 ICO의 재정 지원을 받았으며, Manchester 대학, Southampton 대학, Open Data Institute, 영국 통계청을 포함한 컨소시엄으로 구성
 - Core Network: 소수의 숙련된 전문가들이 익명화 데이터 관련된 문제에 대하여 논의
 - Extended Network가 있는데, 전자는 하며, 후자는 Core Network와 소통하며 관련 사례나 논문을 제출하고, UKAN 주최 워크숍에 참여

② 설립배경 및 과정

- 2012년 영국 정보감독청(ICO)의 주도 하에 설립되었으며 2014년까지 ICO의 재정지원을 받았음.
- ICO가 2012년 ‘익명화 행동강령(Anonymisation Code of Practice)’을 출간하며 이를 기반으로 실용적인 조언을 제공해줄 주체가 필요하였음.
 - 행동강령은 정부에서 권고하는 비식별 관련 규제 및 정책을 정리해놓은 책으로, 이 행동강령만으로는 익명화에 대한 해답을 찾기 어려움.

③ 조직 및 운영

- UKAN은 크게 네 가지의 단계로 조직이 구성되어 있음.
 - (1) 본부(hub): 맨체스터 대학교: Elliot 교수와 Mackey 박사가 전체적인 운영 및 대외 협력을 담당
 - (2) 컨소시엄: 맨체스터 대학교, 사우스햄튼 대학교, 영국 통계청, 오픈데이터 연구소
 - 4개 기관은 서로 협의하는 관계(consultation-based relationship)
 - ICO와는 정기적으로 만나며 긴밀하게 협의(UKAN은 ICO로부터 독립적인 기관)
 - (3) 핵심 네트워크
 - 30명의 다양한 공공 및 민간 분야(법률, 통계, 컴퓨터 등)의 고위급 전문가로 구성되어 있으며 UKAN의 전략을 수립하는 데 큰 역할을 함.
 - 익명화된 데이터는 활용도가 많지만 법률이나 규제 해석에 있어서 모호한 부분도 많기 때문에 다양한 분야의 전문가들을 모으는 것이 UKAN 설립 초창기에 중요한 작업이었음.

- 18개월 간 여러 차례의 워크숍을 통해 ‘익명화란 무엇인가?’라는 질문에 대한 해답을 찾으려고 하였으며, 그 결과 ‘익명화 의사결정 프레임워크(Anonymisation Decision-Making Framework)’를 2016년에 출간
 - * 이 프레임워크는 ‘익명화 행동강령’과 달리 데이터 익명화를 진행하는 단계를 상세히 명시하고 있으며 내용이 일관되도록 ICO에서 상시적으로 피드백을 받고 있음.
 - * 의사결정 프레임워크는 새로운 시도로, 익명화의 여러 분야(법률, 통계, 컴퓨터과학 등) 전문가들의 의견을 수렴하여 발간

(4) 확장 네트워크

- 핵심 네트워크에는 포함되지 않는 기타 전문가들로 이루어져 있으며, 핵심 네트워크와 교류하며 UKAN의 워크숍 및 활동에 참여
- 500여명 규모로 핵심 네트워크는 영국인들로만 구성되어 있는 반면에 확장 네트워크는 유럽 전역 및 북미 지역도 포함

□ ICO와의 관계

- 익명화에 관한 정책 사항을 업데이트 받고 관련 법을 해석함에 있어서 의견을 나눔.
 - * ICO의 행동강령과 UKAN의 프레임워크는 포괄적인 가이드라인이기 때문에 구체적인 사례를 논의함으로써 정책에 대한 해석을 명확히 하는 데 도움을 주고 제도를 정립할 수 있음.
 - * 디지털경제법(Digital Economy Act 2017), 유럽연합의 GDPR과 같이 새로운 정책에 대한 내용과 이로 인해 변경되는 사항들을 전달 받음.
- ICO에서는 익명화 이슈에 대한 문의를 받거나 관련 민원 등 사건이 발생하면 UKAN에게 상담 받을 것을 권고

- * 이 권고를 따르는 것이 의무는 아니며, 익명화 데이터를 활용하고자 하는 개인 및 단체가 UKAN의 조언을 받아야 하는 것이 필수 사항은 아니지만, ICO로부터 UKAN에 문의할 것을 권고 받고 따르지 않은 후 데이터 관련 사고나 법적인 문제가 발생하면 처벌받는 것을 피할 수 없음. 반면에, 상담을 받은 경우에는 처벌이 감해지거나 면제받을 수도 있음.

□ 재정운영

- 영국 국무조정실, 통계청 등으로부터 출연금을 받고 있으며, 기업들과는 계약(프로젝트) 단위로 일정 금액을 받고 프로젝트를 수행
- UKAN의 주 업무인 클리닉 상담이 무료임을 감안하여 다른 자금이나 재정 모델을 고려하고 있음.
 - * 투자자 모델은 UKAN의 독립성에 지장을 줄 수 있음.

□ 운영상의 어려움

- UKAN을 국가 서비스의 하나로 운영할 수 있도록 규모를 키우는 것
 - * 여전히 많은 사람들이 익명화 데이터를 활용하거나 관리하며 많은 실수를 하고 법을 위반하는데, 이는 UKAN이 그만큼 영향력을 행사하지 못했다는 것을 증명
 - * 현재 홍보나 마케팅 등의 활동은 하지 않음.
 - * 재정적으로 규모를 키우는 것 또한 하나의 숙제
- 익명화 관련하여 전문가나 인식이 많이 부족함.
 - * 통계 등 기술을 보유한 전문인력은 영국 통계청을 비롯하여 많은 편이지만, 익명화와 관련된 여러 분야(법률, 통계)를 두루 아는 인력이 부족함.
 - * 익명화에 대한 조언과 지도를 부탁할 전문가를 이전에 섭외하려고

시장 조사를 시행하였는데 영국 전역에 적합한 전문가는 10명뿐이
었음.

* 사람들은 이 분야에 관해 무엇을 모르는지 모르는 상황. 오히려 클
리닉을 찾아오는 사람들은 주도적이고 이 분야에 관심이 있는 사람

* 지식이 부족하기 때문에 발생할 수 있는 두 가지 상황

(1) 법·규제 등을 무시하고 멋대로 데이터를 활용하고 관리함.

(2) 위험을 회피하고자 데이터 공유 및 활용을 하지 않음.

④ 컨설팅 업무

□ 익명화 데이터를 활용하고자 하는 개인, 기업, 단체 등에게 컨설팅을 제공
하는 것이 UKAN의 주요 업무 중 하나.

□ 클리닉(clinic)은 1시간 내외의 1:1 무료 상담 세션으로 익명화에 대한 일
반적인 조언 혹은 상담 요청자가 원하는 구체적인 사례에 대해 논의

○ 내용과 특성에 따라 추후 클리닉이 연속적으로 이루어질 수도 있고,
ICO와 같은 다른 기관들에게 문의할 것을 권유하기도 함.

○ 일반적으로 상담 문의자들은 익명화에 대한 개념이나 그 상황이 익숙지
않아 좀 더 명확한 가이드라인을 받기를 원하기 때문에, 대부분의 클리
닉은 UKAN의 프레임워크에 대해 설명을 해주는 것으로 시작함.

* 70% 이상은 상담을 받은 후에 본인이 다음 단계를 어떻게 진행해야
할지 깨닫고 추가적인 클리닉을 요청하지 않음.

* 상황이 복잡하거나 UKAN 내부적으로도 흥미로운 사례라고 여기면
추가적으로 상담을 진행하기도 함.

○ 데이터 활용자가 UKAN의 컨설팅을 받는 것이 의무는 아니지만 ‘모범

사례(best practice)'를 공유받기 위해 UKAN을 찾고 있음.

- 기업이나 기관이 익명화된 데이터를 활용하고자 하는 경우 UKAN과 계약을 맺어 큰 규모의 프로젝트를 진행하는 경우도 있음.
 - 최근에 한 제약회사가 임상실험에 참여한 환자 정보 익명화 방안 및 보안 유지에 대한 작업을 의뢰하여 UKAN에서 수행
 - 프로젝트는 그 규모에 따라 전문가를 초빙하기도 함.

⑤ 익명화 데이터 활용

- 보건의료 분야는 여러 관계자 간에 이해가 상충되는 부분이 있어 까다로움. 제도적으로 정보를 공유해야 하는 의무도 있지만(NHS Digital), 정보를 공유하지 못하는 제약도 있음(GP, 의사 등 의료 종사자).
 - 데이터를 공유함에 있어서 어느 정도까지 정보를 공개해야 하는지 결정하는 것이 어려운데, 그 정보를 공유하는 대상과 사유를 정확히 아는 것이 필요함.
- 사례: A 제약회사가 임상 실험 데이터를 비식별화하여 공유하려고 했을 때 정보는 병원 차원에서 수집되었고, 이는 유럽 여러 국가의 대학연구소, 제약회사 등에게 공유될 예정이었음. 이 정보를 바탕으로 데이터베이스를 구축하는 것이 목표였는데 크게 두 가지 문제점이 있었음.
 - (1) 여러 기관이 데이터베이스를 관리하고 활용함에 따라 누가 데이터 관리자이고 데이터 활용자인지 명확하지 않아 데이터 거버넌스에 문제가 생김.
 - (2) 데이터 제공자인 환자들이 그들의 정보가 어디에, 어떻게 쓰이는지 정확하게 숙지하고 동의했는지 여부
 - 데이터는 컨소시엄 내에서 공유되고 데이터가 익명화된다고(anonymised)

고 설명했지만, 실은 비식별처리(de-identified)하는 것이었으며 영국 내에서 비식별 데이터는 개인정보로 분류됨.

- 정부차원에서 보건의료 데이터 공유를 독려하고 의무화하는 법·정책 등이 최근에 많이 늘어남(예: Caldicott Guardian, 보건복지법(Health and Social Care Act 2012) 등)

* 이러한 정책이 모든 기관에게 해당되는 것은 아님. 상업 목적이 강한 보험회사, 마케팅회사 등에게는 공유를 금지

□ 금융 분야에서는 정보를 공유하는 것보다 재활용에 대한 문제가 있음.

- 고객 정보를 수집하여 원래 목적이 아닌 내부적인 연구에 활용하면 비식별처리 된 정보는 재식별 가능성이 있기 때문에 이에 대한 우려
- 영국은 모든 금융거래가 신용평가회사에 기록되며 영국중앙은행도 모든 금융회사의 개인대출 및 담보 이력 등을 보유하고 있음.
- 예전과 비교하면 제도적으로 데이터 공유가 의무화되어 있음.

⑥ 비식별화/익명화 관련 인식 제고

□ 2012년에 행정데이터연구네트워크(ADRN)에서 익명화 데이터 활용 연구에 대한 워크숍을 진행

- 이틀 간 20-25명으로 구성된 소규모 그룹을 대상으로 12차례 진행
 - * ‘연구란 무엇인가’라는 기초적인 질문부터 시작해 익명화에 대한 정의, 데이터 보안 및 데이터 남용·악용 방지 등에 대한 내용을 전달
- 교육 전에는 90% 이상이 데이터 공유에 대해 반대했지만, 교육 후에는 90% 이상이 데이터 공유에 대해 긍정적으로 생각하게 됨.
- 워크숍 이후에 행정데이터를 공유하고 활용함에 있어서 두 가지의 원칙

을 수립

- * 영리기업에게는 공유하지 않도록 함.
- * 데이터는 연구 종료 후에 파기해야 함.

□ 일반 대중의 인식 제고

- 데이터를 활용한 연구가 투명하고 공익을 위한 것임을 강조해야 하며, 그 결과가 연구 성과 등으로 나타나야 함.
- 교육제도에 관련 내용을 포함하는 것 외에는 효과적인 방안을 찾기 어려우며, 이는 장기적인 방편
- 영국은 15년 전 정부가 투명성 제고를 위해 많은 노력을 기울였고, 이는 정부의 개인정보 및 데이터 활용에 대한 신뢰도 함께 개선시킴.

⑦ 데이터 보안 및 해킹

□ 클리닉 세션이나 구체적인 사례 상담을 요청 받았을 때 재식별에 대한 위험을 최우선적으로 고려함.

- 영국 내 원칙은 재식별에 대한 위험이 '무시할 수 있는(negligible)' 정도이면 데이터 활용이 가능함. 위험을 아예 없애는 것은 불가능에 가까움.

□ 위험도 판별에 대한 기준(risk standard)

- 데이터 보안 사례도 여러 가지가 있기 때문에 전례가 있거나 그 문제가 흔한 경우에는 위험도 판별이 용이함.
- 그 외에는 영국 인구조사 익명화 기록 샘플(Samples of Anonymised Records, SAR)³⁾을 활용

3) 익명화기법이 적용된 인구조사 마이크로데이터

- * 인구조사는 수차례에 걸쳐 반복되는 작업이며 인구조사가 시행되기 전에 몇 년 동안 기초작업이 수반되기 때문에 위험도의 기준으로 삼기에 최적인 데이터
- * 비교 데이터 상황 분석(comparative data situation analysis)을 통하여 영국 인구조사 SAR과 사례자의 데이터를 비교함.
- 마지막 단계에서는 모의침투테스트(Penetration Test)를 진행
 - * 비용이 많이 소요되기 때문에 항상 모의침투테스트를 수행하지는 않지만 선례도 없고 비교가 불가능하면 데이터 공격(해킹)을 시뮬레이션하여 데이터의 취약점 및 보안 정도를 판단

⑧ 국제협력

- 유럽, 미주 등에 국제적인 네트워크를 보유하고 있음.
 - 매년 익명화, 비식별화 데이터 관련 국제회의에 참여
 - 전 세계적으로 전문가들은 200여명
- 호주 정보감독청(OAIC)과의 협력을 통하여 호주에서 적용될 수 있는 ‘비식별화 의사결정 프레임워크’이 지난 9월에 출간되었음.
 - 법률적인 부분이 많이 달라 호주 전문가들이 본문을 재작성하고 수정하는 작업을 거쳤지만 기본적인 익명화 원칙은 달라지지 않음.
 - 언어가 같아 협력을 하는데 용이
 - 캐나다와도 비슷한 작업을 수행할 계획
- 많은 국가들(터키, 포르투갈 등)에서 협력 요청이 있으나 언어가 걸림돌