

국외출장 보고서
(미국 및 캐나다,
'16.8.31.~9.9.)

2016년 통계작성기관 해외연수 결과보고서

2016. 9.



통 계 청
통계작성기관 연수단

차 례

I. 연수 개요	1
II. 연수 내용	
1. MPC_IPUMS 방문	4
2. ICPSR 방문	17
3. 미시건대학교 RDC	29
4. 캐나다 통계청 및 RDC 방문	43
III. 주요 정책적 시사점	57
[붙임] Meeting Agenda 및 사전질의서 등	61

1. 연수 개요

1. 목 적

- 통계작성기관 간의 업무협력 관계 강화
- 선진외국의 통계제도와 작성실태, 최신동향 등 비교·분석을 통해 실무에 반영 및 외국의 다양한 체험을 통해서 통계 실무의 업무행태 변화 유도

2. 세부 추진내역

- 연수기간: 8. 31. ~ 9. 9.(8박 10일)
- 연수내용
 - 마이크로데이터 통합관리·서비스와 통계자료 공유시스템 관련한 다양한 운영 실태를 파악
 - ◆ 마이크로데이터 서비스 영역과 운영조직 파악
 - ◆ 제공자료의 종류 및 서비스 방식* 파악
 - ◆ 서비스 절차 및 시스템 현황, 제공 후 관리 현황 파악
 - * 이용대상, 서비스 유형·비용, 비밀보호 등
 - 통계청과 유관기관과의 공유 및 협력 관계 파악
 - 집중형과 분산형 국가의 통계제도를 비교하고, 분산형 제도에서 통계생산의 효율성을 제고하기 위한 방안 모색
- 방문기관

국 가	방문 기관
미 국	MPC_IPUMS, 미시건대학 소재 ICPSR 및 RDC
캐나다	캐나다 통계청 및 RDC

○ 연수인원: 22명(통계작성기관20명, 통계청 인솔자 2명)

기관	계	중앙행정기관	지방자치단체	통계작성지정기관
인원	20	8	5	7

- 연구 분임조 구성: 3개 조

구분	연구 기관 및 연구 주제	연구 내용
1조	<ul style="list-style-type: none"> • 캐나다 통계청 및 RDC/ 미시건대 RDC - 국가통계기관 마이크로데이터 서비스 - RDC(Research Data Center) 운영 - 분산형과 집중형 서비스체계 비교 	<ul style="list-style-type: none"> · 미국, 캐나다 마이크로데이터 서비스 및 유관 기관 간 공유,협력체계 · RDC 접근성, 제공자료범위, 절차 등 · 분산형 및 집중형 체계 비교
2조	<ul style="list-style-type: none"> • IPUMS - 국제협력 서비스 사례 	<ul style="list-style-type: none"> · 해외자료 서비스의 사회적 가치 · 제공자료 범위 및 서비스 제공 절차 · 국제협력 및 공유 방식 등
3조	<ul style="list-style-type: none"> • ICPSR - 데이터 아카이브 서비스 사례 	<ul style="list-style-type: none"> · 데이터아카이브 서비스의 사회적 가치 · 제공자료 범위 및 서비스 제공 절차 · 연구데이터, 분석방법 공유 및 교육체계 등 · 통계청과 아카이브간의 협력체계

○ 연수 세부 일정

일 정		방문지역	활동 내용
1일차	8.31(수)	인천-미네아폴리스	출장지(미국 미네아폴리스) 이동
2일차	9.1(목)	미네아폴리스	미네소타대학교 IPUMS 방문
3일차	9.2(금)	디트로이트	미시건대학교 ICPSR, RDC 방문
4~6일차	9.3(토) ~9.5(월)	휴일/ 캐나다 이동	휴일 및 노동절
7일차	9.6(화)	캐나다 오타와	캐나다 RDC 방문 일정이 9.7일로 조정
8일차	9.7(수)	캐나다 오타와	캐나다 통계청, RDC 방문/ 토론토 이동
9~10일차	9.8(수) ~9.9(목)	토론토-인천	출장복귀(인천) 이동

3. 참가자 명단

번호	기관명	직급	성명	성별	분임조	비고
1	국방부 창조행정담당관	소령	김세용	남	1조	
2	산업통상자원부 기업협력과	행정5급	이창원	남	1조	1조 연구장
3	국토교통부 정보화통계담당관	공업6급	남기한	남	1조	
4	한국감정원 부동산통계센터	과장 (4급)	김근석	남	1조	1조 서기
5	한국전자정보통신산업진흥회 조사분석센터	부장	유중현	남	1조	
6	한국형사정책연구원 범죄통계조사센터	연구원	박희정	여	1조	
7	충청북도 법무통계담당관	행정7급 (임기)	정춘영	남	1조	
8	환경부 정보화(담)	환경6급	이준한	남	2조	
9	여성가족부 정보통계담당관	6급	김지연	여	2조	
10	건강보험심사평가원 의료정보관리부	차장 (3급)	임현정	여	2조	2조 서기
11	울산광역시 법무통계담당관실	행정6급	장혜경	여	2조	
12	경기도 미래전략담당관	행정6급	김형락	남	2조	2조 연구장
13	강원도 기획관실	행정5급	최경자	여	2조	
14	제주특별자치도 협치정책기획관실	행정7급	김동환	남	2조	
15	문화체육관광부 창조행정담당관	행정6급	조은호	여	3조	
16	고용노동부 노동시장조사과	통계5급	이은경	여	3조	
17	국민안전처 정보통계담당관실	기술서기관 (4급)	최정례	여	3조	
18	한국환경공단 자원순환정책통계팀	과장 (4급)	한혜영	여	3조	3조 서기
19	정보통신산업진흥원 산업분석팀	책임연구원	이승민	남	3조	3조 연구장
20	한국교육개발원 취업통계팀	실원	손지혜	여	3조	
21	통계청 마이크로데이터과	전산5급	박미숙	여		인솔자
22	통계청 통계조정과	행정6급	한윤희	여		"

II. 연수 세부 내용

1. MPC* 방문

* MPC(Minesota Population Center)는 미네소타대학교 내의 인구센터로서 전 세계 인구센서스 마이크로데이터 제공 등 IPUMS(Integrated Public Use Microdata Series) 프로젝트를 수행하고 있음

가. 개요

- 일 시: 2016. 9. 1. (목) 09:00 ~ 15:00
- 장 소: 미국 미네아폴리스주 미네소타대학교 윌리홀(Willey)
- 참석자:
 - (MPC) Dr. Steven Ruggles(센터장), Dr. Robert McCaa, Dr. Lara Cleveland 등 MPC 관계자 7명
 - (연수단) 통계청 및 통계작성기관 22명, 통역 등 2명
- 상세 회의일정

시간	주제	발표자
09:00 - 09:30	① 미네소타 인구센터(MPC) 소개	Dr. Steven Ruggles
09:30 - 10:30	KOSTAT MDIS 서비스 소개	박미숙 사무관
10:30 - 11:20	② MPC Data 프로젝트 소개	Mr. Josiah Grover
11:20 - 11:30	휴식	
11:30 - 12:00	③ MPC Data 보안 정책	Mr. Fran Fabrizio
12:00 - 13:00	④ IPUMS-International 데이터와 메타 데이터 처리과정	Dr. Lara Cleveland
13:00 - 14:00	점심식사	
14:00 - 14:30	⑤ IPUMS-International MOU 협력기관 ⑥ IPUMS-International 이용방법	Dr. Robert McCaa Dr. Patt Kelly Hall
14:30 - 15:00	⑦ IPUMS 지도와 TerraPop 소개	Dr. Matthew Sobek

- 사전 연구 및 연수자료 정리 : 2분임

나. 기관 개요

(MPC IPUMS 주요업무)

국내 뿐 아니라 세계 각 국의 인구센서스 마이크로데이터 및 관련문서를 보관하고 있으며, 센서스 마이크로데이터 및 메타데이터를 전 세계 연구자들에게 서비스함

○ 서비스 연혁

- 1992년 IPUMS 프로젝트를 시작하였으며, 1999년 10월 IPUMS-International 프로젝트를 시작하였음

○ 운영조직

- IPUMS 프로젝트는 미네소타 인구센터(Minnesota Population Center, MPC)에서 운영되며 100명의 교수진으로 구성되어 있음
- 미네소타 인구센터, 센서스국 및 국제 데이터 아카이브의 공동 작업으로 프로젝트가 추진
- 미국국립과학재단(U.S. National Science Foundation)과 국립아동 건강 및 인구개발기관에서 주요 자금을 지원하며, 미네소타대학 연구부총장부서와 미네소타 인구센터에서 추가 지원이 이루어짐

○ 주요 서비스

- IPUMS 프로젝트는 크게 IPUMS-USA, IPUMS-CPS, IPUMS-International 3개의 범주로 진행됨
- IPUMS-International은 전 세계 약 7만 명의 연구자에게 인구조사 자료를 서비스하고 있으며, 97개 국가통계기관과 양해각서(MOU)를 체결하여 센서스 마이크로데이터 397개 세트를 위탁받아 진행 중임

○ 마이크로데이터 보유 및 주요 제공 자료

- IPUMS-USA

- 15개 연방센서스·미국지역사회조사로부터 수집된 미국 인구 샘플 제공
- 1850년부터 현재까지의 센서스 및 2000년부터 현재까지 미국 인구의 장기간의 자료 보유
- 센서스자료 간 레코드 배치와 코딩 구성 및 문서가 상이한 형태로 이루어진 장기간의 자료를 시계열적 변화를 분석할 수 있도록 균일한 코딩을 적용하여 일관된 형태로 자료를 재구성하여 서비스함

- IPUMS-CPS

- CPS(Current Population Survey, 현행인구조사)는 미국센서스 및 노동통계국이 공동으로 매달 수행하는 가구조사인데, 10년 주기로 시행되는 IPUMS-USA에 부합하도록 데이터를 재구성하여, 센서스가 시행되지 않는 연도에도 연구에 필요한 데이터를 얻을 수 있음

- IPUMS-International

- 79개국의 센서스 데이터를 제공하며, 258개의 국가별 센서스데이터를 통해서 약 5억 6천명의 개인기록을 보유하고 있음(개인수준의 센서스 데이터를 공개적으로 이용할 수 있는 가장 큰 데이터임)
- 국가간 자료를 재구성(harmonization)하여 제공함으로써 연구자는 국가간 비교연구를 쉽게 할 수 있음

○ 홈페이지: <https://international.ipums.org>

다. 연수 세부 내용

① 미네소타 인구센터(MPC) 소개

(발표자: Dr. Steven Ruggles, 센터장)

- 주요 연혁

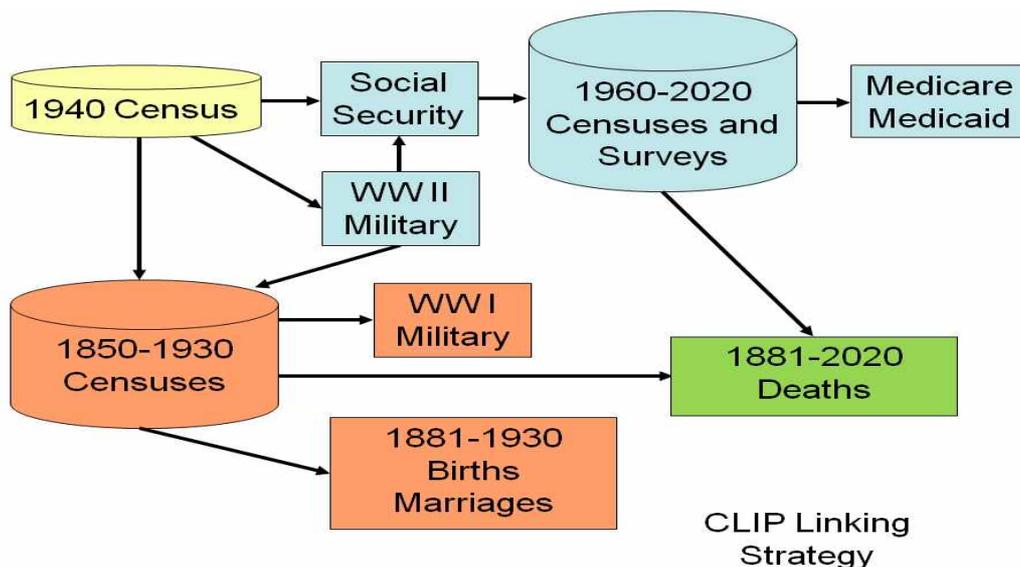
- 1960년 인구주택총조사 0.1%자료로 첫 번째 공공용 Micro Data 만듦
- 1970년대 다양한 Data 형태로 공공용 Microdata를 추가 마련 하였으며, 1960년대 자료를 1970년대 자료와 연계할 수 있도록 재구성하여 비교 분석이 가능하도록 하였음
- 1991년 매년 항목이 달라 시계열분석이 어려웠던 센서스 자료를 코드북(Codebook)을 마련하여 일관성 있는 통합DB화 진행 (1850년, 1880년, 1900년, 1910년, 1940~1990년 자료)
- 1993년 통합DB자료를 시범적으로 자료 제공
- 1994년 IPUMS 웹사이트를 개설
- 1995년 광범위한 자료 제공 시작, 여러 서비스 형태로 제공, 관련된 자료 추출 시스템 제공(연도, 표본, 데이터 형태 등)
- 1999년 센서스 자료와 다른 자료와 연계하는 서비스 시행
- 2003년 NHIS(National Health Interview Survey, 건강조사)와 센서스자료를 연계한 통합 데이터 생성
- 2005년 미국인 시간사용조사와 자료 연계
- 2009년 TERRA POPULUS(인구환경자료와 연계하는 서비스) 구현
- 2011년 IDHS(Integrated Demographic & Health Series, 아프리카, 인디아 등 개발도상국의 지역과 건강자료) 서비스 구현

- MPC의 역사적인 2대 주요 사업

- 1790-1940년대 자료 연계 사업(2019년 완료예정)
- 센서스 종단 프로그램(CLIP, 1850~1950년, 2020년 완료예정)

- * 1940년의 센서스자료를 통해 전후 자료를 연계 가능함
1940년 센서스의 주민번호와 세계제2차대전 군징집자료(WWII)를 통해 1960~2020년 센서스와 연구자료와 연계 가능하며, 메디케어와 메디케어와 연계분석도 가능함
- * 1850년-1930년 센서스자료는 1940년 센서스 자료와 세계제1차대전 군징집자료(WWI)를 통해 연계되며, 출생과 결혼, 사망 등의 자료와 연계분석이 가능함
 - ☞ 출생부터 사망까지의 개인의 삶을 연계하여 초년기 삶이 중년이후 삶에 어떻게 영향을 미치는지 연계할 수 있으며, 대표적인 연구사례로 1930년 수돗물에 납사용이 IQ에 미치는 영향조사를 들 수 있음
 - ☞ 여러 세대의 삶의 자료를 연계하여 보건상태, 삶의 상태 확인이 가능하며, 공업화·가족형태의 변화 등 대규모적인 사회변화 등을 파악할 수도 있음(현재 2백만~3백만 정도의 가구 정보를 보유하고 있음)

<자료연계 구조도>



- 주요 질문 및 토론 사항

Q1. 센서스 자료와 부동산과의 연계 분석이 가능한가

- ☞ 센서스 자료에 인구자료가 있으며, 지리정보와 연계 가능하며, 집의 소득이나 구성 등 상황요인과의 연계하여 분석할 수 있음

Q2. 인구센서스 자료와 환경자료를 연계하는 주요 이유는

- ☞ 미국 연방과학재단에서 자료를 이용하며, 환경과 관련한 인구적 특성 등을 연구하기 위한 것임

Q3. 민감한 개인정보를 수집할 수 있는 이유는

- ☞ 13개주 통계작성기관에서 72년간 공개하지 않았으나, 연방 정부의 명령으로 연계할 수 있었음. 개인정보노출이 있는 승인용 자료는 특정 공간에서만 사용이 가능하므로 제공하지 않으며 별도 제약이 존재하지 않는 공공용 자료에 대한 질을 높이는 것도 무척 중요함

② MPC Data 프로젝트 소개(발표자: Mr. Josiah Grover)

- MPC는 국제적인 연구기관으로 43개 단과대학에서 참여하고 있음. 인터넷 자료제공에 중점을 두고 공공용 data 배포와 사용법에 대한 안내 및 연구를 지원하고 있음. 특히 추가식별키 등을 만들어 제공하여 다양하게 데이터를 이용할 수 있도록 다양한 포맷의 자료를 생성하여 제공함

- IPUMS-USA:

- 1850-2010년까지의 10단위의 USA 인구총조사 자료
- 1850년, 1880년, 1940년 인구총조사 전수자료
- 1850-1880-1930 연계된 센서스 샘플자료
- 푸에르토리코(서인도제도의 섬, 미국령)의 1910-2014년 샘플자료

- 2000-2014년 연간 총조사 샘플자료
- IPUMS-CPS(Current Population Survey):
 - 1989-2015 매월 미국 고용조사자료
 - 1962-2015 매년 사회 경제 공급자료
 - 추자자료: 교육, 선거구, 경력, 흡연여부, 인터넷 사용 자료 등
- IPUMS-International
 - 78개국의 마이크로데이터 및 1960년부터 현재까지 샘플자료
매년 새로운 나라의 자료를 제공하고자 노력하고 있으며,
공공용으로 사용이 가능한 세계에서 가장 많은 자료를 보유하고 있음
- 북대서양 인구조사(North Atlantic Population Project)
 - 1703년에서 1951년까지 캐나다, 영국 , 노르웨이, 스웨덴 등
32개국 과거 센서스 자료 보유, 14개의 센서스 자료가 연계되
었으며, 사업완료시 1억 3천만명의 자료가 생성됨
- 건강조사 연계(Integrated Health Interview Series)
 - 국가 건강조사로 1963-2014년까지 매년 조사자료로 4만5천가구에
대한 샘플조사자료로 1만3천개의 변수로 구성되어 있음
- 시간사용조사(IPUMS Time Use Surveys)
 - 어디서 누구와 무엇을 하며 어떻게 시간을 보내고 있는지에
대한 기록자료, 미국 2003-2015 연간 시간사용조사 샘플자료 및
1965년, 1975년, 1985년과 1990년대의 역사적 자료 보유, 국가간
자료는 네덜란드와 스페인, 영국과 미국자료 보유
- 인구통계학적 건강조사(Demographic and Health Surveys)

- 개발도상국의 260개 조사 자료로 아이를 낳을 수 있는 여자, 동거하는 남자, 5세미만 영유아에 중점을 두고, 가족계획, 아이의 건강과 생존율, 에이즈, 여성의 지위, 영양, 건강 등에 관한 자료
- 국가 역사와 지리정보 시스템(National Historical Geographic Information System, NHGIS)
 - 1790-2010년 센서스 요약자료와 소지역단위 통합적 특징 제공 지역단위 정보와 시계열정보를 제공
 - 지역 인구자료(Terra Populus)
 - 인구 마이크로데이터와 환경자료와 연계한 서비스, 행정구역 자료와 인공위성의 지역정보를 활용하고, 온도·강수량 등의 기후적 자료와 연계하여 자료를 활용할 수 있도록 제공
 - 온라인 추출 시스템
 - 사용자가 자료와 변수를 선택하고, SAS나 SPSS, EXCEL, CSV 등의 형태로 자료를 생성할 수 있도록 서비스하고 있으며, 자료생성 등에 대한 사항을 담은 코드북을 제공
 - 온라인 분석시스템
 - 빠른 속도의 분석 프로그램을 제공하고, 별도의 통계패키지가 없어도 다년간의 마이크로데이터를 활용할 수 있도록 지원하며, 자료 활용을 위한 가이드를 웹사이트에서 제공함

③ MPC Data 보안 정책(발표자: Mr. Fran Fabrizio)

- 데이터 보안: 데이터 자체에 대한 보안뿐만 아니라 올바르게 사용자에게 전달하는 것까지를 의미하며, 데이터 구조를 구성하는 정책과 제공하는 정책으로 구성됨

- 원자료 보유여부: 원자료는 각 나라의 통계청에서 보유하고 있으며, MPC는 익명화 처리된 원시 데이터 파일만 보유함
- 물리적 보안: 미네소타주립대 데이터센터 안에 존재하며 서버 보안을 위해 단 3명만 서버에 접근가능하며 필요한 경우 3명중 1명과 동행하여야 함. 데이터센터 정책과 업무절차는 ISO/IEC 270002:2005에 근거함
- 이용자보안: 미네소타대학 정보보안 틀 안에서 PC이용하고 ISO 27001과 27002에 따른 규정을 준수하여 패스워드 강화, 화면보호기 강화, 관리자계정 등 다양한 보안 기능 구현
- 직원보안: 미네소타대학 계정을 사용하며, 별도의 MPC Local 계정은 존재하지 않음. 직원이 퇴사하면 자동적으로 계정이 삭제됨
- 원자료 독립: 원자료 초기 정제작업을 하는 경우에 5인 미만의 인원만 접근가능하며, 자료보안 알고리즘 적용 후 데이터 추출할 수 있도록 제공함. IPUMS-International은 책임사항을 인지 후 자료 사용이 가능하며, IT 보안허가규정을 준수하여야함. 30개의 서버의 드라이버 안에서 재복사하여 서버가 다운되어도 사용이 가능하며 다른 캠퍼스에 백업됨. 전원공급, 온도조절장치가 구비되어있음. 하루에 4번 정도 데이터복사가 진행된다면 30일후에 파기됨. MPC에서도 2009년부터 매주 복사를 진행하고 있으며, 미시건대학교의 ICPSR과 백업부분을 협의 중에 있음

- 주요 질의 및 토론사항

- ☞ 원자료 보안을 위해 UNIX 그룹별 보안으로 담당직원이 외부와 차단된 개인단말기에서 정제작업을 진행함

④ IPUMS-International: 데이터와 메타데이터 처리과정

(발표자: Dr. Lara Cleveland)

- 첫째: 수집정보(Input) 확보(원자료, 메타정보, 지리정보로 구성)
 - 82개 국가에서 정보를 받고 있으며, 데이터의 형태가 각기 달라 분석을 위해 정제하는 과정이 반드시 필요함. 자료식별 등을 위해 수집정보에 대한 설명자료 확보가 필요하며, 사용자를 위한 메타정보를 제공하기 위해 통계자료 지침서를 반드시 확보하고 있음
- 둘째: 표준화
 - 통계제공국가에서 공공용 자료를 제공하는 경우에는 데이터 정제과정이 필요하지 않지만, 연계 분석을 위한 표준화를 위해 원자료를 확보하고 있음
 - 표준화 과정: 원자료 확보 → 원자료와 공공데이터 점검 → 공공용 자료로 판단되면 제공
 - 데이터 점검 시 개인식별이 가능한 정보가 나오면 프로그램 개발자와 같이 SPSS등 통계기법을 활용하여 비식별처리 진행함
 - IPUMS 자료 사전(Date Dictionary)
각 샘플별·자료별 사전이며, 새로운 변수가 생길 때마다 계속 추가하고 있음. 1700-1800년대 자료정리를 위해 시작되었으며, 연구자들이 연구를 위해 표준코드를 제공하는 것을 목표로 하고 있음. 통계분석 코딩, 메타정보, 연구보조 코딩데이터를 제공

하며 통계집계방식도 관리하고 있음. 메타데이터 제공을 위해 통계질문지 등 원어민 조사표와 영어식 조사표를 제공함. 수집자료와 코딩자료를 점검 후에 자료를 연계하고 있으며, 시계열분석이나 다른 변수와도 연계가 가능하도록 구현하고 있음. 특히 데이터의 분류가 중요한데 의미를 파악하고 보정 후 비슷한 그룹으로 분류하고 있음. 분류를 포함한 모든 데이터 표준화 업무과정은 IT기술자와 연구자가 협력하여 진행하고 있으며 어떤 자료를 활용했는지 어느 부분을 수정했는지 등 모든 처리과정을 문서화하고 있음

예를 들면 결혼의 경우 분류가 각 나라별로 달라 바로 연계할 수 없어 결혼여부라는 큰 분류를 만들고 상태별로 종속변수를 만들고 있음. 이용자는 간단한 분류체계나 세부분류 체계 중 선택해서 요청하고 있음

- 셋째: 데이터 간 조화(Data Harmonization)

- 사용자가 요청 시 처리하는 것으로 연구제안서의 내용에 따라 특정 샘플과 분류자료를 제공하며, 메타정보 등 제공항목 설명자료를 제공함
- 데이터 계정을 사용하는 사용자는 자신이 요청한 자료에 처리 과정을 확인할 수도 있으며 72시간 내에 재 다운로드 가능함. 분류나 샘플을 추가 요청할 수도 있으며, 데이터는 원하는 변수 항목을 선택할 수도 있음. 5-6명의 사회과학대학원생들이 정보를 입력하고 정제하고 있음

- 주요 질의 및 토론사항

- Q1. 데이터의 변수가 무척이나 다양한데 어느 범주까지 변수를 만들지에 대한 판단 기준은
- ☞ 많은 인구조사의 경우 해당국가에 요청하기도 하며, 주로 「국제표준분류」를 적용하여 판단하며, 정확한 판단기준이 있는 것은 아니며 연구자가 상급자와 협의함
- Q2. 표준화를 진행하는 방식은(SAS 등 통계패키지 or 프로그램 개발)
- ☞ 원자료 분석을 위해 R을 활용하지만 이후에는 분석자들이 별도로 개발한 프로그램으로 분류함. 새로운 식별분류가 나오면 사전에 적용해 본 이후에 모든 샘플에 적용하고 연구자들이 최종 확인을 하게 되어 있음
- Q3. 영어버전의 조사표 등에 대해 해당국가의 감수 여부
- ☞ 전문번역가들에게 번역 의뢰 후 MPC사에서 받아서 검증 후 진행하고 있음. 정확한 자료를 위해 2년 정도의 시간을 가지고 진행함
- Q4. 새로운 조사, 새로운 변수 적용 시 식별위험에 대한 관리여부
- ☞ 식별위험 때문에 2만명 미만의 지역에 대한 서비스는 제공하고 있지 않는 등 제공자료에서 식별 위험성은 거의 없음
- Q5. 각국 자료 수집 중 개인식별정보가 포함된 경우가 있는지
- ☞ 각국자료는 공공용자료를 주는 경우가 많고, 자료제공 시 사전에 범위 등에 충분히 논의하고 있음
- Q6. 각국의 자료제공에 대한 협조는 긴밀한지, 협조를 이끌어내기 위해 어떠한 노력을 하는지
- ☞ 컨퍼런스 등을 이용해 파트너십을 확보하고 있으며, 자료

제공이 느린 경우가 있으나 각국의 연구자들의 연구에 어려움이 있음 등에 대해 설명하고 자료를 수집하고 있음

- ⑤ IPUMS-International MOU 협력기관(발표자: Dr. Robert McCaa)
 - 97개 국가통계기관 MOU를 맺음. 각국 제공 자료에 대해 설명함
- ⑥ IPUMS-International 이용방법(발표자: Dr. Patt Kelly Hall)
 - 데이터 이용, 자료추출 방법을 설명함
- ⑦ IPUMS 지도와 TerraPopulus 서비스 설명 및 시연(Dr. Matthew Sobek)
 - IPUMS 지리 연계 : 모든 국가의 1,2차 행정구역 경계 파일 생성한 후 변하지 않는 행정구역을 기준으로 삼아 연계
 - 1960년부터 현재까지의 인구나 환경을 연계한 데이터

라. 주요 활동 사진



2. ICPSR(Inter-university Consortium for Political and Social Research) 방문

가. 개요

- 일 시 : 2016. 9. 2. (금) 09:00 ~ 13:30
- 장 소: 미국 앤 아버 미시건대학교 사회과학연구소 G150AB룸
- 참석자
 - (ICPSR) Mr. Peter Granda, Ms. Stephanie Chardoul, Ms. Patricia Maher 등 ICPSR 관계자 6명
 - (연수단) 통계청 및 통계작성기관 22명, 통역 등 2명
 - (기 타) 통계청 김정란 과장(미시건대학에서 직무 연수중)
- 상세 회의일정

시간	주제	발표자
09:00 - 09:30	ICPSR director와 미팅	
09:30 - 11:00	① ICPSR 개요 및 주요 서비스	Peter Grada (Associate Director)
11:00 - 12:00	② 사회과학연구소(ISR) 소개 및 사회조사 운영(SRO)에 대한 사항	Stephanie Chardoul (Director of SRO)
12:00 - 13:30	점심식사	

- 사전 연구 및 연수자료 정리 : 3분임

나. 기관 개요

(주요업무)

세계에서 가장 많은 사회과학 경험연구 데이터를 보유하고 있는 데이터 아카이브*로 연구를 위한 풍부한 데이터 제공과 데이터 연구를 위한 교육을 실시함

* 데이터 아카이브: 물리적, 기술적 문제로 인해 손상될 수 있는 데이터를 장기보존이 가능한 형태로 저장하고, 이용자가 이를 활용할 수 있도록 체계적으로 정리하여 제공하는 기관

○ 서비스 연혁

- 1962년 미시건대학교 사회조사연구소(Institute for Social Research) 소속으로 설립되었으며, 설립 당시 정치학자를 중심으로 22개 대학 간 컨소시엄으로 구성으로 출발함
- 1963년 사회과학 양적방법 교육 워크숍인 'Summer program' 실시로 비롯되어 정기적인 자료 분석 워크숍으로도 유명
- 1968년 사회과학 영역 전반으로 주제를 확대하여 127개 컨소시엄 기관으로 범위 확대
- 1994년 'Summer program' 웹사이트 개설
- 1996년 FTP를 이용한 데이터 다운로드 서비스 시작
- 2001년 웹 다운로드 서비스 오픈
- 2009년 교육 및 학습 자원 포털인 'TeachingWithData.org'를 개설하여 예제 및 기타 교육용 리소스를 제공함
- 2011년 데이터변수, 관련 문헌에 대한 이행정보 등 전체 내용에 대한 통합 검색 제공
- 2014년 openICPSR을 런칭(무료 자료공유서비스)
- 2016년 현재 정부, 세계 대학, 연구기관 등 766개 기관이 컨소시엄 참여하고 있으며, 세계에서 가장 많은 사회과학 경험연구 데이터를 보유하고 있음

※ 국내는 사회과학데이터센터(KSDC), 고려대, KAIST, 카톨릭대, 국립도서관 등이 참여

○ 운영조직

- 미시건대학교 사회조사연구소(Institute for Social Research)에서 운영
- 회원기관 등의 컨소시엄 형태로 운영되며 저명한 학자와 자료전문가들로 구성된 운영위원회(전세계 회원들의 선출로 12명 선출에, 임기 4년, 연 3차례 회의)에 의해 운영
- 실무부서는 컬렉션제공팀, 개발팀, 큐레이션서비스팀, 특별주제 컬렉션팀, 네트워크서비스팀, 여름교육프로그램팀, 관리행정팀으로 구성

○ 홈페이지: <https://www.icpsr.umich.edu/>

다. 연수 세부 내용

① ICPSR 개요 및 주요 서비스(발표자: Peter Grada, Associate Director)

- 제공자료 종류

- Public use(공공용)와 Restrict(승인용) 데이터로 구분되어 제공되고 있으며, 보건, 건강, 인구통계 등 다양한 연방정부 후원 통계 및 개인제공 데이터를 제공
- Public 데이터는 모든 사람이 열람이 가능하나, Restrict 데이터는 회원에 한하여 제공되며 '제한자료 이용서약서'를 작성 제출하면 이를 검토하여 승인한 경우에만 이용 가능함
- 특히, ICPSR의 Data Discovery Tools는 키워드 검색, Data와 관련 논문, 출판물 등의 하이퍼링크를 제공하여 사용자 편의 제고

- 자료수집 방법

- 연구기관과 개인연구자가 산출한 자료를 기탁 받거나, 미국 센서스국 자료를 수집
- 매년 400~500건의 자료를 수집하고 있으며 미국 센서스국으로부터 7종의 마이크로 데이터를 제공 받음

<미국 센서스국이 제공한 마이크로데이터>

자료명	제공년도 범위
U.S. Census Data	1940-2000년 (10년 주기)
American Housing Survey Series - American Housing Survey - Annual Housing Survey	· 1984-2007년 (1년 주기) · 1973-1983년 (1년 주기)
Current Population Survey Series	1968-2012년 (월별 또는 분기별)
County and City Data Book Series	1944-1977, 1983, 1988년
County Characteristics	2000-2007년
Survey of Income and Program Participation(SIPP) Series	1984-1993, 1996, 2001, 2004년 Panel
TIGER(Topologically Integrated Geographic Encoding and Referencing) database/Line Files	1990, 2000, 2002, 2004-2006년

- 비밀보호 방법

- 통계적 분석 목적에 한하여 데이터 사용이 허용되며, 제공 전 모든 데이터세트를 검수하고, 이용자가 응답자의 프라이버시를 침해하지 않아야 할 책임을 부여
- 자료내의 직접적인 식별정보(이름, 주소, 전화번호 등) 뿐만 아니라 간접적인 식별정보(거주지, 교육기관, 직업명 등) 역시 타 변수와 연계하여 식별가능성이 존재함으로 포괄적인 리코딩 등 비밀보호 적용

- 서비스 비용

- 회원제로 운영하고 있어, 대학·정부 및 민간기관 등이 연회비를 납부하면 해당 기관의 소속인은 자료를 무료로 이용 가능

<ICPSR 회원의 종류와 연회비>

구 분		2016년	2017년, 2018년
US academic institutions	Doctoral Research Extensive	\$ 16,540	\$17,400
	Doctoral Research Intensive	\$ 9,450	\$9,900
	Master's Comprehensive I and II	\$ 6,825	\$7,100
	BA Universities	\$ 2,100	\$2,200
	Specialized Institutions	\$ 2,100	\$2,200
	Associate/Community Colleges	\$ 1,680	\$1,765
· Nonprofit, government, and academic · institutions outside the United States	Associate I	\$ 16,550	
	Associate II	\$ 11,000	
	Associate III	\$ 5,550	
	Associate IV	\$ 2,200	

- Summer Program

- 6,7,8월 4주, 1주 과정의 교육프로그램(81개 과정, 101명 교사와 40명 조교)을 운영하며, 사회과학 데이터 관련 교육을 제공하고 있음. 매년 1,000여명의 교육생을 배출(4개국, 300개 기관, 30개 교육 기관)

- 보유 예산

- 3,600만불 (2016년), 4,066만불 (2017년) (400만불 규모의 예비비 보유)
(정부 지원금, 법무부(NACJD) 펀드 및 회원 연회비로 충당)

- 기타 활동

- 논문 경진대회, 인턴십, VDE(Virtual Data Enclave)* 제공 및 컨소시엄 체결 추진

* VDE(Virtual Data Enclave) : 사이버데이터 저장소로 50개 단말기 운영, Remote Data 접근에 대한 서면 요청 및 승인을 득한 후 열람토록 함

- 제공시스템

- ICPSR웹사이트(<http://www.icpsr.umich.edu>)

- 검색 서비스

소장 자료를 19개 주제로 분류하여 제공, 시리즈 자료들을 자료명 알파벳순으로 제공, 조사대상 국가 및 지역을 알파벳순으로 제공, 연구자 이름순으로 제공, 국제비교자료를 8개의 주제별로 제공 등

- 메타데이터 서비스: 개별 조사자료별로 메타데이터를 제공함

- 공유 및 협력

- 데이터 관리 및 디지털 큐레이션 분야, 특히 응답자 정보 노출 위험, 응답자 보호, 사회과학 자료의 디지털 보존 및 메타데이터와 관련 문서에 대한 연구를 주로 수행하며, 역사인구학 및 환경에 관한 연구, 20개 주제 컬렉션 구축 등 관련 연구기관과 긴밀한 협력관계 유지

- 연기기관 외에 DDI(Data Documentation Initiative) 이사회 멤버로 참여하고, Data-PASS(Data Preservation Alliance for the Social Sciences)의 파트너로 활동하고 있으며, 미국 범죄 재판

데이터 아카이브(NACJD), 건강 보건 아카이브(HMCA) 등 다양한 아카이브와 협력관계를 맺고 있음

- 주요 질문 및 토론 사항

Q1. ICPSR의 데이터 제공 승인 기준은

☞ ICPSR의 제공 데이터는 연구목적의 자료 활용에 큰 제약사항은 없으며, 다만 기밀자료 혹은 개인 식별이 포함된 자료의 경우 서면 요청 후 데이터관리 책임자의 허가 및 RDC를 통해 데이터 분석 및 결과 값을 수령함

Q2. 데이터 제공 승인관련 기구는

☞ 승인 위원회 등 별도 행정기구는 없지만, 데이터 제공기관에서 1차로 판단하고 2차로 ICPSR의 데이터 책임자가 심의를 함

Q3. ICPSR의 데이터 콜렉션 방식은

☞ 연구자의 자발적 제공과 ICPSR에서 특정논문을 선정 후 연구 경로 등을 문서화하는 것을 연구자와 협의 후 자료를 제공(연구 시작단계에 컨택하여 통계 획득 장려) 받거나, 인터넷 검색을 통해 최근 이슈가 되는 연구에 대한 자료를 직접 수집

☞ 제공되는 자료는 SPSS, Access 기반이나, SAS, R 등으로 서비스 확대 중임

Q4. 회원기관, 연구자들의 자료 기탁 협력을 이끌어내기 위한 ICPSR의 노력은

☞ 초창기 많은 노력이 필요했을 것이라는 예상과 달리 초창기 사회과학분야의 아카이브가 ICPSR 밖에 없어서 회원의 자발적 데이터 등록이 활발하였고, 회원 가입이 활발했으나, 최근 구글 아카이브 등 경쟁 서비스 출현, 다양한 경로를 통한 정보 획득

등으로 자료 등록 독려에 어려움이 있음

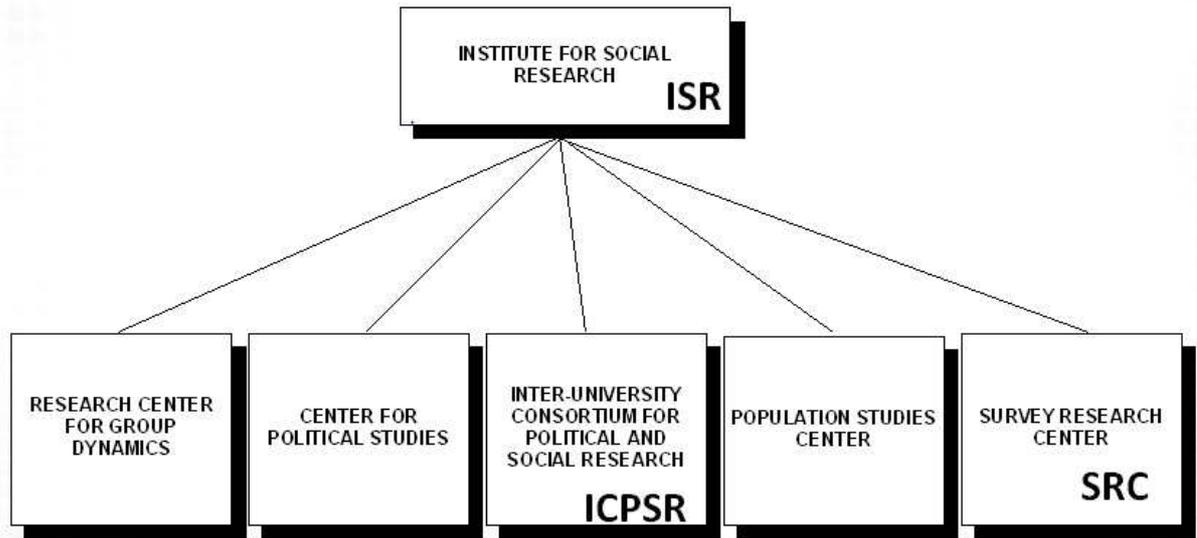
- ☞ ICPSR만의 특화된 노하우로서는 연구에 편의성 제공이 장점으로 작용하고 있음(연구자 URL로 지속적으로 업로드되어 있어서 해당 데이터가 어디에 인용되었는지 어느 연구에 사용되었는지 확인이 가능)

Q6. 데이터 보관 기한은

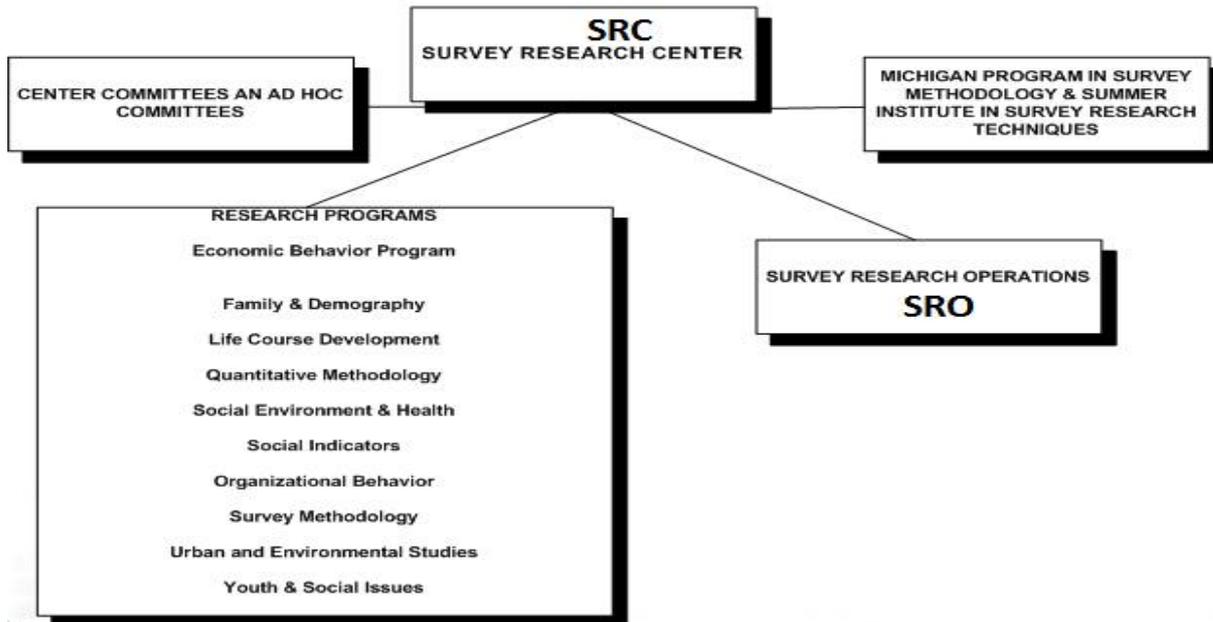
- ☞ 모든 데이터는 영구 보관을 원칙으로 함
 - ☞ 데이터의 생성 방식이 변경될 경우 과거년도 자료에 대한 업데이트가 중요한 사업으로 추진되고 있음
- 예) 1800년 미국 인구센서스 책자 자료를 파일화하여 보관,
1900년대의 DSPSS 자료를 SPSS 자료로 변경 보관 등

② 사회과학연구소(ISR) 소개 및 사회조사 운영(SRO)에 대한 사항
 (발표자: Stephanie Chardoul, Director of SRO)

- 사회과학연구소(ISR) 조직 구성



- 조사연구센터(SRC) 조직 구성



- 조사연구센터(SRC) 개요

구분	주요 내용
개요	<ul style="list-style-type: none"> · 미시건대학교 사회조사연구소의 부설로 통계조사 특화기관이며 설문조사 및 통계조사 방법론 연구 수행, 통계조사 컨설팅 및 조사대행 서비스를 제공 * 통계 설계 전문연구자 145명, 2016년 현재 80개 프로젝트 수행 중 * 미시건대 부설기관 중 정부지원금 최대 수령 기관(의대와 동급)
기관 기능	<ul style="list-style-type: none"> · 연구제안, 표본설계, 통계분석, 통계조사의 기술적 방법 개발, 통계조사(대면조사, 인터넷, 이메일 조사 등 다양한 방법으로 실시) 등
조사 방법	<ul style="list-style-type: none"> · 면접조사(면접 Station 60개 운영 중), 콜센터, Web 기반 설문조사센터 운영 중 * 경제적 효율성, 조사 성격, 목표 응답률 등을 고려하여 조사 방법을 선정

- 조사연구운영부(SRO) 개요

구분	주요 내용
개요	<ul style="list-style-type: none"> · SRC내에 Data Collection Operation을 위한 부설 조직으로서 면접원의 구직, 채용, 교육훈련, 급여 등 종합적인 인사 관리 수행 조직
예산/인사	<ul style="list-style-type: none"> · '16년 36백만불, '17년 40백만불 예산 규모이며, 펀드의 절반이상이 연방정부 지원금으로 구성 · 직원 절반 이상이 외부인사로 구성

- SRO 데이터 수집 방법

- 분산형 데이터 수집(De-centralized Data Collection) 및 집중형 데이터 수집(Centralized Data Collection) 방식 혼용
- De-centralized Data Collection(분산형_면접원 조사)

<ul style="list-style-type: none"> • CAPI: 대면조사 • 청경, 관찰 • 특수조사: 물리학적/생물학적/지식경험적/그룹면접 등 	<ul style="list-style-type: none"> • CATI: 전화조사
--	--

- Centralized Data Collection(집중형_SRC내 위치한 Survey Services Laboratory (SSL) 빌딩 안에서 조사)

<ul style="list-style-type: none"> • CATI: 콜센터 • 통계품질 관리 	<ul style="list-style-type: none"> • 데이터 입력 • 이메일 조사 	<ul style="list-style-type: none"> • Coding • CATI programming
--	--	--

- 조사원 채용 시 체계적 교육 수료 후 시험을 통과해야만 고용되며, 온라인 교육 등을 이용하여 수료가 가능함
- * 교육 등 전문 지식을 요하는 조사의 경우 조사원에게 관련자격증 및 학위를 요구하는 경우도 있음

<ul style="list-style-type: none"> • 코딩 방법 • 편향되지 않도록 조사하는 방법 등 	<ul style="list-style-type: none"> • 설문지 해석 법
---	--

- 모든 조사원은 표준화된 면접용 도구(Equipping Iwers)를 지참하고 면접 실시
- * 면접용 도구: 체중계, 혈당계, 혈압측정기, 줄자 등
- 조사원 사후 관리 철저(면접 내용 녹음 후 사후 평가, 조사원 네트워크 교육/Web Base Meeting/컨퍼런스 콜 등)

- 질의 및 토론 사항

Q1. 민감정보 데이터 획득을 위한 노하우는

- ☞ 대면인터뷰보다 단독 공간에서의 컴퓨터 설문 답변 방식이 민감정보 데이터 획득률을 높인다는 연구 결과에 따라 컴퓨터를 이용한 A-CASI 방식으로 집계하고 있음

Q2. 조사원에 훈련 기간은

- ☞ 일반 조사의 경우 2일간 교육, 전문조사는 7~8일간 교육이 진행되고, 교육 후 테스트를 통해 숙지된 조사원에 한하여 채용함. 조사원과 이메일, 온라인 회의를 통해 업무 효율성을 제고하고, 시범 조사를 통해 조사원의 일일 업무량을 할당함.

Q3. 인터뷰에 참여하는 사람의 보상체계 및 조사원 보상체계가 궁금함

- ☞ 케이스에 따른 다양한 금전보상체계가 구성되어 있으며, 대개의

경우 60~90분 인터뷰의 경우 일반적 조사에 50~80불, 특수한 조사에 100~150불에 달하는 보상이 주어짐. 조사원의 경우 시간당 10~18불의 최저보상체계에 그 밖의 비용이 더해지며 시간에 쫓겨 조사품질이 낮아지지 않도록 건당이 아닌 시간에 비례하여 정산함

Q4. 목표 응답률은

☞ 패널 조사(기참여)는 85~ 95%, 신규 조사는 60 ~ 70%, 전화 조사는 20 ~ 40% 수준

라. 주요 활동 사진



3. 미국 미시건대학교 RDC 방문

가. 개요

- 일 시: 2016. 9. 2. (금) 13:30 ~ 14:00
- 장 소: 미국 앤 아버 미시건대학교 사회과학연구소 G150AB룸
- 참석자:
 - (MRDC) Clint Carter(MRDC Administrator)
 - (연수단) 통계청 및 통계작성기관 22명, 통역 등 2명
 - (기타) 통계청 김정란 과장(미시건대학교 공무연수 중)
- 상세 회의일정

시간	주제	발표자
13:30 - 15:30	미시건대학교 RDC 소개	Clint Carter (Administrator, MRDC)
15:30 - 16:00	MRDC 견학	

- 사전 연구 및 연수자료 정리 : 1분임
- 미국의 마이크로데이터 서비스 전반에 대한 연구는 기 연구 보고서, 센서스국 홈페이지 등을 통해 사전 연구로 진행

나. 미국 센서스국 마이크로데이터 서비스 개요

(주요업무)

미국은 전형적인 분산형 통계체계로서 전적으로 통계활동만을 수행하고 있는 기관은 상무부 센서스국, 노동부 노동통계국, 보건복지부 보건통계센터 등 11개 기관이 있음

센서스국(Bureau of the Census)은 인력 규모가 가장 방대하며 사업체모집단 DB의 구축 및 운영, 10년 주기의 인구·주택센서스, 5년 주기의 경제센서스 실시, 타 기관에서 의뢰 받은 통계조사의 실시 등 중앙통계기관으로서의 역할을 수행하고 있다고 할 수 있음

센서스국은 산하에 경제연구센터(CES, Center for Economic Studies)를 두고 마이크로데이터 관리 및 서비스를 수행하고 있음

○ 서비스 연혁

- 1964년: 마이크로데이터 제공 시작
- 1983년: 센서스국 산하에 경제연구센터(Center for Economic Studies, CES) 설립. CES 내부에 마이크로데이터 이용센터인 연방통계 연구데이터센터(Research Data Center, RDC) 설립
- 1998년: RDC 프로그램 확대

○ 운영 조직

- 마이크로데이터의 공개 여부는 센서스국의 자료 공개 검토위원회(Disclosure Review Board, DRB)에서 결정하고, 마이크로데이터의 제공 및 관리는 센서스국 산하의 CES에서 담당
- 미국 센서스국은 자체적인 서비스와 더불어서 대학 등 타 기관과 협력하여 다양한 형태로 마이크로데이터를 제공하고 있음

○ CES 주된 기능

- 센서스, 기업체와 가구조사, 고용인-피고용인 결합자료, 정부기관의 행정자료 등을 이용하여 다양한 연구를 진행하며, 센서스와

서베이, 행정자료의 마이크로데이터를 사용한 공공용 자료를 생성·관리하고 RDC를 관리함

○ 제공 자료의 종류

- 센서스국에서 직접 수행한 조사의 마이크로데이터가 주된 자료이며, 최근에는 RDC를 통해서 AHRQ(Agency for Healthcare Research and Quality)와 NCHS(National Center for Health Statistics)의 자료도 제공. 센서스국에서는 현재 약 7개 분야 60여종의 마이크로데이터를 제공 중

○ 제공 자료의 유형

- 익명화 수준에 따라 크게 공공용(Public Use Microdata Sample, PUMS)과 승인용으로 구분되며, 인구조사와 같이 많이 활용되는 마이크로데이터는 높은 수준으로 익명화하여 공공용으로 제공되며, 대부분의 자료는 승인용으로 분류되어 CES에 의해 엄격하게 관리됨

○ 서비스 이용 대상

- 외국인을 포함한 미국 내 거주민과 미국 내 정부정책연구 공무원
- 상업적 이용은 엄격히 제한됨
- 영토 외 거주 미국인 및 외국인은 상담을 통한 심의과정 거침

○ 서비스 유형

- 공공용은 웹 등 온라인 서비스하며, 일부는 원격접근서비스 방식으로도 제공됨
- 승인용은 RDC를 통해 제공함

○ 서비스 요금 체계

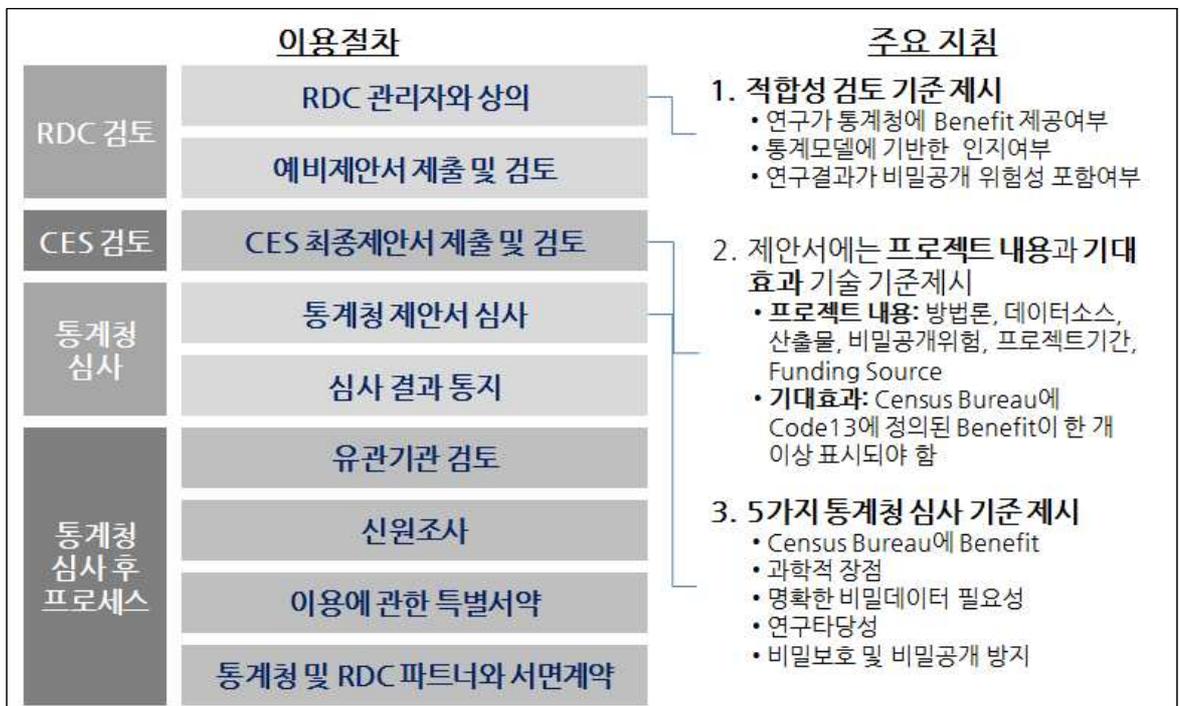
- 공공용은 무료로 제공되며, CD-Rom, DVD-Rom의 경우 장당 \$100 받음

- 승인용은 RDC별로 기본요금에 대한 과금체계가 다르며 캘리포니아 UCLA RDC는 표준요금을 데이터 건당 사용기간 단위로 산정하는데, 월 \$3,125에 연간 최저 비용이 \$7,500임

○ 승인용 자료이용절차

- 승인용 자료의 경우 RDC → CES → 통계청 → 유관기관의 순서로 심사절차를 거침

<승인용자료 이용절차>



○ 공유 및 협력

- 센서스국은 마이크로데이터 서비스를 위하여 대학과 비영리 연구기관 및 정부기관을 포함한 50개 이상의 연구기관과 RDC 파트너 협력 관계를 맺고 있음

다. 연수 세부 내용

① 미시건대학교 RDC 소개(발표자: Clint Carter, Administrator of MRDC)

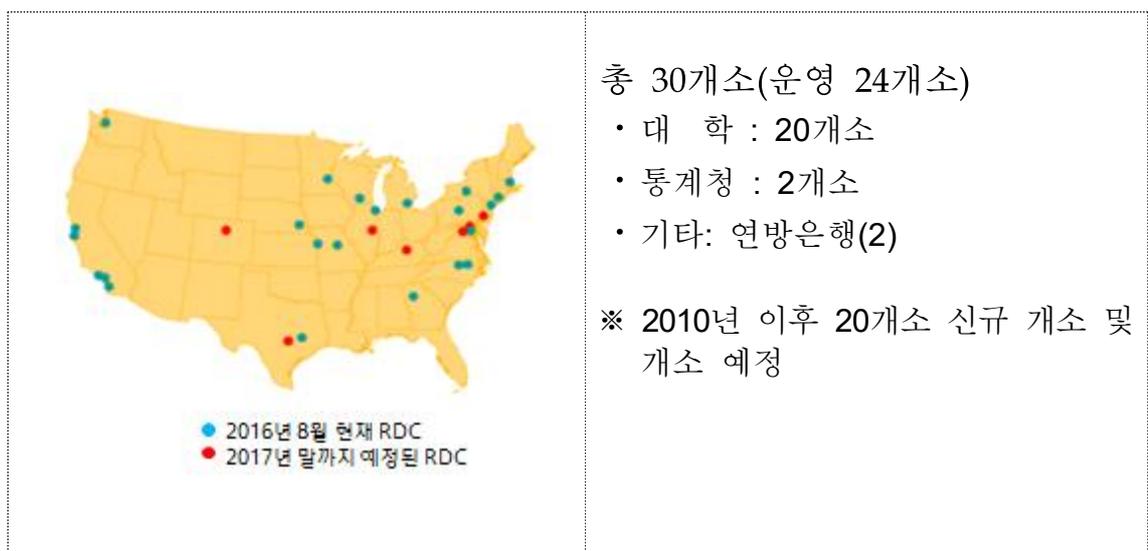
- RDC란

- 이용자 제한 자료를 연구자가 안전하게 이용할 수 있는 연구소임.
센서국을 비롯한 타 정부로부터 제공 받은 공공용 이외 자료를 제공하고 있음
- MRDC는 ISR(Institute for social Research) 기관 내에 위치함

- RDC를 사용하는 이유

- 응답자의 지리정보 등 보다 구체적인 자료를 제공받을 수 있음
- 자료 연계 및 기존 연계된 데이터셋 이용 가능
- No top-coding자료 이용 가능
- 더 큰 규모의 샘플자료를 이용할 수 있음
- 더 상세한 변수 정보(인종, 출생지, 질병코드 등)를 제공 받음
- 어떤 자료는 공공용 버전이 없음

- 미국 RDC 현황



- RDC에서는 자료 간 연계 가능

- PIK(Personal Identification Keys), 사업체·법인식별Key가 있는 내부자료 간 연계
- 외부 자료를 가져와 연계 가능(개체식별정보인 SSN, EIN, 이름, 주소, 지리정보 등 활용 가능)

- RDC 제공 자료

인구가구자료
<ul style="list-style-type: none"> · 10년단위 인구조사 자료: 1960, 1970, 1980, 1990, 2000, 2010 · 미국 커뮤니티 조사자료(ACS, 매년 마이크로자료): 1996-2011 · Current Population Survey(CPS) · Survey of Income and Program Participation (SIPP) · 미국가구조사(AHS) · 국가범죄 및 희생조사(NCVS) · National Survey of College Graduates (NSCG) · 국가 장기 사망률 연구 (NLMS) · National Longitudinal Surveys (NLS) <ul style="list-style-type: none"> · Young/Mature Men/Women · 행정자료 등
사업체자료
<ul style="list-style-type: none"> · Economic Census <ul style="list-style-type: none"> · ASM, ARTS, SAS · LBD and ILBD <ul style="list-style-type: none"> · Quarterly Financial Report (QFR) · Survey of Business Owners (SBO) · MEPS-IC · Annual Capital Expenditures Survey (ACES) · Business Research & Development and Innovation Survey (BRDIS) · Manufacturers' Shipments, Inventories, and Orders (M3) · Survey of Pollution Abatement Costs and Expenditures (PACE) and Manufacturing Energy Consumption Survey (MECS) · Commodity Flow Survey (CFS) · Longitudinal Foreign Trade Transactions Database (LFTTD)

보건자료
<ul style="list-style-type: none"> · Agency for Health Care Research and Quality (AHRQ) <ul style="list-style-type: none"> ··MEPS-HC · National Center for Health Statistics (NCHS) <ul style="list-style-type: none"> ··National Health Interview Survey (NHIS) ··National Health and Nutrition Examination Survey (NHANES) ··National Survey of Family Growth (NSFG) ··National Vital Statistics System (NVSS) ··National Health Care Surveys - NAMCS, NHAMCS, NHDS, NNHS, NNAS, NSRCF, NSLTCP

- RDC 자료 접근 및 보안환경

센서스자료 열람 방법
<ul style="list-style-type: none"> · 자료의 필요성 확인 <ul style="list-style-type: none"> ··www.census.gov/ces/ · RDC administrator 와 협의 <ul style="list-style-type: none"> ··사전제안서 ··2페이지 요약서 · 제안서, PPS · 통계청 및 타 기관 검토
미 통계청 데이터의 특이성
<ul style="list-style-type: none"> · 자료제공 시 응답자 고지의무 없음 · 통계청 시행조사에 의무 참여 · 자료 이용에 법적 근거 존재 <ul style="list-style-type: none"> ··자료 이용의 "주된 목적"이 미 통계청에 호혜적이어야 함
자료이용 주목적 성명
<p>※ 자료 이용자는 자료 이용 주목적을 성명해야 함</p> <ol style="list-style-type: none"> ① 미 통계청 시행 정책과 일관되고 유관한 자료 수집 목적 및 체계에 대한 평가 ② 미 통계청 시행 업무에 영향을 줄 수 있는 인구학, 사회 혹은 경제 과정에 대한 분석, 특히 통계청이 공개하는 제품의 품질을 검증하거나 제고하는데 유용한 방법일 것 ③ 미 통계청 데이터의 공공 정책, 인구학, 경제학, 사회연구 등 연구에 이용하기 더욱 용이한 방법 개발 ④ 행정보조, IT 보조, 프로젝트 관리감독, 적법한 감사 등을 통한 센서스 및 서베이 데이터획득시행 보조 정제, 배포 <p>* ①~④는 Demographic projects only</p>

- ⑤ 데이터 품질 이해 및 제고
- ⑥ Title 13, Chapter 5에 적시된 서베이, 센서스 혹은 추정치에 사용되는 방법론의 제고 혹은 개발에 기여
- ⑦ 수집된 자료 품질 제고
 - 미응답 대체 향상
 - 시간/변수간 연결
- ⑧ 데이터 수집에 필요한 사업체등록, Master 주소기록, 산업 및 지리분류체계의 한계 발견 및 발전방향 제시
- ⑨ 수집된 자료 및 수집방안의 한계 발견 혹은 신규 자료 수집 수요 문서화
- ⑩ 표본 frame 형성, 확인, 발전
- ⑪ 인구수 추정 및 인구 특성 파악
- ⑫ 미응답 추정을 위한 방법론 개발
- ⑬ 통계적 가중치 개발

자료이용을 위한 특별서약

- 오프라인 서류작성 및 기록
- 온라인 트레이닝 수료의무
- 신원조회
 - ..온라인 제출 후 대면면접
 - ..거주기록
 - ..해외 방문기록
 - ..교육 및 직장 기록
 - ..주변인
- 지문 채취

연구 환경

- 모든 기밀자료 이용 연구는 RDC 내에서 시행해야 함
- 컴퓨터 환경
 - ..초기 산출물 RDC 내부서 열람
 - ..인터넷 접근 불가
 - ..유닉스 환경
 - ..SAS, Stata, R, Matlab 제공 (기타도 가능)
- NCHS 이용 연구원들은 Admin이 상주 시 이용할 수 있음

기밀보안 프로세스

- 모든 산출물은 검토 전 반출 불가
- 미리 검토 준비해야 함
- RDC Admin 이나 기관의 기밀검토관이 담당
 - ..제안서 단계에서 기밀 검토 계획 작성
 - ..1~2주 걸리지만 3~4주 걸릴 때도 많아
 - ..Descriptive data 이용 시 기밀검토 불통과 가능성 높아

- 주요 질의 및 토론 사항

Q1. 카네기멜론대 RDC 패쇄 사유를 혹시 아시는지

☞ 카네기멜론대학교의 사례는 유일한 RDC 패쇄 사례입니다. 패쇄 이유에 대해서 잘 모르겠으나, 수요가 낮아서일 수도 있습니다.

Q2. 가구조사와 비즈니스데이터 연계가 유용하다고 하셨는데, 연계에 필요한 고유키가 있는지 그리고 유용하게 활용되는지. NCHS 현재 이용 현황이나 연구가 어떻게 진행되는지

☞ LEHD의 경우, 피고용인에 대한 데이터는 각 주의 사회보장보험 등을 제공하는 고용센터에서 얻은 행정데이터를 사용하고, 고용자 데이터 또한 이에서 얻습니다. 이 자료들을 센서스 자료와 연계합니다.

Q3. 기존에 연계한 자료를 RDC에서 제공하나요, 혹은 연구자들이 직접 자료들을 연계하나요

☞ 연구자들이 직접 고용정보, 사업자, 개인 데이터 등을 연계합니다. NCHS의 경우 보건학 연구이기에 보건학, 경제학, 공공정책학 연구자들이 NCHS 자료를 사용합니다. RDC에서 사용하는 자료의 절반이 NCHS 자료입니다.

Q4. RDC 이용 연구들이 센서스에 어떤 긍정적인 영향을 줬나요

☞ 거의 모든 RDC 시행 연구들은 전체 인구의 추이에 대한 추정치를 제공하였습니다. 다른 benefit으로는 외부 자료를 내부자료와 연계하여 bridge file을 만든 것입니다. 또, 외부 자료의 변수들을 가져와 Census Bureau의 설문 변수체계와의 차이를 파악하여, Census Bureau의 설문조사 등에 기여를 할 수 있습니다. 본 RDC에서 시행하는 LEHD자료 이용 연구의 경우, ORGUS라는 외부 상업데이터와 LEHD데이터를 연계하여 미 통계청에서 잘 파악하지 못하는 다국적기업을 식별하고

있다고 하였습니다.

Q5. 연구자들이 RDC를 이용하는데 매우 긴 연구제안서 평가 기간에 대해서 complain을 한 적이 없습니까? 연구자들이 심사과정 중 어떤 과정에 대해서 가장 불편해합니까

☞ 이용자들이 이에 대해서 항의한 적이 없고, 이용신청기간이 오래 걸린다는 것에 대해서 당연하게 여깁니다.

Q6. RDC 이용신청서가 기각되는 경우는 있습니까?

☞ 처음 연구제안서를 제출할 때 연구제안서를 잘 쓰지 못해서 이용신청이 기각된 적이 많습니다. 보통 이에 대해서는 수정을 기하면 이용신청이 허락됩니다. 보통 신원조회에 60일이 소요 되고, 연구제안서 심사에 6개월에서 1년까지 걸립니다.

Q7. 최근 RDC가 급격히 증가하고 있는데, 이에 대한 이유가 있습니까?

☞ 미국에서 RDC의 증가는 대학교들의 수요에 의해 (Demand-driven) 이뤄졌습니다.

Q8. RDC 운영에 많은 비용이 들텐데, 인프라 구축, 인원, 통계분석 프로그램의 비용부담주체는 누구입니까?

☞ 제가 알기로 물리적인 인프라는 유치기관에서 부담하고, 통계청에서 IT 소프트웨어 제반, 예를 들어 분석프로그램, 보안 연결, thin client 등을 조달한다고 알고 있습니다. Census Bureau가 제 생각으로 유치기관에서 설치비를 받아서 이를 구매한다고 알고 있습니다. 본 대학 RDC 컨소시움에는 Michigan State 대학교와 본교 내 각 단과대에서 자금을 조달한다고 들었습니다. 인원의 월급의 경우, 기관에서 통계청에 overhead 비용을 주고, 미 통계청에서 이 돈을 다시 각 직원에게 지급함으로써 알고 있습니다.

Q9. RDC Analyst, Administrator 및 Executive Director의 소속 기관은 어떻게 됩니까?

☞ RDC Analyst, Administrator는 모두 통계청 직원이지만, 유치 기관에서 이들의 고용을 전담합니다. Analyst와 Administrator들은 각 RDC Executive Director밑에서 일합니다. Executive Director의 경우, 각 유치기관 소속이고 통계청 공무원이 아닙니다. RDC Administrator들은 통계청에만 보고합니다. 카네기 멜론대학교에서 10년 전에 RDC Administrator들은 각 기관 소속이었습니다. 갑자기 정책이 바뀌어서 통계청 소속 공무원이 각 RDC에 배치되기 시작했습니다.

Q10. RDC Administrator와 Executive Director의 업무 역할이 어떻게 되나요?

☞ RDC Administrator인 저는 RDC 보안유지, 이용자 및 연구자 보조, 자료 기밀노출심사, IT 인프라 관리, 이용자의 제공 문의 응대, 신규 이용자 유치를 위한 활동 등을 합니다. Executive Director는 자금조달 및 운영비 관리, 유치기관과 연방정부 통계작성기관간 관계 유지, RDC에서 더 많은 데이터가 제공가능하게 협의하는 등 더 큰 관리역할을 맡습니다.

Q11. RDC Administrator를 신규 채용할 때 요구되는 사항, 예를 들어 통계적 지식, 학위 등

☞ 각 RDC Executive Director들의 각 RDC에서 필요에 의해 Administrator를 고용합니다. 어떤 곳에서는 학사학위만 있으면 되고, 어떤 곳에서는 더 높은 학위를 요구합니다. 이는 모두 각 기관의 연구 수요에 의해 결정됩니다. 요구되는 학사학위 전공은 없고, 사회학, 경제학, 인구학, 보건학 등을 전공한 사람들이 많습니다. 모든 RDC에는 한명의 연구자가 위치해 있습니다. Branch RDC에는 프로젝트가 적기 때문에 part-time

administrator들이 있습니다. 현재 저는 26개의 연구를 관리하고 있습니다.

Q12. RDC Administrator들이 심층적인 통계분석도 도와줍니까?

☞ 이는 연구자들이 미리 가지고 있어야 하는 소양이고, 각 연구자들이 시행해야 하는 부분이기 때문에 제가 돕지 않습니다.

Q13. 공동연구 지원 기능이 어떻게 됩니까?

☞ RDC를 이용할 경우, 공동연구자들이 공유하는 연구 프로젝트 directory에 접근이 가능합니다. 각 연구자들은 본인 명의의 계정으로 이 directory에 자료를 업로드해서 열람할 수 있고, 공개범위 또한 설정할 수 있습니다. 각 dataset은 고유의 directory가 있고, 연구제안서에서 신청한 dataset과 directory만 열람할 수 있습니다.

Q14. 여러 RDC가 있으므로 중복연구가 있을 수 있는데 연구제안서의 중복 검사 등 RDC간 협조 프로세스가 있습니까?

☞ 각 연구는 개별연구이기 때문에 중복연구라 볼 수 없다고 판단하여 이러한 협조는 없습니다. 다만, 연구 결과 및 산출물 승인을 할 때, 비슷한 주제에 대한 다른 산출물을 비교하여 합칠 때 노출위험이 있을 경우에는 이에 대한 관심을 가질 수 있습니다.

Q15. 중단연구 시행의 경우 지속적으로 연구를 시행해야 할텐데, 이에 대한 중간 산출물 저장 등은 합니까?

☞ 우리는 자료 열람 및 산출물 생성 등 모든 자료를 5년동안 미 통계청 중앙서버에 저장하고, 그 이후 이를 압축하여 영구적으로 저장합니다. 한 프로젝트는 평균적으로 5년 동안 RDC를 이용할 수 있고, 만약 지속적으로 RDC를 이용하고 싶으면 새 신청서를 제출하여 지속적으로 사용할 수 있을텐데, 새로 이용하는 연구에 기존 연구자료를 이어서 사용하고 싶다면 산출물, 분석

자료, 추출자료 등 RDC 내에서 사용한 자료를 모두 다시 제공할 수 있습니다. 압축된 백업자료를 다시 획득하는데 2일 정도가 걸릴 것입니다.

Q16. RDC를 통해서 만든 연구결과들을 공유합니까? 만약 공유한다면, 인센티브나 공유 체계 등은 어떻게 됩니까?

☞ 연구결과는 모두 미 통계청에 제공해야 하지만, 이러한 연구결과를 통계청이 민간에 공유할 의무는 없습니다. 모든 RDC 이용 연구들은 Working Paper를 RDC Annual Report에 공시해야 합니다. 또, 연구자들은 논문을 작성할 것이고, 이들 논문은 어차피 출간되기 때문에 공유된다고 인지할 수도 있습니다. Working Paper들은 CES Working Paper Series에서 열람할 수 있습니다. 만약 RDC를 이용했는데, 연구결과를 내지 못한다면, 향후 동일한 연구자가 RDC 이용신청 시 기각될 수도 있습니다. 이는 강제적인 연구결과 도출 및 Census Bureau와의 연구결과 공유 인센티브로 볼 수도 있습니다.

Q17. RDC별 이용자 과금체계가 동일합니까?

☞ 아닙니다. 각 RDC에서 이를 따로 관리합니다. 미시간대의 경우, 외부연구 프로젝트 당 연간 1만 5천불입니다.

Q18. RDC 제공 자료의 연계 키는 어떤 수준에서 제공합니까?

☞ 주민등록번호, 사업자번호 등과 같은 민감한 정보는 거의 없이 PIK를 통계청에서 생성해서 제공합니다. 연구자가 개인식별자료가 들어 있는 자료를 반입할 때에도 Census Bureau에서 PIK 작성비를 받고 대신 만들어서 제공합니다. 연구자는 결국 개인이 누구인지 모른 상태에서 자료를 접근하게 됩니다. 이를 만드는 유관부서는 기밀사항으로 알고 있습니다. 한 dataset에서 PIK를 만드는데 평균적으로 1만 5천불이 드는데, 보통 사업체 관련 외부자료를 연계할 때, 이미 사업자등록번호가 있을 가능

성이 높기 때문입니다. 인구통계 자료에서 이름과 SSN이 없는 경우, 이름, 주소, 생년월일이 있으면 개인식별키를 만들 수 있겠지만 비용이 더 비쌀 것입니다.

Q19. PIK를 미리 만드는지, 그리고 만약 식별자료가 이미 만들어져 있다면 연구자가 이를 맞춰서 반입해야 하는지 궁금합니다.

☞ 모든 Census에 적용되는, 미리 만들어놓은 PIK가 있습니다. 연구자들은 PIK에 대해서 모르기 때문에 연구자가 이러한 PIK에 대해 미리 양식을 맞출 수 없을 것입니다.

라. 주요 활동 사진



4. 캐나다 통계청 및 RDC 방문

가. 개요

- 일 시: 2016. 9. 7. (수) 09:00 ~ 15:00
- 장 소: 오타와 캐나다 통계청 25층 컨퍼런스룸
- 참석자:
 - (캐나다 통계청) Jodi-Anne Brzozowski(Director of Microdata Access Division), Donna Dosman(RDC Chief), Nicolas Rahal(국제협력과) 등 5명
 - (연수단) 통계청 및 통계작성기관 22명, 통역 등 2명
- 상세 회의일정

시간	주제	발표자
09:00 - 09:15	환영인사	국제협력과장
09:15 - 09:45	한국의 마이크로데이터 설명	박미숙 사무관
09:45 - 10:30	①마이크로데이터 제공 개요 ②DLI(Data Liberation Initiative) 소개	Jodi- Anne Brzozowski ChantalRipp(마이크로과장)
10:30 - 10:45	휴식	
10:45 - 12:00	③RTRA(실시간 원격서비스)	David Price (마이크로 기반조성팀 Chief)
12:00 - 13:00	점심식사	
13:00 - 14:00	④RDC 소개	Donna Dosman(RDC Chief)
14:00 - 14:45	공공용 자료생성 가이드라인	David Price (마이크로부 기반조성팀 Chief)
14:45 - 15:00	FRDC(연방RDC)* 견학	통계청 1층

※ 당초 오타와대학에 위치한 Cool-RDC를 방문할 계획이었으나 캐나다 통계청의 권유로 FRDC를 방문

- 사전 연구 및 연수자료 정리 : 1분임

나. 캐나다 통계청 마이크로데이터 서비스 개요

(주요업무)

산업부(Industry Canada) 산하기관으로 전형적인 중앙 집중의 강력한 통계 관리 체계를 유지하고 있는 통계청은 인구 및 농업 센서스와 경제, 사회, 문화 등 국가통계 대부분의 마이크로데이터를 관리하고 서비스함

○ 서비스 연혁

- 1971년: 공공용 마이크로데이터(Public Use Microdata File, PUMF) 제공 시작
- 1996년: 공공용 자료제공 협조 네트워크인 'Data Liberation Initiative(DLI)' 시작
- 1998년: DLI가 마이크로데이터 이용시설을 설립할 것을 권고하면서 RDC 서비스를 본격적으로 추진

○ 운영 조직

- 캐나다 통계청에서 마이크로데이터 관리 및 제공을 담당하고 있음
- 통계청에 DLI 커뮤니티 운영을 위한 DLI팀이 있음

○ 제공 자료의 종류

- 캐나다 통계청이 조사한 자료로서 공공용 90여 종, 원격접근 30여 종, RDC를 통한 110여 종을 제공하고 있음

○ 제공 자료의 유형

- 익명화 수준에 따라 크게 공공용(Public Use Microdata Sample, PUMS)과 승인용으로 구분되며, 대부분의 자료는 승인용으로 분류되어 엄격한 승인절차를 통해 제공됨

○ 서비스 이용 대상

- 캐나다 거주 연구자와 캐나다 정책 개발 공무원
- 외국인 연구자는 캐나다 국적 또는 영주권 소지 연구자가 공동 연구자일 경우에만 이용 가능

○ 서비스 유형

- DLI를 통해 공공용 제공
- 실시간 원격접근(RTRA)을 통해 가상자료 제공
- 승인용은 RDC를 통해 보다 상세한 자료 제공

○ 서비스 요금 체계

- PUMF(공공용)는 2011년 초부터 무료로 제공하고 있음. 이용자들은 DLI에 가입한 학교나 도서관을 방문하면 PUMF를 무료로 이용할 수 있음
- DLI 기관은 연간 회비 방식의 정액제 요금을 내야 함
(연간 회비: 연구기관 CA\$12,000, 기타 기관 CA\$3,000)

○ 승인용 자료이용절차

- 연구제안서를 제출받아 통계청에서 심사하고 연구결과물을 검토
<승인용자료 이용절차>



○ 비밀보호 방법

- 이름, 주서 등 개별 식별이 직접 가능한 정보는 미제공
- 교육 년수나 소득과 같이 연속 값을 갖는 정보는 큰 범주로 묶어서 제공

- 극한값은 모아서 하나의 범주로 코딩하는 top-coding
- 가장 상세한 지역 단위는 일부 도시권(metropolitan)이며, 대부분의 경우는 주(province)나 연방 직할지(territories)까지만 제공
- 공유 및 협력
 - 캐나다 통계청은 DLI 네트워크와 RDC 네트워크 등 대학 및 지역도서관 또는 대학 연구기관과 마이크로데이터 서비스를 위하여 긴밀한 협력 관계를 맺고 있음
- 홈페이지: <https://www.statcan.gc.ca/>

다. 연수 세부 내용

① 마이크로데이터 제공 개요

(발표자: Jodi- Anne Brzozowski ChantalRipp, 마이크로데이터과장)

- 자료를 제공하는 캐나다 통계청의 주목적
 - 국가 및 국민의 경제, 사회, 기타 제반 데이터 수집, 정리, 분석, 출간 의무
 - 캐나다 국민에게 중요정보 제공 의무
 - 마이크로데이터 접근 제고와 동시에 기밀자료 보안 노력
- 정부 자료 공개의 중요성
 - 실증적 자료 기반 정책 형성
 - 자료 공개를 통한 정부 청렴성 및 효율성 제고
 - 인간, 기술, 협력 중심의 혁신정책 추진 도움
- 마이크로데이터과(Microdata Access Division)의 업무원칙
 - 기밀정보를 보호하면서 캐나다 통계청의 마이크로데이터 접근을 도움
- 마이크로데이터 접근 원칙
 - 자료 접근의 민주화

- 보안준수 문화 형성
- 협력관계 강화
- 서비스 신뢰도 형성
- 캐나다 마이크로데이터 연구 지원
- **유용성과 노출위험에 따른 마이크로데이터 및 서비스 분류**
 - 기밀데이터 자료
 - 서베이 시행 시 마스터파일에서 개인식별자료 삭제 후 제공
 - 높은 강도의 접근심사 시행
 - 자료 이용 후 노출위험분석을 통한 산출물 분석 후 외부반출 가능
 - RTRA(Real Time Remote Access), RDC, FRDC(Federal RDC), CDER(Canadian Centre for Data Development and Economic Research)를 통해 서비스 제공
 - 공공용 자료
 - 변수 collapsing, suppression, outlier 제거 등 SDC 적용된 자료
 - 변수 합계 및 연결 등을 시도하더라도 개인식별 불가능한 자료
 - DLI, PUMF 접근서비스 통해 제공
- **서비스 성공요인**
 - 이용자 수준 및 수요에 맞춘 서비스 개발 및 제공
 - 정부, 대학 및 연구기관, 사기업 등의 지속적인 이용량 증가
 - 자료제공기관, 연구자들과 강한 협력관계 형성
 - 정량적 조사방법론에 통달한 연구자 증가
 - 정책 관련 연구결과 및 출간물 증가
- **당면 과제**
 - 기술적 문제
 - 금전 등 자원 부족 문제
- **향후 계획**
 - 자료제공기관, 서비스제공기관, 데이터 이용 연구자들과 협력을 통해 마이크로데이터 서비스 제공 기능 향상

- 마이크로데이터 파일 신규 수집, 접근방안 제고 모델 및 지침 형성
- 메타데이터 양식 및 구조 개선
- Tabulation 서비스 제공
- 안전한 원격접근 서비스 구축
- 제공자료 수 및 범주 확대

② DLI(Data Liberation Initiative) 개요

(발표자: Jodi- Anne Brzozowski ChantalRipp, 마이크로데이터과장)

- **DLI란:** 1996년에 시작된 공공용 자료제공 협조 네트워크로서 캐나다 주민들은 DLI에 참여하는 학교, 도서관 등 공공기관을 통해서 별도의 절차 없이 무료로 공공용 PUMF에 접근할 수 있으며, DLI EFT(Electronic File Transfer) 등 데이터 다운로드 및 웹기반 분석 서비스를 이용할 수 있음
- **DLI 관리 및 의사결정 위원회**
 - 외부: DLI 외부 자문위원회
 - 내부: 운영위원회, 마이크로데이터 접근관리 위원회, 마이크로데이터 배포자문위원회, 마이크로데이터 접근에 대한 정책, 마이크로데이터 배포에 대한 권고
- **DLI 개선관리**
 - 자문단 : 각계 대학 교수진으로 구성
 - DLI에 대한 질문 및 개선의견: 매월 50-60개 발생
 - 고등교육기관 협력기관 참여자(DLI관리자): 2015년 133명
- **자료이용 교육**
 - 협력기관들은 세미나나 교육을 통해 데이터 사용방법 및 검색방법 등을 교육하고 있으며, 추후 자료이해도에 대해 검증함

③ 실시간 원격 서비스(Real Time Remote Access, RTRA)

(발표자: David Price, 마이크로데이터 기반조성팀 Chief)

- **RTRA란:** 온라인으로 실시간 데이터 분석 가능하며, 이용자 계정을 통해 접속하여 SAS 프로그램 이용 가능하며, 이용 컴퓨터 단말기가 아닌 서버에 자료가 저장된 상태로 이용
- **RTRA 사용 목적**
 - 복잡하지 않은 연구 방법론이 요구되는 프로젝트에 기술통계학적으로 기여 가능
 - RDC를 이용하기 힘들지만, 공공용 자료보다 높은 유용성 있는 자료 필요 시 용이
 - 공공용 자료 버전이 없는 자료 이용
 - RDC 이용 전 연구타당성 등 검토 가능
- **RTRA의 장점**
 - 인터넷을 통해 언제나 자료를 이용할 수 있음
 - 105개 이상의 데이터셋에 접근할 수 있음
 - 1시간 내에 요청자료의 결과를 수취 가능
 - 자동화된 산출물 기밀보안 체계
 - 1주일이면 사용자계정이 생성됨
 - 프로젝트 심사 불필요
 - 웹 출판 및 리포트용 대규모 table 생성 가능
- **RTRA를 통해 제공하는 행정자료**
 - 범죄리포트, 암등록현황, 중등교육학생자료, 살인자료 등
- **RTRA 결과물로서 수취할 수 있는 자료**
 - Frequencies, Means, medians, Percenties, Ratios, Proportions 등

④ RDC 소개

(발표자: Donna Dosman, RDC Chief)

- RDC 연혁

- 1998년 사회통계에 대한 이용을 위해 Research Data Centers(RDC) 설립 추진하였으며, 2000년 RDC 네트워크가 형성됨
- RDC는 캐나다 통계청, Canadian Institutes for Health Research (CIHR), Social Sciences and Humanities Research Council (SSHRC), Canadian Foundation for Innovation (CFI) 간의 파트너십으로 운영

- RDC란

- 대학 캠퍼스에 위치한 안전한 캐나다 통계청 시설
- 캐나다 통계법 하에 운영
- 캐나다 통계청 마이크로데이터 파일저장
- 캐나다 통계청 직원 관리 하에 운영
- 허가받은 프로젝트 소속연구원들만 서약" 후 "공무원 신분" 취득 후 이용 가능
- 모든 연구자들은 자료에 직접 접근 가능
- 자료 내 모든 개인식별정보 제거

- RDC 프로그램 목적

- 자료의 비밀 보호를 추구함과 동시에 캐나다 통계청의 microdata를 이용한 사회과학 연구의 발전 도모

- 주체별 업무

- 캐나다 통계청
 - 보안기준 설정 및 보안감사 시행
 - 기밀누출방지 반출물 심사시행
 - 자료제공
 - RDC 센터 인력 파견 및 운영

- 유치 기관
 - .. 안전한 연구시설 제공
 - .. 운영비 제공
- 지원 단체
 - .. 대학에 연구보조비 지급

- 캐나다 RDC 현황



- Centre RDC와 Branch RDC 차이

- Center RDCs
 - .. 1주 최소 5일 운영
 - .. 산출물 기밀누출평가 및 연구원의 "공무원 신분" 부여 가능한 RDC Analyst가 운영
 - .. 1-3명의 Analyst 및 몇 명의 Statistical Assistant 상주
- Branch RDCs
 - .. 단축된 운영 시간
 - .. 산출물 기밀누출평가 및 공무원 신분 부여 불가능한 Statistical Assistant 상주

- RDC 데이터 보안

보안요소
<div style="text-align: center;"> <pre> graph TD A[컴퓨터 보안] --> E((보안문화형성)) B[물리적 보안] --> E C[기밀누출심사] --> E D[개인 책임] --> E F[법적 책임] --> E </pre> </div> <p style="text-align: center; background-color: #e0e0e0; padding: 5px;">정기적 보안감리 및 보안감사 실시 통해 보안절차 준수 및 보안체계 확인</p>
물리적 보안
<ul style="list-style-type: none"> · 건축 요구사항: <ul style="list-style-type: none"> · 4면을 감싸는 콘크리트 혹은 철망 벽 · 강철 혹은 강철 프레임 출입문과 고강도 보안시설물 · 외부 창문에 쇠창살 및 센서 · 자동 출입기록이 저장되는 출입 기록 시스템 · 캐나다 통계청 직원 상주 · 센터 운영시간 외 내부 모션센서 작동
컴퓨터 및 네트워크 보안
<ul style="list-style-type: none"> · 로컬 저장기능 없음. 모든 외부 저장매체 접근불가 · RDC 내부 안전한 서버에 데이터 보관 · 보안 WAN 연결에 인터넷 연결 없음 · 연구제안서에 적시된 자료만 Active Directory를 통해 접근가능 · 유치 대학 IT직원들 전원 캐나다 통계청 공무원 신분 부여 · 중앙에서 강력한 보안 프로토콜 유지
반출물 보안심사
<ul style="list-style-type: none"> · 모든 반출물은 RDC Analyst를 통해 기밀누출심사 · Survey data 반출심사 기본적 기준 <ul style="list-style-type: none"> · 최소 cell size · 가중치 적용값 배포가능 · 최소 배포가능 행정지역단위 설정 · 행정자료 심사기준

<ul style="list-style-type: none"> · ·최소 셀 크기 설정 · ·자료별 scoring method나 controlled rounding 적용 · 기밀심사 후 이용자에게 산출물 이메일로 전송
개인적 책임감
<ul style="list-style-type: none"> · RDC 내 스마트폰 포함개인컴퓨터기기사용불가 · 모든 인쇄물은 지정된 색의 용지로 인쇄, RDC에서 반출불가 · 방문자 기록 후 캐나다 통계청 직원 인솔하에 RDC 방문 · 인쇄기록, 계정기록, 보안 등 모든 과정 감시
법적 보호
<ul style="list-style-type: none"> · 통계법 하에 공무원 신분 획득 후 보안등급 획득 및 기밀보호 서약 · 자료 기밀보호 및 conflict of interest가 없음이 명시된 마이크로데이터 연구 계약서 서명 · 자료 노출 시 통계법, 소득세법, 내국소비세법 처벌대상이 됨
보안수준문화 형성
<ul style="list-style-type: none"> · RDC 이용 전 보안 오리엔테이션 실시 · RDC Analyst들의 지속적인 감시 및 이용자 책임 적시 · 산출물 반출 불가 시 RDC Analyst와 지속적인 상담
보안점검 및 감사
<ul style="list-style-type: none"> · RDC 설립 전 물리 및 IT시설 보안점검 · 지속적인 RDC 물리 및 IT시설 보안점검 · 연간 1~3개 RDC 보안감사 실시 · 현재까지 보안감사 결과 긍정적

- RDC 이용신청 과정

- 학술적, 정부소속, 비영리기업, 사기업 소속 연구자 경우
 - · 연구제안서 제출 → 연구제안서의 연구실현가능성과 연구중요성 검토 → 신원조회 → 보안 오리엔테이션 → 공무원 신분 부여 서약 → 연구계약서(행동강령 등 포함) 작성
- 연방, 주, 자치구역 연구자의 경우
 - · 각 지자체별 통계담당 연구공무원 제안서검토 → 통계청 연구제안서 심사 → 신원조회 → 보안 오리엔테이션 → 공무원 신분 부여 서약 → 연구계약서(행동강령 등 포함) 작성

- 외국인 이용신청에 대한 추가적인 요구사항
 - 신원조회를 위한 작업비자와 캐나다 등록 주소지 증빙서
 - 연구 공저자로 캐나다 시민권자 혹은 영주권자 참여의 필요성
 - 외국인 학생의 경우, 소속 대학 증명서 제출
- RDC 제공 데이터
 - 개인식별자료 제거한 자료 제공
 - 500여개 cycle 제공
 - Social Survey data:
 - Longitudinal and cross- sectional
 - Topics include: social, health, income, labour, education
 - Census: 1911 - 2011
 - Administrative data: Vital Statistics, Cancer Registry, Tax, Justice, Provincial Social Assistance data, Provincial Health data
- RDC 내 자료 연계
 - RDC 내 개인식별자료 없음
 - 고위공무원급 허락 하에 자료 연계
 - 연구자가 자료 합치기 가능
 - 통계청 내부 부서에서 연계 가능
 - Data 통합 목적으로 외부 기관 작성 마이크로데이터 반입 가능
- 자료 획득
 - 통계청 내 타 부서와 협조하여 자료 획득 협의
 - 연방정부 고용 및 이민데이터 획득 협의 중
 - 지방정부 보건 등 자료 획득 협의 중
- 시범 제공자료
 - 법무, 국세, 사회보조자료 현재 시범 제공 중
 - 기존에 종료된 종단연구와 국세, 사망, 암 자료 연계 프로젝트 진행 중
 - 입원자료와 인구동태통계 연계
 - Canadian Community Health Survey와 사망데이터베이스 및 입원자료 연계
 - 2006 센서스와 입원자료 연계
 - 1996, 2006 센서스와 출생 데이터베이스 연계

- RDC 이용료

- RDC 유치기관의 학술자들은 이용료 없음
- 정부 및 NGO 연구자들은 22일 이용료로 \$5800 요금을 냄
- 사기업 연구자의 경우 프로젝트 규모 및 연구팀 규모에 따라 이용료 산정
- 사기업 연구자의 이용료에는 데이터 준비 가격, 연구제안서검토, 이용계약서 준비, 보안등급 부여, 계정생성, PC 사용 및 기술적 보조비, 반출물 심사비 등이 포함됨

- 기타 사항

- 제공 데이터셋 형태와 종수가 지속적으로 증가하고 있음
- RDC를 통한 연구프로젝트 수 및 연구자가 지속적으로 증가하고 있음
- 연구 결과물이 지속적으로 증가하고 있음

- 주요 질의 및 토론 사항

Q1. RDC와 Federal RDC의 제공 자료는 다른지

- ☞ 같은 자료를 제공하며, FRDC에서는 특정 선행 연구일 경우 다른 자료를 추가로 제공할 수 있음

Q2. RDC 이용 확대를 위한 홍보 노력은

- ☞ Web 사이트, 트위터 등을 통해 RDC 홍보
- ☞ 공공용자료 이용 지원을 하는 DLI 사서를 통해서도 RDC 홍보를 하게 함
- ☞ RDC간 워크숍, 학술발표대회를 개최함

Q3. 공공용 자료의 경우에도 DLI를 통하지 않는 제공사례가 있는지

- ☞ 있음. DLI에 속하지 않은 개인이 공공용자료를 이용 신청하는 경우가 있으며, 이 경우에는 유료임

Q4. 공공용자료에 대한 연구자 지원 방법은

- ☞ 대학, 지역도서관 등과 네트워크(DLI)로 형성하여 다양한 서비스를 제공하고 있으며, 해당기관 사서(史書)에게 주기적으로 데이터에 대한 활용방법 등을 교육하여 연구자를 가까운 현장에서 지원하도록 함

Q5. DLI 네트워크 체계가 발전한 요인은 무엇인가. 대학, 지역도서관 등이 자발적으로 참여한 것인가 통계청의 권한으로 의무화시킨 것인가. 영토가 넓은 요인도 있는가

☞ DLI 회원 기관의 필요에 의한 자발적인 참여이며, 영토가 넓은 요인도 커다란 요인이 됨

Q6. 지방정부로부터 행정자료 또는 조사자료를 수집하고 있는가

☞ 주 안에 통계청이 있으며, 주 통계청으로부터 행정자료, 조사자료를 수집하고 있으나 협조가 원활하지 않아 공유협력을 이끌기 위한 노력을 하고 있음

Q7 FRDC 견학 시 질문과 답변

☞ 개인전산물품에 대한 휴대 불가, 안에서 사용하는 메모용지 조차 노란색 등 색깔이 강하여 어떻게 활용하는지 즉시 확인할 수 있도록 하여 자료의 유출 방지

☞ 모든 RDC방문자는 방문내역을 기록함. 실제로 연수단도 머문 시간과 방문자 성명을 기록하였으며, Federal RDC 견학을 인솔한 마이크로데이터과장도 본인이 직접 방문대장에 작성함

☞ 현재 FRDC에서 진행 중인 프로젝트는 50~60건임

라. 주요 활동 사진



Ⅲ. 주요 정책적 시사점

가. 마이크로데이터의 사회적 가치와 개인정보 보호 의식

- 세계적으로 정책 및 학술 연구 자료로서의 마이크로데이터의 가치를 공감하고 있었으며, 동시에 자료보안 의식이 상당히 높았음
- 모든 방문기관이 공공용 자료 뿐 아니라 승인용 자료의 제공 체계가 잘 구축되어 있었음
- 조사자료, 행정자료를 이용하여 자료 간 연계·조화한 종적, 횡적 결합 마이크로데이터의 연구가 활발하게 진행되고 있고 이러한 자료의 안전한 제공 체계를 중요시하고 있음

나. 마이크로데이터 제공을 위한 협력 네트워크

- 자료 제공과 연구 활동에 있어 정부 통계청은 대학, 연구기관과 연구기관은 정부 통계청과의 긴밀한 협력 관계를 유지하고 있음
- 자료제공기관은 풍부한 자료를 연구자에게 제공하는 대신 연구자로부터 연구 성과물을 환류 받아 통계자료의 품질을 높여 가며 연구자들을 통계발전을 위한 협력자로 인식하는 문화가 형성되어 있음

다. MPC IPUMS를 통한 국제협력 서비스 사례의 시사점

- 세계 각 국의 인구센서스 자료의 통합제공 사례로써 이는 국내 마이크로데이터 통합서비스를 추진하는데 있어 좋은 협력사례가 됨
- 아주 오래된 과거자료에서부터 최근자료까지를 연계·조화하여 장기간의 인구 변화를 연구할 수 있는 자료를 제공하는 것은 모범 사례로 들 수 있으며, 인구자료와 환경자료, 보건자료 등 타 자료와의 연계 연구도 활발함

라. ICPSR을 통한 국제 아카이브 서비스 사례의 시사점

- 마이크로데이터 뿐 아니라 질적 자료, 역사자료 등 다양한 형태의 자료를 소장하고 있음
- 자료생산 기관 및 연구기관들과 광범위한 자료공유 네트워크, 교육 및 이용자포럼, 데이터 외 부가자료 서비스를 통한 마이크로데이터 활용증진 활동은 한국의 마이크로데이터 통합서비스 개선에 모범 사례가 될 수 있음

마. 미국, 캐나다 사례를 통한 RDC의 미래 발전 가치 및 시사점

- 민감하고 상세한 마이크로데이터의 이용 요구를 만족시키면서 데이터 보안을 유지하기 위한 RDC의 가치를 미국, 캐나다에서는 기 인식하고 오래전부터 RDC를 설립하여 운영해 오고 있음
- RDC는 대학이나 대학에 기반을 둔 데이터 아카이브 기관을 통해 확산되고 있는 추세임. 이는 잠재적 이용자와의 거리를 좁히고 대학의 시설 및 자원, 교육체계, 전문성을 활용할 수 장점이 있어서임
- RDC의 이용절차가 상당히 까다롭고(승인절차 통과한 연구자에게 공무원 신분 부여), 물리적 보안시설, 법·제도적 보안 장치 등이 엄격하여 연구자 스스로가 자료 보안의식을 느낄 수 있는 문화가 형성되어 있음

바. 미국, 캐나다 사례를 통한 분산형과 집중형 통계체계의 서비스 비교 및 시사점

- 미국과 캐나다는 마이크로데이터의 수집과 제공에 있어 증거기반 정책 수립과 경험적인 사회과학의 발전을 위한 정보인프라 수립이라는 보다 종합적이고 장기적인 계획 하에 발전되어 왔음

- 분산형 통계체계인 미국의 자료 관리·제공에 있어서는 센서스국이 노동통계국, 보건통계국 등의 통계자료를 수집하여 통합적으로 제공하는 방향을 보이고 있으며, 집중형 통계체계인 캐나다도 지방 정부에서 생산한 지역통계자료를 연방 통계청에서 수집하여 통합적으로 제공 관리하고 있음
- 미국과 캐나다 모두 자료 제공체계에 있어서는 대학, 연구기관 등을 기반으로 분산화하면서도 상호협력체계를 강화하고 있음
- RDC 운영에 있어서는 먼저 발전한 미국의 모범 사례를 캐나다가 적용하여 발전시키고 있음

사. 해외 연수 참여기관 간 협력관계 증진

- 외국 선진 마이크로데이터 서비스 사례의 사전연구, 기관 방문 및 토의 등을 통해 마이크로데이터 가치와 서비스 증진에 대한 공감대가 형성되었으며, 연수 참여기관과의 유대관계 및 협력관계를 공고히 하였음

아. 해외 우수사례로 본 우리나라의 마이크로데이터 서비스 발전방향

- 우리나라도 공표된 통계에서 얻을 수 없는 세부 마이크로데이터에 대한 수요가 정책·학술연구자, 기업, 개인으로부터 지속적으로 증가되고 있는 상황
- 이용자의 입장에서는 더 다양하고 상세한 자료를 공공용으로 제공받기를 원하고 통계청 및 통계작성기관에서는 민감 정보의 노출 위험도 문제로 이용자의 요구수준에 부응할 수 없는 대립적인 요소가 있음
- 100여개 통계작성기관과 협업하여 추진 중인 264종의 마이크로데이터 통합서비스(MDIS) 프로젝트는 서비스 방법, 서비스 자료 유형,

비밀보호 방법 측면에서 해외 사례와 견줄만한 서비스 기반을 갖추었다고 볼 수 있음

- MDIS 서비스 체계를 기반으로 이용자가 쉽게 접근할 수 있는 공공용자료의 대상과 항목을 지속적으로 발굴·확대해야 함은 물론, 보안자료인 승인용·특수목적용 자료를 이용할 수 있는 원격접근서비스 및 RDC를 체계적으로 발전시키고 활성화해야 하며, 이용자의 개인정보보호 의식도 제고되어야 함
- 또한, 통계청에서는 보안에 더 민감한 행정통계자료, 빅데이터 등을 제공하기 위한 통계데이터센터 설립을 활발하게 추진하고 있는데 통계데이터센터와 RDC를 연계하여 통합적인 시각에서 발전시키면 좋을 것이며, 연구자들의 연구 성과물을 공유하고 환류 받을 수 있는 채널 및 학술네트워크의 활성화 노력이 필요함

□ 미네소타 MPC IPUMS

UNIVERSITY OF MINNESOTA *Twin Cities Campus*

MPC
 Minnesota Population Center
 Robert McCaa, Research Professor
rmccaa@umn.edu

IPUMS-International
 Archiving, Integrating and Disseminating
 the World's Census Microdata
<http://www.ipums.org/international>

50 Willey Hall
 225 19th Avenue South
 Minneapolis, MN 55455
 Tel: 1+952.334.7867
 Fax: 1+ 612-626-8375

KOSTAT technical expert site visit, Thurs. Sept 1, 2016.

Draft Aug 25

Location: 50 Willey Hall, 225 19th Ave. S., West Bank, University of Minnesota

- 9 Virtual Tour and Introduction of MPC (22 visitors) - Dr. Steven Ruggles
- 9:30 KOSTAT Microdata Integrated Services – Ms. Misuk Park
- 10:30 Over-view of MPC integration projects: Steve Master of Ceremonies
 IPUMS-International (Lara), IDHS, CPS, IHIS, ATUS, etc.
- 11 coffee/tea break
- 11:30 Archiving, security of anonymized source and integrated microdata – Fran/Willy
- 12:00 IPUMS-International: Data processing, sampling, anonymization - Matt
- 12:30 IPUMS-International: Integration/evaluation of microdata and metadata - Lara
- 13:00 lunch hosted by MPC at MPC
- 14 IPUMS-International: MOU (Bob), Dissemination (Kristen) and user certification
 - Patt
- 14:30 IPUMS geography and TerraPop - Steve
- 15 coffee/tea break/adjourn

Documents (provided before visit):

- KOSTAT-IPUMS MOU
- Proposed IPUMS integration of high precision samples of Korean censuses, 1960-2015
- MacDonald 2016 external review of IPUMS-International
- UNECE IPUMS-International Case Study
- Journal of Demographic Economics article on IPUMS-International
- Chinese Journal of Sociology: statistical coherence of IPUMS integrated Asian samples
- Privacy in Statistical Databases 2012: confidentiality protections of IPUMS-International

Contact: Robert McCaa, cell 1+952-334-7867 rmccaa@umn.edu skype rmccaa98

□ 미시건대 ICPSR 및 RDC

Statistics Korea
Friday, September 2, 2016
Room G150AB Perry
330 Packard, Ann Arbor, Michigan

ICPSR

9:00 – 9:30 a.m. Meeting with **Margaret Levenstein**, Director, ICPSR
Room 1120 Perry

**per Dr. Kim this is with “two delegates”*

9:30 a.m. – 11:00 a.m. Room G150 AB Perry

- ❖ **Peter Granda**, Associate Director, ICPSR and Director, General Archive
- ❖ **AsmatNoori**, Assistant Director, Computing and Network Services
- ❖ **ArunMathur**, Data Services Specialist
- ❖ **Matthew Richardson**, Computing and Network Services
- ❖ **Justin Noble**, Acquisitions Manager

11:00 a.m. – 12:00 noon **Survey Research Operations**

- ❖ **Stephanie Chardoul**, Director, SRO Proposal Development Group
- ❖ **Patricia Maher**, Director, Survey Research Operations

12:00 - 1:30 p.m. **per Dr. Kim, they are making their own lunch plans**

2:00 p.m. – 4:00 p.m. *Clint Carter, Census RDC.*

- ❖ 3:30 ~ Tour about RDC with two members.
- ❖ Other members will tour about MiCDA (Michigan Center on the Demography of Aging).

□ 캐나다 통계청

**Microdata Service Division of Statistics Korea Study visit
Wednesday, September 7th 2016
RHC 25 A Conference Room**

Wednesday, September 7th 2016	Topics	Presenter (S)
9h00 – 9h15	Welcome/Round table	
9h15 – 9h45	KOSTAT Microdata Service Presentation	
9h45 – 10h30	MAD Presentation 1. Microdata Access programs overview 2. Data Liberation Initiative	1. Jodi-Anne Brzozowski, Director, MAD 2. Chantal Ripp, Unit Head, MAD
10h30 – 10h45	Break	
10h45 – 12h00	MAD Presentation 3. Real Time Remote Access	3. David Price, Chief, Access, Infrastructure and Innovation, MAD
12h00 – 13h00	Lunch	
13h00 – 14h00	MAD Presentation / Discussion 4. Research Data Centres	4. Donna Dosman, Chief, RDC program, MAD
14h00 – 14h45	MAD Presentation / Discussion 5. Data Documentation (DDI)	5. David Price and Chantal Ripp
14h45 – 15h00	Closing remarks	

<질의서> About IPUMS

- Q1. IPUMS-Internatinal은 103개 국가통계기관 및 16개 국제기관의 센서스 자료를 제공하고 있는 것으로 알고 있다. 이 기관들과 자료 제공을 위한 공유 및 협력을 이끌어내기 위해 어떠한 노력을 해 왔는가
- 또한, 현재의 기관 협력도는 만족할 만한가 아니면 협력도를 높이기 위한 추가 노력을 기울여야 하는가
- Q2. 자료의 수집 및 업데이트는 원활하게 이루어지는가
- 마이크로데이터 외 설명정보(조사목적, 조사방법, 조사기간, 샘플링 방법, 가중치, 코드북 등)는 충분히 수집되는가
 - 자료의 주기적 업데이트에 제공기관이 적극 협력하고 있는가
- Q3. 마이크로데이터 및 관련 설명자료 송수신 방법은
- 또한, 자료 보안기술 적용 수준은
- Q4. 수집된 자료의 품질점검 작업을 수행하는가. 그렇다면 품질점검절차는
- Q5. 국제자료 간 레코드 배치, 코딩 구성, 설명자료 문서 등이 상이한 형태일 거라 생각되는데, 균일한 코딩을 적용하고 일관된 형태의 문서로 정리·제공하는 노하우는
- Q6. 시계열적 분석, 데이터간 비교 분석이 가능하도록 데이터들의 조화 (harmonization)하는 노하우는
- Q7. 데이터 및 코드북의 DB모델에 대한 개괄적으로나마 설명해 줄 수 있는 지
- Q8. DDI 기반의 메타시스템과 온라인 통계분석도구인 SDA(Survey Documentation Analysis) 적용 현황을 설명해 주시길

<Questions to the IPUMS>

Q1. It is known that the IPUMS-International project is a collaboration of 103 national statistical offices and 16 international organizations for the dissemination of census microdata.

- Is collaboration among agencies satisfactory? Or should more efforts be made for better collaboration?

Q2. Are data smoothly collected and updated?

- In addition to microdata, are metadata sufficiently collected, including survey purpose, survey method, survey period, sampling, weight, code book, etc.?
- Are data provision agencies are cooperative for periodic data updating?

Q3. How to send and receive microdata and metadata?

- Security methods that are applied when sending and receiving data?

Q4. Does the IPUMS carry out quality control for collected data? If so, the procedures of quality control?

Q5. As for international data, differences in record layout, coding schemes and metadata documents are expected. If so, expertise in consistent coding and coherent documentation?

Q6. Expertise in harmonization for comparative analysis of time-series data?

Q7. (If possible) Overview on the DB model for data and code book

Q8. Application status of DDI-based metadata system and SDA (Survey Documentation Analysis) for online statistical analyses?

<질의서> About ICFSR(미국 미시건대학 소재)

- Q1. 현재 ICFSR에는 전 세계의 대학, 정부기관, 연구기관 등 750여개 기관이 컨소시엄에 참여하고 있는 걸로 알고 있다. 이 기관들과 자료 제공을 위한 공유 및 협력을 이끌어내기 위해 어떠한 노력을 해 왔는가
- Q2. 자료의 수집 및 업데이트는 원활하게 이루어지는가
- 마이크로데이터 외 설명정보(조사목적, 조사방법, 조사기간, 샘플링 방법, 가중치, 코드북 등)는 충분히 수집되는가
 - 자료의 주기적 업데이트에 제공기관이 적극 협력하고 있는가
- Q3. 제공 자료의 유형(On-line(WEB), secure dissemination, VDE, Physical enclave)을 구분하는 표준화된 기준이 마련되어 있는가
- 또한, 서비스 유형 구분과 비밀보호를 위한 익명화처리는 누가 하는가 ex) 자료 제공기관 or ICFSR
- Q4. 외국인이 secure dissemination, VDE, Physical enclave 자료를 이용할 수 있는가
- Q5. 자료 이용 과금체계는
- Q6. ICFSR 데이터 관리 및 서비스시스템(ICT 분야)을 담당하는 인력 규모는
- Q7. 마이크로데이터 및 설명자료의 송수신 방법은
- 또한, 자료 송수신시 적용하고 있는 보안기술은
- Q8. 데이터 및 코드북의 DB모델에 대한 개요 설명을 해 줄 수 있나
- Q9. DDI 기반의 메타시스템과 온라인 통계분석도구인 SDA(Survey Documentation Analysis) 적용 현황을 설명해 주시길

<Questions to the ICPSR (within the University of Michigan)>

- Q1. It is known that 750 universities, governmental agencies and research institutes participate in the ICPSR all over the world. What efforts are made for data sharing and cooperation among agencies?
- Q2. Are data smoothly collected and updated?
- In addition to microdata, are metadata sufficiently collected, including survey purpose, survey method, survey period, sampling, weight, code book, etc.?
 - Are data provision agencies are cooperative for periodic data updating?
- Q3. Did the ICPSR establish standards for the classification of data provision types (On-line(WEB), secure dissemination, VDE, physical enclave)?
- Who is responsible for classifying service types and making data anonymized for confidentiality, for example, data provision agencies or ICPSR?
- Q4. Are foreigners allowed to make access to secure dissemination, VDE or physical enclave?
- Q5. Service charges for data use?
- Q6. Size of the personnel responsible for data management and services of the ICPSR?
- Q7. How to send and receive microdata and metadata?
- Security methods that are applied when sending and receiving data?
- Q8. Overview on the DB model for data and code book
- Q9. Application status of DDI-based metadata system and SDA (Survey Documentation Analysis) for online statistical analyses?

<질의서> About RDC(미국 미시건대학 소재)

- Q1. 센서스국(Bureau of the Census) 산하기관인 CES(Center for Economic Studies)와 협력체계는?
ex) RDC는 CES 소속 공무원이 운영하는 지, 아니면 대학교 등 협력기관이 운영하는 지
- Q2. 미시건 대학에 있는 RDC 운영 인력 규모는
- Q3. RDC를 통해 제공되는 자료의 종류는
- Q4. RDC를 통한 자료 제공의 상세 수준은? 예를 들어, 특별한 경우에 개별 식별이 가능한 이름, 주소, 생년월일 등이 제공되는 경우가 있는 지
- Q5. RDC 이용절차는
- Q6. 외국인이 RDC를 이용할 수 있는가
- Q7. RDC를 통한 자료 이용 과금체계는
- Q8. RDC별로 과금체계가 동일한 지, 아니면 다른 지
- Q9. RDC의 물리적 보안체계는
- Q10. 서로 다른 마이크로데이터 간 연계하여 2차자료 생성, 분석이 가능한지

<Questions to the RDC (within the University of Michigan)>

- Q1. Cooperation status with the CES, which is affiliated with the U.S. Census Bureau?
e.g.) Is the RDC operated by CES officials or by the university?
- Q2. Size of the personnel involved in the operation of RDC within the University of Michigan?
- Q3. What kinds of data are provided by the RDC?
- Q4. The detailed level of data provision through the RDC? For example, does the RDC provide data from which users can identify names, addresses or dates of birth?
- Q5. Procedures for using RDC?
- Q6. Are foreigners are allowed to use the RDC?
- Q7. How to impose charges on the data utilization via RDC?
- Q8. Is there a difference in charging systems among RDCs?
- Q9. Physical security system of the RDC?
- Q10. Is it possible to derive or analyze secondary data by combining different microdata?

<질의서> About 캐나다 통계청

- Q1. 캐나다 통계청은 마이크로데이터 제공을 위해 대학, 지역도서관, 연구기관 등 DLI, RDC 네트워크와 긴밀한 협력관계를 맺고 있는 것으로 알고 있다. 이런 협력 기관들은 기관 자발적으로 참여한 것인가 아니면 의무적으로 참여한 것인가
- Q2. 제공 자료의 유형인 공공용(Public Use Microdata file), 승인용(Confidential microdata file)을 구분하는 표준화된 기준이 마련되어 있는가
- Q3. 자료 제공 방식인 DLI, RDC, RTRA, 기타에 대해 간단한 설명을 부탁드립니다
- Q4. About RDC
- 26개 대학에 RDC가 설치된 것으로 안다. 한 곳의 RDC에서 근무하는 인력 규모는
 - RDC는 통계청 소속 공무원이 운영하는 지, 아니면 대학교 등 협력 기관이 운영하는 지
 - RDC를 통한 자료 제공의 상세 수준은? 예를 들어, 특별한 경우에 개별 식별이 가능한 이름, 주소, 생년월일 등이 제공되는 경우가 있는 지
 - RDC에서 상주하며 연구자들의 자료 분석, 분석결과 검증 등을 도와주는 RDC 직원이 있는가
 - RDC에서는 다른 통계의 자료 간 연계하여 분석할 수 있는 지(연계 Key 제공 여부 등)
 - 외국인이 RDC를 이용할 수 있는 지
- Q5. RTRA 방식에 대해서는 이용자들이 불편함을 느끼지 않나. 그리고 이용률은 어느 정도인가
- Q6. 이용에 대한 과금체계는
- Q7. DDI 기반의 메타시스템과 온라인 통계분석도구인 NESSTAR 적용 현황을 설명해 주시길
- Q8. 데이터 및 코드북의 DB모델에 대한 개요 설명을 해 줄 수 있나

<Questions to Statistics Canada>

- Q1. It is known that Statistics Canada is in a close cooperative relationship with universities, local libraries and research institutes for microdata provision through DLI and RDC network programs. Do these agencies participate in data provision voluntarily or obligatorily?
- Q2. Are there standards to classify public use microdata files and confidential microdata files?
- Q3. Brief explanation of data provision methods including DLI, RDC, RTRA, etc.
- Q4. About RDC
- It is known that RDCs are located in 26 universities. Average size of the personnel per RDC?
 - Are government officials of U.S Bureau of Census responsible for the operation of RDC? Otherwise, are partner agencies including universities responsible for the operation of RDC?
 - The detailed level of data provision through the RDC? For example, does the RDC provide data from which users can identify names, addresses or dates of birth?
 - Are there RDC personnel who stay in the RDC and help researchers to analyze data and verify analysis results?
 - Is it possible to link and analyze data of different statistics in the RDC? (Whether to provide link keys)
 - Are foreigners allowed to use the RDC?
- Q5. Do statistical users regard the RTRA as a convenient system? Utilization rate of the RTRA system?
- Q6. Service charges for data use?
- Q7. Application status of DDI-based metadata system and NESSTAR for online statistical analyses?
- Q8. Overview on the DB model for data and code book