

## 미국의 행정자료 활용 연구 경험과 시사점

### □ 출장 목적

- 미국에서 정책 등 각종 연구에서 행정자료를 활용한 경험과 관련 이슈에 대하여 논의, 시사점을 탐색

### □ 과제명

- 사회보장제도의 소득 서비스 보장 현황과 정책과제 연구

### □ 출장기간

- 2018.05.13.(일)~2015.05.19.(토)

### □ 출장지역

- 미국(시카고, 뉴욕)

### □ 출장자

- 이현주 연구위원

### □ 면담자와 관련 자료

○ 면담자: 시카고 대학에 설치된 아동보호 전문연구기관의 연구자. 행정자료 활용의 오랜 경험을 가진 연구자로 행정자료의 활용에 대한 연구물을 발간하기도 함.

- 다양한 원천의 행정자료를 연결하여 연구에 활용. 예를 들어 위탁아동의 생활을 분석하기 위하여 사회보장행정자료, 경찰, 학교, 상담 자료를 입수하여 연결한 후 분석에 활용

○ 면담자: 시카고 대학의 NORC(National Opinion Research Center: NORC),의 노인의 사회적 생활과 건강 연구(National Social Life, Health, and Aging Project: NSHAP) 관계자

- 미국보건연구소(National Institutes of Health: NIH)로부터 많은 연구를 수주 받은 바 있음.
- 관련 자료: National, Social Life, Health and Aging Project (NSHAP). 대학을 중심으로 한 행정자료와 서베이자료의 연결 사례
  - NSHAP는 조사된 자료로 구성되는데 표본추출은 보건과 퇴직조사(Health and Retirement Study: HRS)를 기반으로 하여 이루어짐.

- HRS는 1992년 이후 2년마다 미국의 약 23000가구를 대상으로 하여 실시되는 조사. HRS는 개인수준에서 사회보장과 보건, 군퇴직자(Veteran) 자료, 사망자료, 고용주 제공의 연금자료 등 행정기록과 연결되기도 함.
- HSHAP 자료는 현재까지 세 차례의 조사로 구성된 패널 자료로 세 개의 조사로 구성됨. 개인단위의 개별적 질의응답/생체적인 자료/보조적인 자기응답의 질의응답
- NSHAP 자료는 National Archive of Computerized Data on Aging(NACDA)의 소유, ICPSR(Inter-university Consortium for Political and Social Research)을 통하여 접근, 사용이 가능

○ 면담자: 시카고 대학의 사회학과 교수로 다년간 세대 간 사회이동, 불평등을 연구한 연구자

- 종단연구를 위하여 센서스 자료 등 과거의 가족 계보를 추적할 수 있는 자료를 구성
- 센서스 자료 연결을 위한 기술과 문제 등이 해당 연구자의 전문영역임.

○ 면담자: 콜롬비아 대학부설 인구조사센터(Population Research Center)의 연구자

- 아동에 대한 연구에서 행정자료를 활용. RDC(Federal Statistical Research Data Centers)의 자료 활용

## □ 주요 논의 내용

○ 자료의 공개 수준

- 공개수준으로 공개자료(public data)와 공개제한자료(Restricted data)로 구분가능
- 연구자들은 접근이 쉬운 공공자료를 주로 사용하는데 예를 들어 센서스 자료도 공개되고 있는 1%자료를 주로 사용함.
- 연구를 위하여 행정자료와 연결된 자료도 비식별조치를 한 상태로 공개 자료가 되기도 함.
- 공개제한자료는 접근 가능 연구자의 자격, 해당 자료에 대한 사용 상 동의, 신청과 승인과정 등 까다로운 조건과 절차를 두고 있음.
- RCD 등의 공개 제한 정보의 경우에는 보안이 매우 강하여 결과표는 연구자에게 상당기간이 걸려 전달되고 모든 표는 검토되며 연구단계에서도 이러한 점이 고려됨. 물리적으로 분석 공간에 카메라가 설치되어 있고 휴대폰등도 주의가 두어지고 관리됨.

- 연구에 행정자료를 사용함에 있어 미국의 특징은 다양성과 절차의 공식화임.
- 다양한 관리주체, 자료의 생산과 구성이 존재하며, 절차는 공식화되어 연구자가 연구의 질과 해당 자료의 필요성을 입증하면 자료에 대한 접근성이 공개적으로 보장될 수 있다는 것이 특징
- 이 과정에서 연구경험이 짧은 연구자라 하여도 자료에 대한 접근 기회를 가질 수 있는 장점이 있으며 자료에 대한 비공식 접근보다 시간을 더 소요될 수 있으나 자료접근의 형평성을 일정 수준 확보할 수 있다는 장점을 지님.

#### ○ 자료의 관리주체들

- 자료의 관리주체는 자료마다 매우 상이
  - 예를 들어 아동관련 행정자료 연계자료는 주로 South California 대학에 두는 방식이 이러한 사례
- 주로 아래의 세 가지 자료 제공처가 중심적 역할을 수행
  - RDC: 미국 통계청 산하 연구자료센터
  - ICPSR(Inter-university Consortium for Political and Social Research): 대학 차원의 협의 조직
  - IPUMS(Integrated Public Use Microdata Series): 공개가 가능한 인구자료로 일부 표본으로 제한된 자료 허브
- RCD 사무소는 미국 전역에 일정수로 배치되어 있으나 어디서나 접근이 쉬운 정도로 많은 것은 아님. 주로 대학에 설치되어 있으나 대학의 위치 등을 고려 특정 지역(예 시카고)은 은행과 같은 조직에 설치되기도 함
  - RDC 에는 자료 전문가를 두고 있음. 이들은 자료의 연결, 자료 설명, 연구의 지원, 자료의 활용에 대한 모니터링 등을 담당
- ICPSR 대학의 협의조직으로 각 대학의 자료 접근성 증개.
  - 해당 조직은 미시간 대학에 설치되어 있음. 해당 조직은 정부조직 등 여러 조직으로부터 스폰서를 받는 조직이기도 함.

#### ○ 자료에 대한 신청과 승인과정

- 위의 세 가지 자료 접근 허브는 자료를 활용하는 절차 등이 모아진 허브로 자료 소유자는 이 조직들과 상이. 자료에 대한 접근을 위해서는 자료소유자의 승인이 필요

- 해당 자료의 소유는 자료 원 소유조직이 갖고 있어 해당 자료에 대한 접근은 이 조직의 허가를 필요로 함.
- 자료 소유자는 대체로 자체적인 협의체를 구성하여 전문가들이 지원한 연구에 대한 전문적 판단을 할 뿐 아니라 자료의 질, 관리 등에 대해서도 논의하도록 함.
  - 예를 들어 CDC의 자료는 CDC가 전문가 6인으로 협의조직을 구성, 운영함.
- 민감 정보를 가진 비공개 자료(국세 조사국 (IRS), 국세청 (IRS), 연방 세금 정보 (FTI)등이 이러한 대표 사례)를 활용하는 연구는 자료 활용 신청을 하여야 함. 신청과 승인절차는 아래와 같음.
  - 일정한 자격을 갖춘 연구자가 연구제안서를 제출-협의조직의 검토-수정과 보완-승인-자료분석
  - 자료를 활용하고자 하는 연구자는 제안서 양식에 따라 제안서를 작성
    - 해당지역의 RDC 관리자와 사전에 비용, 제안서 내용구성, 형식 등에 대하여 상담하기를 권고
    - 이 담당자는 자료활용이 승인될 때까지 연구자를 지원함
  - 제안서는 아래와 같은 기준으로 검토됨.
    - 인구 조사국 프로그램에 이익을 제공하는가
    - 과학적 가치를 입증하는 것인가
    - 비공개 데이터를 필요로 하는 연구인가
    - 데이터가 주어진다면 실현 가능한 연구인가
    - 공개 위험이 없는가
  - 국세청자료 활용을 위한 신청과 승인의 특징
    - 국세청의 자료는 매우 제한된 접근성을 유지하고 있음.
    - 공식적으로 국세청의 자료는 경쟁을 거쳐 매우 우수한 연구계획에 한정하여 자료 접근 기회를 제공
    - 실제 이 자료를 활용하여 분석은 매우 희소하고 해당 영역에서 매우 영향력이 있는 연구자의 연구로 한정되는 경향
  - 자료 소유주체로부터 지원을 받는 연구의 경우에는 자료의 활용까지의 단계가 빠르게 진행되는 경향이 있음. 보통 자료의 활용까지 1년이 소요되지만 이 경우는 6-9개월 정도로 그 소요시간이 줄어들기도 함.

## ○ 자료의 연결

- 각기 다른 자료의 연결은 다양한 방식으로 진행됨. 예를 들어 아래의 두 가지 방식이 가능
  - 연구자가 자료를 받아 연결이 가능하도록 가능한 수준의 정보를 활용하여 자료 연결을 시도. 주로 성, 연령 등 다수 주요 변수들을 기준으로 연결.
  - 다른 방식은 자료를 소유한 조직이 연구자의 신청에 따라 자료를 연결하여 제공하는 방식. 예를 들어 RDC는 연구자의 전문성을 기초로 자료 연결 방식을 결정한 후 연구자의 자문에 따라 RDC가 자료 연결을 하고 연구자에게 제공하기도 함.
- 사회적 이동의 연구에서 소득이 아니라 직업을 이동을 판단하는 변수로 활용하였음. 이러한 변수선택의 이유 중 하나는 소유가 다른 조직의 자료를 연결하여 동 자료를 활용하는 연구의 진행이 어렵기 때문이기도 함을 지적

## ○ 분석 방식

- 가상 자료 보호(virtual data Enclave)를 주로 사용하여 자료 분석
  - 대학의 경우, 특히 경제적으로 RDC와 컨소시엄이 가능한 대학은 대학에서 이를 활용하여 자료 분석이 가능
  - 자료에 접근하는 기제를 확보하려면 높은 비용을 지불하여야 함. 이러한 비용은 제공자의 관리 비용에서 비롯되는 것
- RDC를 방문, 자료 분석
  - 수주 받아 진행하는 연구의 연구자도 센서스 자료에 대한 분석은 RDC에 가서 분석
  - 분석을 위하여 방문하는 곳은 주로 대학에 설치된 RDC 지소이지만 경우에 따라서는 대학이 아니라 다른 공간을 활용. 시카고의 경우에는 통계청 자료의 주요 사용자인 연구자가 속한 대학이 공간적으로 분산되어 접근성을 고려한 지역의 Federal Reserve Bank에 지소를 두고 있음.
  - SED Security Enclavement data center에서 분석결과에 민감 정보가 포함되어 식별이 가능한지 등을 판단. 전문가 배치하여 역할을 하도록 함.