

등록센서스 도입을 위한 양국간 실무자 회의 결과 보고

I

출장개요

□ 출장 목적

- 인구주택총조사의 고비용 구조 개선 및 응답자 응답부담 경감을 위하여 2015년부터 ‘등록센서스’를 도입하도록 검토, 연구 중
 - 추진과정에서 야기될 수 있는 문제점과 제약사항을 해결하기 위한 구체적인 전략과 실천방안 모색이 필요
- 최근 등록센서스를 도입하거나 추진 중에 있는 국가를 방문하여, 온라인을 통해 확보하기 어려운 세부적인 기법 적용 프로세스를 직접 확인하고 구체적인 경험 사례와 실무 노하우를 토론

□ 기간 및 장소

- 기 간 : ‘11. 11. 16(수) ~ 11. 24(목)
- 장 소 : 스페인 통계청, 영국 통계청

□ 출장자 : 등록센서스과 5급 윤미선, 8급 안미숙

II

주요내용

< 스페인통계청 >

1. 2011년 스페인 센서스 일반 개요

□ 일반전략

○ 센서스 자료원(유럽 연합 규정 763/2008)

- 전통적인 센서스
- 전통적인 센서스와 표본 조사 조합
- 등록자료기반 센서스
- 등록자료기반 센서스와 표본조사의 조합(스페인 센서스 방식)
- 등록자료기반 센서스와 표본조사, 전통적인 센서스의 조합
- 등록자료기반 센서스와 전통적인 센서스의 조합
- 롤링센서스

○ 인구 등록대장(Padron)

- 지자체(municipality)마다 1개 존재하며 5년마다 삭제하고 새로 구성하였음
- 1996년 이후 매월 지자체는 인구대장의 변동내역을 통계청으로 발송
- 통계청은 중앙 데이터베이스에 변동내역을 반영하고 품질을 향상
- 불법체류자도 인구대장에 등록해야하며, 등록된 사람들만이 교육, 보건혜택 등을 받을 수 있음
- Padron은 인구 변화를 추적하기 위한 필수적인 자료원이며, 중앙 데이터베이스는 15년 동안 지속적으로 개선되어 왔음
- 센서스를 위한 좋은 기초 자료임

○ 센서스 전략의 주요 요소

- 사전 센서스 파일(FPC) : 행정 및 통계 정보
 - 인구파일 : Padrón를 기초로 다른 행정자료와 연계하여 생성

- 지역(주소)파일 : 2001년 센서스와 인구대장을 기초로 하고, 부동산대장을 활용하여 항목을 추가
- 목적 : 집계될수 있는 레코드수를 결정, 초기 표본 생성, 건축물 센서스의 초기프레임으로 사용, 인구/성별/나이/국적/출생국에 대한 데이터 제공
- 건축물 총조사의 목적은 거처의 프레임 갱신, 지리적 좌표 획득, 건축물의 항목 수집
- 인구 표본조사는 사전 센서스 파일에 없는 항목에 대한 정보를 수집하고 인구집계를 완료하는데 필요한 정보를 획득하기 위함

○ 사회시설 및 노숙자

- 호텔을 제외한 기관들은 전수조사하고 거주자들은 표본조사
- 노숙자는 음식 등 서비스를 제공하는 기관에서 조사

□ 표본 설계 및 데이터 수집

○ 표본설계

- 스페인에서는 각 섹션에 표본단위들이 있으며, 각 섹션에 프레임은 다음과 같다
 - .프레임 A : 충분한 데이터를 포함한 주소를 가진 주거지로 사람이 거주하고 있는 경우는 인터넷 id와 패스워드를 우편으로 발송 (약 13.6%)하고 빈집은 조사원이 조사
 - .프레임 B : 정보가 부족한 주소를 가진 주거지로 건축물센서스를 통해 만들어지며 표본으로 뽑힌 경우 설문지와 인터넷 패스워드를 줌
- 표본에서 지자체의 크기는 조정되는데 이는 프레임 품질이 표본 크기에 영향 받는 것을 피하기 위함

POBLACIÓN

Tramos de Población	Municipios	Población media por municipio	Fracción de muestreo	Muestra media por municipio	Muestra total
menos de 50	399	34	100,0	34	13.720
De 50 a 99	642	74	100,0	74	47.758
De 100 a 199	1197	145	100,0	145	173.148
De 200 a 499	1562	325	63,9	208	324.655
De 500 a 999	1062	711	45,7	325	345.653
De 1.000 a 1.999	926	1.414	26,9	380	352.074
De 2.000 a 4.999	1011	3.150	17,9	564	570.078
De 5.000 a 9.999	564	6.984	13,5	945	532.867
De 10.000 a 19.999	354	14.170	9,1	1.287	455.757
De 20.000 a 49.999	252	29.599	9,2	2.727	687.284
De 50.000 a 99.999	83	71.183	9,3	6.591	547.060
De 100.000 a 199.999	33	144.681	9,3	13.524	446.286
De 200.000 a 499.999	23	271.363	9,3	25.191	579.399
De 500.000 a 999.999	4	697.624	9,4	65.328	261.313
1.000.000 y más	2	2.452.185	9,4	230.186	460.373
TOTAL	8.114		12,3		5.797.425

○ 데이터 수집

- 일정

- .2011년 9월 19일 : 인터넷으로 조사표를 작성할 수 있도록 사용자id와 비밀번호를 우편으로 발송
- .2011년 11월 10일 : 조사표를 우편 발송(인터넷 무응답자)
- .2011년 11월 28일 : 노트북을 이용한 건축물 센서스 실시
- .2011년 12월 중반 - 2012년 3월 : CAPI

- 프레임 A :

- .주거지 : 인터넷조사 -> 우편조사 -> 무응답이면 부표본
- .빈집 : 건축물센서스시 전수조사 -> 무응답이면 CAPI

- 프레임 B(건축물 센서스에서 검출) : 표본으로 선정되면 인터넷과 조사료로 응답할 수 있도록 우편발송하고, 무응답이면 CAPI

□ 주요 개념과 항목

○ 조사범위

- 인구 : 일반적으로 스페인에 살고 있는 사람.
- .조사기준 이전 12개월 동안 거주지에 살고있는 사람.
- .12개월 동안 거주지에 살고 있지는 않았지만 12개월 이상 거

지할 예정인 사람

- 어떤 구체적인 상황

.주중에 다른 장소에서 일을 하고 주말에 집에 오는 사람들은
주말에 있는 집에서 조사.

.떨어져 있는 학생들은 부모님 집에 포함

- 거처종류

.주거전용 거처 : 상시주거지, 빈집, 계절적 주거지, 사회시설들

.주택이외 거처 : 영업용만 있는 건물은 제외되며 1개 이상의
주거전용 거처를 포함 건물이 해당

□ 자료 공표

○ 공표 계획

- 2012년 4분기 : 인구와 인구 구조 (출생의 성별, 나이, 국적, 국가)

- 2013년 4분기 : 전체 공표

○ 공표 방법

- 데이터웨어하우스

- 분석을 위한 대한 지리적 도구 : GIS

- 센서스 그리드

- 센서스 허브

- 정보의 기타 요구

2. 인구등록대장(Population Register)

□ Padron

- 지방 정부에 의해 관리되는 지자체 주민대장

- 국가통계연구소(NSI)에 의해 관리되는 전국의 인구등록대장으로
구성된 데이터베이스

- 스페인의 국가 인구 등록 대장은 행정 관리의 관점에서는 존재하

지 않지만, 8116개의 기초지자체 주민대장은 서로 다른 지역의 자료와 중복을 피하기 위해, 국가통계연구소에 의해 조정이 되어짐

□ 국가통계연구소(NSI)의 국가 인구대장 소유·관리

○ 스페인의 행정 기관

- 17개의 자치주(autonomous community), 52개의 주(province), 8,116개의 기초자치체(municipalities)로 구성

○ 지자체 주민대장(Padron)

- 지자체에 있는 Padron은 일반적으로 지자체에 살고 있는 모든 주민을 포함한 지방 행정 인구 등록 대장이며, 해당 주민의 거주사실과 주소의 증명자료임
- Padron은 1856년부터 시작되었으나, 관리가 어려웠음
- 기초지자체(municipalities)에서 Padron을 만들고, 변경 등 관리
- Padron의 필수 등록 항목은 성명, 성별, 1차 거주지, 국적, 생년월일, 국가 신분카드 또는 외국인 신분증(거류증, 여권)의 개인 식별번호, 교육수준, 선거인 명부를 작성하기 위한 필요 항목 등
- 5년마다 Padron 정비 작업 수행 : 해당 지자체에서 더 이상 살지 않는데 등록대장에 남아있는 사람들에 관련된 정보 디버깅

○ 1996년 법률 개혁

- 5년마다 시행되는 등록대장 정비 작업 제거할 필요성 대두
 - 매우 비싼 작업
 - 대량의 정비작업이 조사원에 의해 작성되거나 시민 스스로가 작성하기 때문에 많은 실수와 중복을 야기함으로써 지역인구등록대장과 선거인명부에 왜곡요소가 됨
- 큰 데이터 파일에 대한 저장 및 관리 기술의 발달
 - 컴퓨터 취급이 용이하고 가격의 하락
 - 지자체가 해당 지역 인구대장 정보를 지속적인 디버깅 과정에 전산화된 관리를 수행하는 경우 증가

○ 새로운 관리시스템의 특징

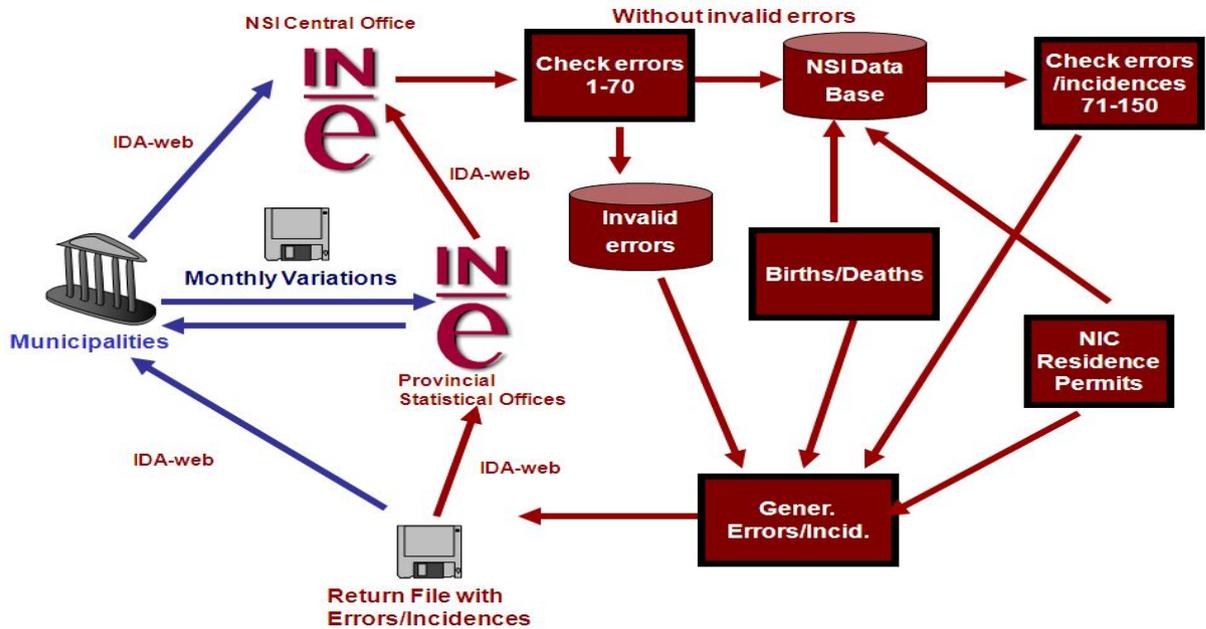
- 모든 지역 레지스터(지자체 주민대장) 관리의 전산화(경제적 및

- 관리 능력이 불충분한 지자체의 경우 주정부가 지원)함으로써 5년마다 수행하던 레지스터 갱신 작업이 없어짐
- NSI는 매월 발생하는 변경내역을 받아, 각 지자체의 인구대장 간의 중복을 방지하기 위해 조정하고 있으며, NSI와 지자체 사이의 모든 정보 교환은 전산화된 교류파일을 사용
 - 인구 등록위원회의 설립
 - 인구 등록위원회는 공식 인구 통계를 알려주고, 지자체 사이 또는 지자체와 NSI 사이의 불일치를 해결, 지역의 인구 등록 관리를 위한 기술적 지침을 제안하며, 지역 및 일반 정부의 대표로 구성
 - 공식 연간 인구 통계의 가용(매년 1월 공표)
 - 해외 스페인 인구 등록대장(RASPR) 생성
 - 스페인에 사는 모든 사람(외국인포함)의 등록은 의무임
 - 외국인의 합법적인 체류 자격과는 별개로 스페인에 거주하는 외국인 등록이 안정적으로 관리됨

□ 새로운 관리시스템

- 1996년 법의 개정으로 지자체에서는 해당 지역 인구 등록 사본 파일을 NSI를 제공했으며, 매월 변동사항을 제공하고 있음
 - 편입(거주지의 변화, 출생, 누락 등)
 - 제거(거주지의 변화, 사망, 중복, 부당한 포함, 만료 (2006 년))
 - 수정(주소 변경, 개인 정보의 변경, 지역 특성 변경 등)
- IDA - Padron 시스템
 - 지자체는 IDA - Padron 시스템을 통해 인터넷으로 NSI 파일에 있는 해당 지역자료를 액세스할 수있음
 - 지자체는 해결해야할 오류와 거리, 시정촌 또는 국가의 사전 보조 정보, 정보를 보내기 위한 표준을 참조할 수 있음
 - 2004년 1월부터 지자체는 매월 인터넷으로 파일 전송 가능

○ 자료처리 흐름도



□ 통계적 활용

- 연간 공식 인구 통계
- 인구의 통계를 사용 등록
- 인구이동 통계
- 인구 센서스 기초자료 (2001, 2011)
- 인구 현황
- 인구 구조 설문 조사 프레임
- 인구 통계학적 중단면 연구

3. 사전 센서스 파일(Pre-census file)

□ 사전 센서스 파일 작성

- 사전 센서스 파일(FPC)은 인구 등록 대장(PADRON)을 기본으로 하고 다른 행정 등록자료와 통계적 작업으로 생성함
- 사전 센서스 파일은 영토 파일과 개인 파일로 구성

□ 사전 센서스 지역 파일(Territorial file)

- 지역파일의 목표는 거처에 사람을 배치하는 것으로, 명부에서 주소를 찾음
 - 명부에서 주소가 찾아지면 사람들을 해당 거처에 할당
 - 명부에서 주소가 발견되지 않으면 새주소를 생성하고 거처에 사람을 할당
- 지역파일 생성 과정
 - 2001년 인구 및 주택 센서스에서 지역구조를 생성
 - 2008년 12월까지의 인구대장으로부터 정보 업데이트
 - 선거인 명부와 연계
 - 2011년 1월 부동산 등록대장과 자동 연계
 - 지방통계청은 지도제작 정보를 사용하여 교차 점검하고 개선

□ 사전 센서스 개인 파일(Personal file)

- 연계과정 : 가장 큰문제는 적합한 연계키를 찾는 것이고, 키워드는 각 사람마다 고유해야 함.
 - 개인 키워드(식별번호)는 스페인 사람은 DNI(국가 ID)이고 외국인은 NIE(외국인 식별번호)임
 - 식별번호를 확인하기 위해 내무부 정보를 이용
 - 식별번호 확인이 어려운 등록자료는 다른 식별 항목(이름, 성, 생년월일, 출생지, 성별, 국적)을 기반으로 확률적 및 결정적 절차를 적용
- 새로운 행정자료 연계 절차
 - 행정자료 항목 연구
 - 데이터의 표준화 및 코딩
 - 개인식별번호는 내무부로 확인
 - 새 행정자료를 인구 등록대장에 연계
- 개인 파일에 들어있는 정보

- 2010년 12월까지 인구등록대장에 포함된 정보(완료)
- 2011년 11월까지 인구등록대장의 정보로 업데이트됨
- 연계정보
 - 1) 2001년 인구 및 주택 센서스 정보
 - 2) 인구동태통계 - 사망(2002~2008년 완료, 2011. 11월까지 갱신)
 - 3) 인구동태통계 - 출생(2002~2008년 완료, 2011. 11월까지 갱신)
 - 4) 사회 보안(2009년 완료, 2010~2011년 갱신)
 - 5) 세금 기관(2008년 완료, 2011. 11월까지 갱신)
 - 6) 기타 소스 : Muface (공무원의 등록) 및 거주 허가
 - 7) 계획 : 교육 레지스터 등

□ 사전 센서스 파일 작성 목적

- 특정 등록자료가 센서스 인구에 포함 여부를 결정하기 위해, 각 사람으로부터 정보 획득 (의심, 계수조정요인, 인구 통계)
- 센서스 조사 항목의 직접적인 정보 제공
- 표본조사에 대상이 될 사람과 거처를 선택하는 초기 프레임으로 사용
- 건축물센서스의 초기 명부(디렉터리)로 사용
- 데이터 처리에 대한 추가 정보 제공(에디팅, 임퓨테이션)

4. 2011년 건축물센서스와 인구조사

□ 2011년 센서스의 목표

- 인구의 기본 구조 보급
- 건물, 주거, 사람에 대한 특성 획득
- 기초 지자체에서 이용가능한 정보 생성
- 지리적 참조 정보
- 사람과 관련된 행정 등록자료의 강한 사용

- 지역 디렉토리
- 국민의 협력을 보다 쉽게만들기
- 조사원에 의한 자료 수집 양 줄이기
- 정보 보급을 위한 다양한 방법 추진

□ 사전 센서스 파일

○ 개요

- 행정 등록 자료 사용 : 인구대장과 다른 등록 대장 간의 결합
- 참조 날짜 : 2011년 11월 1일(조사기준일)
- 목표 : 조사 항목에 대한 정보와 건축물센서스 초기 프레임, 인구표 본조사의 초기 프레임을 확보하고, 조사된 자료의 편집과 보정을 쉽게 하기 위함

○ 사전 센서스 지역 파일

- 초기 포인트 : 2001년 센서스 정보 (새로운 건물의 정보 부재)
- 인구대장 정보로 업데이트(문제점 : 사람이 살고있는 거처만 포함됨)
- 스페인통계청 정보와 토지대장과의 연계를 통해 건축물의 좌표 정보, 조사항목 정보뿐만 아니라 더 많은 주택들을 보완

○ 사전 센서스 개인 파일

- 인구대장(PADRON)을 기본으로 하고 개인키를 사용하여 다른 행정자료(사회 보장, 출생 및 사망, 세금 관리, 교육부 등)와 결합
- 결합 작업 중 발생하는 문제점에 대해 내부 절차를 개발

□ 건축물 센서스

- 조사기간 및 조사원 : 2011년 11~12월 / 4,000 조사원(전수조사)
- 건축물 센서스는 주택 및 그 건물 내부에 있는 주거지(가구)를 파악하고, 건물의 특성 확보하기 위함(휴대용 전산장비를 이용한 전자조사표 사용 -> 유료 자료 사용)
- 건축물 센서스를 통해 발견된 주택은 별도 표본을 선정하여 인구조사 실시

□ 인구 표본조사

○ 개요

- 조사기간 : 2011년 9월부터 2012년 3월까지
- 목표 : 전체 인구 추정, 상세 수준의 인구와 거처의 특성 추정
- 각 센서스 섹션에서 표본 추출

○ 인력 자원

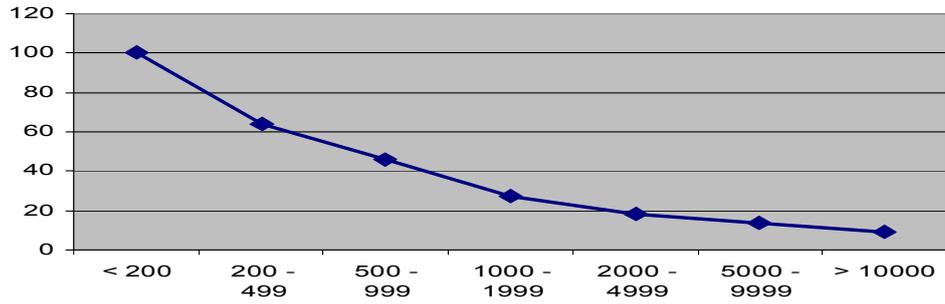
- 조사원(4,000)
 - .건물 센서스 : 2011년 11월 - 12월
 - .인구 조사 (CAPI) : 2012년 1월에서 2012년 3월까지
- 그룹 코디네이터 (1,000) : 조사원 관리 및 감독
- 현지 코디네이터, 현지 검사관, 지역 검사관, 글로벌 검사관

○ 주택 단위 프레임

- 사전 센서스 파일[인구대장, 토지대장, 다른 행정자료들]
- 사람과 가구에 관한 정보
- 사전 센서스 파일에서 주소가 확인된 거처(타입 A)와 건축물센서스를 통해 확인된 신규 거처(타입 B)로 구별

○ 인구표본조사의 구성

- 프레임 A (표본조사 전에 설정)
 - .거처용도인 건축물 단위로 주 거처로 판단된 거처에 대해서는 인터넷이나 종이조사표를 우편 발송하여 조사하나, 무응답시 CAPI 조사
 - .계절주거지, 2차 주거지, 빈집의 경우 전수조사하나, 미해결시 CAPI 조사
- 프레임 B (인구 조사 중 설정)
 - .사전 센서스 파일의 새로운 주거지 중 랜덤하게 선택하여, 조사원이 주택의 종류를 결정
- 표본크기 : 약 3백만의 거처(11.9%), 6백만 명 (12.3%)
 - .기초자치체 수준의 정보획득, 가용할 수 있는 예산 고려
- 샘플추출률은 지자체 규모에 따라 달라짐



○ 센서스 항목

	핵심항목	전체항목
개인	15	42
가족	6	11
주거	12	13
건물	2	15

□ 집단시설과 노숙자

- 독립적인 신규 조사 실시 (10종류의 기관보호시설)
- 2012년 1월에서 3월까지
- 프레임의 정의
- 집단시설에 관련된 질문과 그곳에 살고있는 사람들에 관한 질문
- 노숙자를 위한 특별 조사 실시

□ 아웃소싱

- 도움말 데스크
- 데이터 수집을 위한 IT 시스템
- 인구조사 조사표 스캐닝 및 에디팅
- 우편 서비스
- 자료 수집을 촉진하기 위한 교육적인 캠페인
- 조사원 교육
- 개인 센서스 조사표 인쇄 및 배달
- 홍보 캠페인

□ 결과공표

○ 목표

- 사용 가능한 정보의 양을 (통계 공개) 극대화
- 지리 정보의 활용
- 공표 기간 단축

○ 공표범위

- 2001년과 유사한 공표
- 건축물 관련 항목 - 전체 공개
- 주거 및 개인 관련 항목
 - .사전 센서스 파일 항목 - 전체 공개
 - .표본조사 항목 - 표본오차가 허용되는 경우에만 정보를 보급

○ 공표수단

- 데이터웨어하우스 : 친밀성, 온 디맨드, 통계 공개
- 지리 정보 시스템 : 테마 지도, 센서스 그리드, 개인 요청

○ 공표시기

- 인구 통계 , 인구의 기본 특징: 2012년 말
- 상세 정보 및 마이크로데이터 : 2013년

5. 인구 조사(표본)

□ 2011년 인구주택총조사의 핵심요소

- 인구등록시스템을 핵심으로 하여, 입수 가능한 행정자료를 최대한 이용하여 만들어진 “센서스 사전 파일(pre-census file)”
- 두 가지의 규모가 큰 현지실사:
 - 모든 건축물의 지리 좌표화(geo-referencing)와 건축물 특징의 확인을 가능하게 하게 하는 건축물 센서스(Building Census)
 - 사람과 주거의 특성을 확인하기 위하여, 모집단의 구성요소를 상당히 높은 비율로 추출하여 실시하는 대규모 표본조사

□ 표본조사

○ 목적

- 전체 인구수 추정
- 인구와 거처의 특성 추정

○ 계수요인(Count factors)

- 인구등록대장과 여러 행정자료(인구동태통계, 사회보장, 조세자료)의 정보에 따라, 사람들을 분류
 - 품질 양호하면 계수요인 1로 할당
 - 품질 양호하지 않으면 계수요인 0으로 할당
 - 결정하기 불가능하면 계수요인 Blank로 처리
- 계수요인이 Blank인 그룹에 대한 계수요인을 할당 과정
 - 해당 레코드들은 클러스터로 그룹화
 - 표본조사 결과로 각 클러스터에서 인구추정
 - 각 레코드의 계수요인은 이 추정값과 사전센서스파일에 있는 해당 클러스터의 레코드수의 비율로 정의됨

□ 표본 프레임

○ 각 기초지자체에서 표본 선정을 목적으로 두개의 프레임 사용

- 프레임 A(Framework A) : 사전 센서스 파일에서 정확한 우편주소가 확인되는 거처
- 프레임 B(Framework B) : 프레임A에 포함되지 않는 거처

○ 표본은 각 센서스 섹션안에서 이용 가능한 프레임에 따라 랜덤하게 선정됨

□ 표본

○ 프레임A에서의 표본(사전센서스파일에서 정확한 우편주소를 가진 거처)

- 표본선정 : 무작위 계통추출
- 주택용도의 만들어진 주거처와 2차주거지나 빈집 등 주거처가 아닌 거처로 표본이 만들어짐

- 프레임B에서의 표본(우편주소가 정확하지 않음)
 - 프레임B 현지실사를 전수조사하면서 등록된 거처로 구성
 - 표본선정 : 동등한 추출률을 가지고 베르누이 절차를 사용
 - 임의 샘플크기를 갖지만 규모는 큼
 - 조사원은 프레임A안에 포함되지 않은 거처를 등록하고 샘플 선정 여부를 결정할 수 있음

□ 표본 크기

- 약 3백만 거처와 6백만 인구를 표본으로 선정
 - 기초지자체 수준의 추정을 가능하도록 고려
 - 이용가능한 예산 고려
- 약 인구 2만명인 기초지자체의 특정치가 10%인 변수의 추정치는 7.7%의 변이계수를 가져야함
- 효율적 표본설계를 위해 주거처가 아닌 건축물의 추출률은 프레임A는 40%, 프레임B는 60%로 하며, 200명보다 작은 지자체는 전수를 샘플로 함

□ 추정

- 승수 불편추정량
 - 표본 프레임, 표본추출률, 하위표본에 따라 표본에 포함될 확률
- 무응답의 보정
 - 보정은 표본과 지리적으로 분리된 다른 수준에서 이용가능한 표본과의 이론적관계 기반임
 - 지자체 수준에서 보조변수로 가구의 총수를 가지고 비추정 사용
- 보조정보(사전 센서스 파일)를 활용한 캘리브레이션
 - 캘리브레이션 변수 : 성, 연령, 국적
 - 지리적 분해 : 기초지자체(Municipal), 주(Provincial), 자치체(Autonomous community), 전국(National)
 - 외적 소스 : 계수요인에 의해 보정된 사전센서스파일

6. 정보 보안

□ 스페인 국가 보안 프레임 워크 (ENS)

○ ENS의 목적

- 정보시스템의 안전을 보장하기 위한 조치를 구현함으로써 전자적 수단의 사용에 있어 신뢰를 높일 수 있는 필요한 조건들을 만드는 것
- 행정에서 시민들이 자신의 권리를 행사할 수 있게 하는 것
- 스페인 국가 보안 프레임 워크는 시민들이 행정과의 관계에서 사용되는 정보시스템을 보호

○ ENS의 요소

- 보안 6원칙 :필수 보안, 위험 관리, 예방.반응 및 복구, 심도 방어, 정기 재평가, 직무의 분리
- 15 최소 보안 요구 사항
- 분류 : 어떤 보안사고가 정보시스템에 주는 영향 정도에 따라 기본, 중급, 고급으로 분류
- 보안 조치
 - .조직 프레임 워크 -> 국제적 조직
 - .운영 프레임 워크 -> 운영 체제
 - .보호 조치 -> 특정 자산을 보호
- 감사

□ 기관 보안 전략

○ 2011년 센서스의 정보시스템

- CAWI : 인터넷을 통해 시민이 접근
- CATI : 시민들이 센서스에 대한 정보를 받음
- CAPI : 조사원이 노트북을 가지고 직접 조사

○ 정보보안을 위한 조치

- 정보 분류 : 정보에 대한 책임과 서비스에 대한 책임
- 위험 분석

- .보안 요구 사항
- .필수 보안 문서의 보안 요구 사항
- .감사 : 총조사 시스템 개발업체에 대한 외부 감사 필요, 서로 다른 3개의 감사업체들을 선택
- .서비스 수준 계약 (SLA) : 서비스 수준과 요구 사항을 관리하기 위해 필요하며, SLA와 맞지 않을 경우에는 처벌
 - 용역업체에 대한 계약

○ 용역업체 계약과 관련된 기관 보안 전략

- 기술 계약 체결
- .필수 법제도
- .통계청과의 협상
- .보안 요구 사항 및 필수 보안 문서
- .서비스 수준 계약
 - 관리 계약 체결 : 기술 계약 체결에서 용역업체에 요청된 정보를 평가하는 방법에 대한 세부 사항을 다룸

□ 모니터링

○ 모니터링을 위한 조치

- 계약을 체결 후 필수 보안 문서
- 보안위원회의 정의
- 감사의 결과를 분석, 평가 및 구현

○ 계약을 체결 후 필수 보안 문서

- 감사보고
- 기밀 계약
- 보안 규정 및 절차
- 인증 과정
- 정기적인 테스트 결과
- 보안 사건의 기록

○ 보안위원회의 역할

- 보안 요구 사항을 보장

- 합작 회사를 대신하여 보안에 대한 책임
- 보안 사고와 충돌 처리
- 보안 확인 회의
- 보안 문서 보고
- 보안 상태 보고

○ 감사

- 웹 감사 (CAWI), 노트북 감사 (CAPI), 데이터 서버 감사, 종이 감사 (PAPI)

□ 보안에 대한 권고사항

- 정보 분류
- 위험 분석 및 관리, 방법론
- 책임 지정 (정보, 보안 ...)
- 감사
- 보안 의식
- 인증

< 영국통계청 >

1. 인구추정을 위한 방법론

□ 2011년 센서스 개요

○ 목적

- 품질이 우수한 정확한 추정인구 및 인구학적 특성 제공
 - 전국 추정인구는 95%신뢰구간에서 $\pm 0.2\%$ 신뢰도
 - 모든 지역수준의 인구추정은 95%신뢰구간에서 $\pm 0.32\%$ 신뢰도 (전국센서스 응답률 94%이상, 지역수준의 응답률 80% 이상, 지역과 인구그룹들 사이의 감소된 변동성, 센서스 사후조사 90%의 응답)
- 이용자 요구 충족 측면 : 이용자 요구사항을 충족하고 결과(결과의

- 질, 적시성, 접근성과 이해도, 메타데이터와 같은 지원정보 등을 포함)에 대한 자신감을 보장하는 산출물과 전달 메커니즘을 제공
- 보안과 기밀 측면 : 센서스에 포함된 개인정보 보호(독립적인 검토는 보안 및 기밀을 보장)
- 돈의 가치 측면: 투자에 대한 가치 제공(독립적 검토는 센서스가 예산범위 내에서 식별된 혜택을 전달하는 것을 확인)

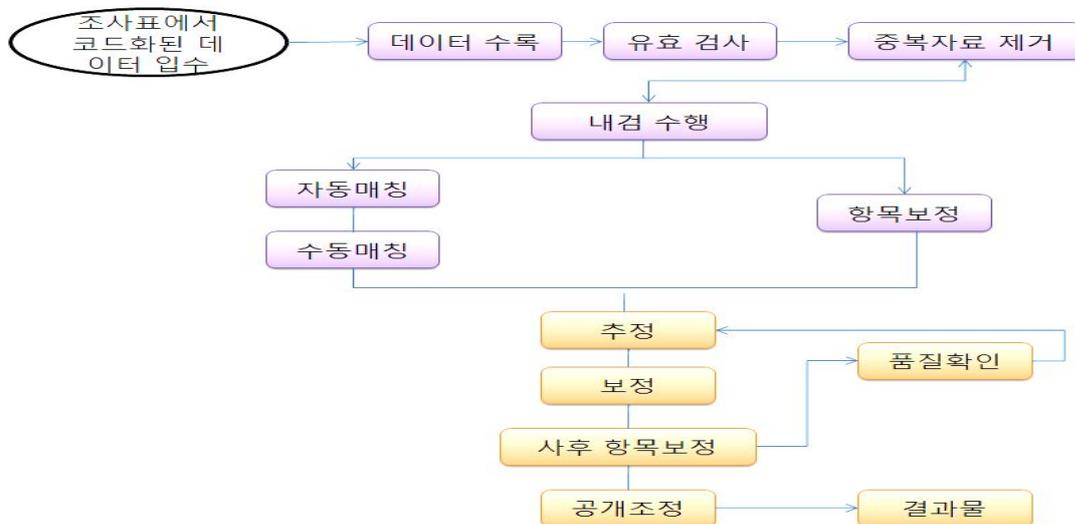
○ 센서스가 갖고 있는 문제점

- 사회와 인구의 변화(인구이동, 다양화되고 고령화되는 인구)
- 인구조사에 대한 국민적 반감 증가
- 자료 기밀에 대한 의문들
- 전국에 걸친 응답의 차이
- 경제침체에 따른 국민들 정서
- 새정부의 지출 감소
- 정부의 광범위한 마케팅 및 채용 동결

○ 2011년 센서스의 주요 혁신

- 주소등록 대장(Address Register)
- 새로운 요구사항을 반영한 질문지
- 조사표의 우편 발송 및 바코드 이용한 조사표 추적
- 혁신적인 직원 및 유연한 현장 인력
- 인터넷 완성

□ 통계적 절차



□ 인구 추정

- 커버리지 평가(Coverage Assessment)와 품질확인
 - 목적 : 지방정부 단위의 성별 연령별 인구추정치를 얻고, 추정치에 대한 고품질을 보장하기 위함
 - 방법 : 센서스 범위조사 실시, 추정, 품질확인
- 커버리지 평가(Coverage Assessment) 주요 단계
 - 개정된 센서스범위조사 설계와 표본 할당
 - 개선된 추정 프로시저
 - 기본 가정이 실제적이지 않은 추정치를 보정하기 위한 부가적 방법 추가(이원시스템 추정법의 편향 보정)
 - 센서스 데이터베이스를 보정하기 위한 방법들 재검토
- 집계 난이도(Hard To Count) 지표
 - 주어진 지역의 가구들이 응답할 가능성
 - 2001년에 열악한 응답과 결합된 요소(민족, 실업, 연령, 주택밀집도)들의 현재 값을 기준
 - 응답 가능성 정도에 따라 집계난이도를 5개 집단으로 분류
- 센서스 범위조사 - 표본선정
 - (Reminder) : 작은지역들의 독립적 표본조사(우편번호), 어떠한 센서스정보나 주소명부를 사용하지 않음, 방문조사(13개 항목), 신생아 등 누락된 인구의 대한 질의
 - 5,800 집계 지역에서 18,000 우편번호 표본 추출(340,000 가구) : 2001년보다 약간 큰 규모임
 - 집계난이도에 따라 각 지방정부 단위의 표본 추출, 응답률이 작은 지역은 더 많은 표본 추출

집계난이도(분포도)	2011	2001
1(40%-Easiest)	2.0%	3.4%
2(40%)	3.0%	3.7%
3(10%)	4.8%	4.5%
4(8%)	6.2%	
5(2%-Hardest)	7.2%	

○ 센서스와 사후범위조사 매칭

- 추정은 이원시스템 추정법을 기본으로 함
- 개별수준 매칭 필요
 - 가구 및 개인 단위
 - 센서스나 사후범위조사에 누락 되거나 중복된 사람들 식별
 - 정확성이 중요
 - 누락된 매치 최소화 요구됨
- 매칭 작업 : 정밀매칭과 확률매칭(자동) => 수동검토 => 수동매칭 => 품질확인
- 매칭후 데이터(약 5,500 군락의 표본지역)
 - 센서스와 CCS에서 모두 있는 경우(매치됨), 센서스에만 존재하는 경우, CCS에서만 존재하는 경우의 가구 및 인구 데이터를 보유
 - 모든 조합의 매치는 범위 추정 및 보정에 사용됨

○ 추정

- 이원시스템 추정
 - 표본지역의 실제 인구 추정
 - 양 조사(센서스와 CCS)에서 모두 누락된 것에 대한 보정
 - 성별/연령별 군락에 대하여 적용
- 비추정
 - 표본으로 뽑히지 않은 지역에 대한 추정
 - 이원시스템 추정과 센서스 결과 사이의 비에 의해 커버리지 비율을 획득
 - 인구추정은 전체 센서스 결과에 해당 비율 적용
- 지방정부 수준의 추정
 - 지방정부(LA) 수준의 추정치를 얻기 위해 추정지역(EA) 수준에 집계난이도 팩턴에 의한 성-연령별 인구 추정치 사용

○ 추정에서 품질확인 단계로 이동

- 추정은 신뢰구간 안에서 지방정부 단위의 성.연령별 인구 추정치 생산

- 보정작업은 가구, 개인별로 보정함
 - 보정은 누락된 사람들의 다른 특성들을 제공(사후범위조사에서 조사되지 않은 변수들에 대해)
 - 모든 집계지역이나 모든 주소에 사는 사람들의 정확한 수를 얻는 것이 목적이 아님
- 보정작업이 끝난 성.연령별 인구는 품질확인 작업으로 넘어감

□ 품질 확인

○핵심 목표

- 방법론이 정확하게 적용되었는지 검사
- 계통오차의 증거가 없는지 검사
- 비교자료에 대해 가장 큰 불일치를 가진 지방정부 식별
- 부가적 분석이 필요성 파악
- 최상의 인구 추정치 생산

○주요 단계

- 지방정부 수준에서 좀더 광범위한 핵심 품질확인
- 부가적 작업 착수 시점에 대해 명확성
- 결과 공표전에 부가적 품질확인
- 요구시 사용할 수 있도록 검토되고 준비된 개선 방법들
- 최근의 전국적 추정치와 지방정부 수준의 추정치의 중요성 인식
- 추가적 품질확인 패널들 : 높은 수준의 패널은 외부 전문가를 포함

○품질확인 프레임워크

레벨	핵심체크	부가체크	개선
지방정부 (LA)	.다른 자료원 비교 .작성정보 .LA에서 제공한 증거	.낮은 지리적 수준에서 비교 .행정자료와 매칭	.외부자료원에 의한 보정 .커버리지 추정에 대한 조정
주정부	.다른 자료원 비교 .작성정보 .LA에서 제공한 증거	.부차적 주거지를 상용한 매칭 .센서스 무응답 연계 연구에 대한 평가	.외부자료원에 의한 보정 .커버리지 추정에 대한 조정 .센서스 무응답 연계 연구를 사용한 조정
잉글랜드/ 웨일즈	.다른 자료원 비교 .작성정보 .LA에서 제공한 증거	.부차적 주거지를 상용한 매칭 .센서스 무응답 연계 연구에 대한 평가 .종단적 연구 평가	.외부자료원에 의한 보정 .커버리지 추정에 대한 조정 .센서스 무응답 연계 연구를 사용한 조정 .종단적 연구를 사용한 조정

○핵심 품질확인 체크

- 핵심 점검에 따라 모든 지방정부 센서스 인구추청
.대체 가능한 추정치 비교
.인구학적 분석
.기능적 체크
- 비교 데이터를 통해 확실한 경우 체크에 가장 높은 가중치 부여
- 2011년 7월에 인쇄된 논문은 부가적 분석이 이 가중치를 이용하여 실행되어질때 시작됨

○핵심 품질확인 체크 - 자료원

- 성-연령별 LA 추정치를 위한 자료원 : 환자 기록부, 중간년도 추정치, 학교센서스, 육아수당, CIS 보조금
- 가구를 위한 자료원 : 주소등록대장(우편 센서스), VOA, 지방정부가 제공한 지방세
- 기타 다른 인구 추정치를 위한 자료원 : HESA(학생들), MoJ(수감자), DASA(군인)

○부가 품질확인 체크

- 핵심 체크의 결과에서 발견, 일반적으로 지리적으로 지방정부 아래 수준에서 수행
- 가구 단위의 분석
.정제된 환자 기록부의 가구수와 비교
.지방정부에서 제공된 지방세 자료와 비교
- 개별 단위의 분석
.비교시 단기 이민자를 포함
.정제된 환자 기록의 인구수와 비교
- 모든 부가 분석이 누락된 사람들을 나타내는 것은 아님
- 2차 주거지, 단기 이민자 체크는 행정자료에 있다고 하더라도 최종결과에 포함되지 않는 사람들을 식별함

○센서스 인구 추정치에 대한 동의

- 내부 품질확인 운영위원회
.주 품질확인 패널로 가는 관문

- .상세한 부분을 검토하는 소규모 실무자 그룹
- 주 품질확인 패널
 - .지방정부 센서스 인구 추정치 와 증거 검토
 - .통계청 전문가들
- 고위 품질확인 패널
 - .새로 만들어진 전구 및 지역 추정치 검토
 - .독립적 학문과 사용자의 통찰력
 - .주 품질확인 패널로부터 언급된 지방정부 평가

○ 품질확인 패널 프로세스



○ 인구추정을 위한 센서스 설계

- 센서스는 모든 지방정부 수준에서 비슷한 고품질 추정치를 일관성있게 생산하도록 설계됨
- 인구수를 집계하기 어려운 사람들에 대한 자료원
 - .주소등록대장
 - .현장 자원 할당
 - .지역사회와의 의사소통, 지방정부와의 의사소통
 - .사후범위조사 표본설계
 - .품질확인 전략
- 상호관계된 작업 및 과정들은 성.연령별 추정치 생산을 위해 함께 움직임
- 품질확인 과정은 모든 것들이 함께 잘 작동되었는지 평가

- 만약 작동되지 않은 증거가 있으면 이것을 식별하고 보정하기 위한 과정이 필요

○ 인구추정 요약

- 지방 정부 수준에서 비슷한 품질을 갖은 성.연령별 추정치를 획득하는 것이 목표
- 추정 과정에서 단일단계로 모든 일을 하는 경우는 없음
.단계의 조합은 필요한 품질을 제공
.지방정부의 총수와 주요 특성은 자원할당에 대해 질적으로 충분함

2. 주소등록대장(The Address Register)

□ 2011년 센서스에서 주소등록대장이 중요한 이유

- 조사표의 95%를 전달하기 위해 우편 발송
 - 조사원 채용의 위험 감소
 - 불응을 대상으로 절감예산 허용
- 조사표의 추적
 - 현장 기능의 관리와 조정을 개선
- 유연성 있는 현장 인력
 - 좀 더 효과적으로 자원을 추적
- 센서스 추정치의 품질보증을 위한 핵심 자료

□ 주소등록대장의 목적

- 이용할 수 있는 자원과 현장조사 작업으로부터 가구들이 최상의 품질의 등록대장 생성
- 핵심 이해당사자, 특히 지방정부들이 자신 있어 하는 등록대장
- 예산대비 가장 가치 있는 등록대장(99%이상의 완성과 1%미만의 중복)

*센서스를 위한 명부는 일반적인 공공사용 목적이 아님

□ 주소 확인을 위한 기준

- 우편주소화일(PAF, Postcode Address File)과 전국부동산주소목록(NLPG, National Land & Property Gazetteer)의 차이(10%)
 - 주소 복잡성에 대한 대응
- 다양한 meters/ 주소들내 선거역활에 따른 다양한 이름들(4%)
 - 복잡한 구조의 전국적 대응
- 성층 표본 추출법(1%)

□ 주소등록대장 작성을 위한 주요 작업

- 매치 및 유효성
 - 전국 주소 파일들의 매치
 - 미해결된 주소는 공급자나 지방정부에 보내짐
 - 통계청은 누락 및 유효하지 않은 데이터를 잡기 위해 E&W의 15%를 현장 조사
- 데이터의 최신화
 - 변화가 있을때 계속적 매치(우편주소화일과 부동산대장)
 - 늦은 가로채기
 - 지방부동산주소목록(Local Land & Property Gazetteer)의 갱신, 분류, 개선 등 매우 중요
- 연구, 데이터 정제, 다른 자료원들
 - 사무 프로세스 : 데스크탑 비교, GIS, 인터넷, 항공사진
 - 자원 자료원 : 빈집, 선거인 명부, 수급계기, HMOs

□ 집단시설

- 우선순위 : 인구, 단체의 크기, 대상 인구(집계난이도)

- 종류
 - 복잡 : 대학, 감옥(100% 현장조사)
 - 표준 : 병원, 호텔 등 (100% 현장조사)
 - 단순 : 호텔, B&Bs(전화조사)
- 자료추적과 현장조사 계획
- 지방정부 입력/품질보증의 기회

□ 요약

- 센서스의 새로운 접근방식
- 주소 등록 대장의 생성 : 여러 데이터셋의 매칭, 현장에서 대상 체크, 기준과 규칙의 증거로 사용
- 집단시설 : 별도의 처리, 학생들 다른 처리
- 성공적인 결과 예상 : 과다, 과소 커버리지에 대한 목표점

3. 2011년 이후 인구통계의 미래

□ 2011년 이후 프로그램

- 변해야하는 이유 : 빠른 사회변화, 이용자 요구사항 진화, 자료공유 등 새로운 기회, 전통적인 센서스에 대한 고민(비용, 횡수)
- 프로그램 목적
 - 장래에 소지역 인구와 사회인구학적 통계들을 제공하기위한 최상의 방법 식별
 - 2014년 9월에 권고사항(비용편익분석과 구현을 위한 높은 수준의 설계에 의해 뒷받침된)제공
 - 합의에 따라 권고사항 구현

□ 2011년 이후 통계적 선택들

○ 잠재적 자료원

- 인구데이터(국민건강보험의 환자기록, 노동연금부 등 고객 정보, 18세 이상의 선거인명부, 학교 센서스 자료(5-16세), 고등교육 통계기관 자료, 출생 및 사망자료, 자동차운전면허 기록 등)
- 사회인구학적 자료원(표본조사들, 자동차운전면허 기록, 사기업 소스들, 전기 및 가스 자료, TV시청 등)

□ 행정자료를 활용한 센서스 대안의 핵심 위험 요소

- 국민여론 : 공공의 수용가능성 연구, 공개적 자문 및 투명한 접근
- 기술적(최첨단, 큰변화) 한계 : 고품질의 팀 구성, 외부 보증이 되는 전문가와의 대화
- 행정자료의 변화 : 탄력적인 솔루션 구축(입법화가 필요할 수 있음)
- 의사결정 보장(정치적 동의 얻기) : 센서스를 행정자료로 대체하는 것은 중대한 결정이며, 실제적 노력이 필요함

□ 프로그램 주요 원칙

- 공개적, 혁신적이며 질 높은 연구
- 이용자와 다른 이해관계자들과 지속적인 대화
- 분명하고 명확한 기준 - 시간이 지나면서 정제
- 통계적 옵션과 방식에 대한 반복적인 평가
- 광범위한 독립적 보증(학계, 전문가 등)

4. 정부의 행정자료 공유

□ 소개

○ 행정자료의 보다 효율적 사용 및 자료 공유의 중요성

- 응답자의 부담 및 비용
- 새로운 통계 및 기존 통계의 개선에 대한 수요
- 표본 조사 및 센서스의 응답률
- 통계청 비용과 효율성

□ 데이터 공유 계획 및 우선순위

○ 공유 데이터에 대한 주요 요구 사항

- 핵심 우선순위 식별
- 작업에 대한 동의 프로그램 개발
- 비밀 보호
- 윤리 문제의 처리
- 투명성 확보

○ 데이터 공유 우선순위

- 잉글랜드의 학교 센서스
- 웨일즈의 학교 센서스
- 이주 노동자 스캔
- 학생 데이터
- 세금 및 정부수당 데이터 (진행)

□ 행정자료 이용의 장점과 한계

○ 행정자료 사용의 장점

- 완전한 범위 (해당 인구)
- 데이터 수집 비용 감소 ("한 번 수집, 여러 번 사용")
- 전국에서 지역까지 일관성있는 정보
 - 중앙 디자인, 중앙 수집 (사용 가능)
 - 중앙 디자인, 로컬 수집 (사용 가능)
 - 로컬 설계, 로컬 수집(사용 불가)
- 빈도

○ 한계

- 데이터의 연계와 매칭이 어려움
 - 영국에서는 키로 사용할 단일 식별자가 없음
 - 중복된 레코드 및 레코드 누락 존재
 - 주소와 같은 데이터 기록방식 차이
 - 시간 경과에 따른 비교 가능성

- 데이터 정의와 분류
- 다른 지리정보의 사용

□ 교훈 및 결론

- 데이터 공유를 성공적으로 수행하기 위해서는
 - 개별 소스를 이해
 - 파일럿 연구 / 타당성 연구를 수행
 - 통계적 또는 방법론적인 한계 극복
 - 시설 공유를 위한 보안 데이터를 개발
 - 기밀 보호 및 신뢰 구축
 - 공동 작업
- 영국의 자료 공유는 시작단계이나, 지금이 적시라고 판단
 - 새로운 법적 제도
 - 새롭고 개선된 사회 인구 통계에 대한 요구 증가
 - 기술적 발전
 - 재정적으로 꼭 필요한 사업에 대한 정부 정책 및 계획

Ⅲ

시사점

- 법적기반 마련 시급
 - 스페인에서는 통계청의 인구등록대장 관리, 타기관의 행정자료 제공 등이 법으로 되어 있어 등록센서스 추진에 어려움이 없어 보였음
 - 현재 우리나라는 각 행정기관으로 자료를 입수하고 있으나, 그 절차가 복잡하고 입수자체가 어려움. 성공적인 등록센서스 추진을 위해서는 무엇보다 법적 기반 마련이 시급하다고 판단됨.
- 행정자료의 관리 방안 필요

- 스페인 통계청은 지자체에서 입수한 인구등록대장을 통합·관리하면서 지속적인 피드백을 통해 조정 역할을 함으로써, 센서스의 기반이 되는 고품질의 인구대장을 보유하고 있음
- 현재 우리나라는 각 기관에서 자료 입수에 급급한 환경에 있으나, 향후 행정자료의 질을 높이기 위한 통계청의 역할이 필요하다고 판단됨

○ 조사방법에 대한 연구 필요

- 각 나라마다 표본조사의 목적에 따라 표본크기, 추정방법 등이 다르므로, 우리나라도 우리 실정에 적합한 표본설계 및 추정방법에 대한 계속적 연구 필요
- 집단시설, 노숙자 등 난조사집단에 대한 조사방법 연구도 필요

○ 지속적 선진사례 연구 및 기술 습득 필요

- 다른 나라에 비해 짧은 시간에 등록센서스를 수행하는 만큼, 앞서서 행정자료의 기반으로 센서스를 수행한 국가들의 사례를 심도있게 연구하고 자료의 연계, 추정, 보정, 품질확인에 이르기까지 각 단계에 관련된 기술을 습득할 필요 있음