
인구주택센서스에 관한
UNECE/Eurostat 합동회의 참가 결과
(Joint UNECE/Eurostat Meeting on Population and Housing Censuses)

2012. 5.

인구총조사과



목 차

I. 회의개요	1
II. 일정 및 발표주제	2
III. 주요 발표 내용	3
IV. 참고자료	41

I. 개요

1. 회의 개요

□ 회의 명칭

- 행정자료를 이용한 센서스 관련 워크숍
(UNECE-UNFPA Training Workshop on Censuses Using Registers)
- 인구주택센서스 UNECE/Eurostat 합동 회의
(Joint UNECE/Eurostat Meeting on Population and Housing Censuses)

□ 회의기간 : 2012. 5. 21(월) ~ 5. 25(금)

□ 회의참가국 : ECE 회원국을 비롯한 약 40개국 및 UNFPA등 5개 기구

□ 회의구성

- 등록센서스 관련 설명회(네덜란드, 노르웨이, UNECE 발표)
- 5가지 주제 관련 약 20개국 및 UNSD 등 4개 기구 발표

2. 출장 개요

□ 출장기간 : 2012. 5. 20.(일) ~ 5. 27.(일), 6박 8일

□ 출 장 자 : 인구총조사과 조수정 주무관

□ 출장장소 : 스위스 제네바 UN사무소(Geneva, Switzerland)

□ 주요활동 : 워크숍 참석, 2010 라운드 센서스 실시 경험공유 및 발전방향 논의

II. 일정 및 발표주제

일 자	주 요 내 용
5.21. (월)	<ul style="list-style-type: none"> . 행정자료를 이용한 센서스 관련 훈련 워크숍 등록센서스 소개 및 관련 기본적인 개념
5.22. (화)	<ul style="list-style-type: none"> . 의제채택 및 의장 선출 - Eric Schulte Nordholt(Netherlands) . UNECE 회원국의 2010 라운드 센서스 개관 Information session on the 2010 Round of Population and Housing Censuses . 등록센서스 실시 현황 결과 보고 Experiences with the use of registers in censuses . 이후 센서스 실시계획(등록센서스 시행 포함) 보고 Should we use registers for our census?
5.23. (수)	<ul style="list-style-type: none"> . 등록 행정자료의 이용가능성, 완전성, 품질 Availability, completeness and quality of data from registers and other sources . 등록 행정자료의 결측치 추정 방법론 Methodology used for estimating Census tables based on incomplete information
5.24. (목)	<ul style="list-style-type: none"> . 전세계 2010 라운드 센서스 실시 현황보고 Information session on the 2010 round of population and housing censuses . 전수조사와 표본조사 관련 실시 결과 보고 Full enumeration versus sample surveys . 마이크로 데이터 접근 Access to microdata
5.25. (금)	<ul style="list-style-type: none"> . 주소/거처 목록 작성 Address/dwelling listing . 인터넷 조사 Internet data collection

III. 주요 발표내용

1. UNECE-UNFPA 행정자료를 이용한 센서스 관련 트레이닝 워크숍

행정자료를 이용한 센서스(등록센서스)에 사용된 일반적인 개념과 장단점 등에 대한 설명하고 기존 센서스와 비교

1) 등록 기반 통계의 장단점 Why register-based statistics?

발표자 : Eric Schulte Nordholt(Netherlands)

데이터 수집

종류

- 센서스/전수조사
- 표본조사
- 행정자료 등록부

행정자료는 행정적 목적을 위해 수집됨

데이터 수집방법 선택 시 비용 효율성, 응답부담, 통계데이터의 질, 이 사이에서 절충이 필요

통계작성 예산(Cost)

조사환경

- 각국 통계기관의 예산 삭감 및 제한
- 세부적이고 새로운 통계에 대한 이용자들의 요구

=> 통계생산의 효율성 증가 필요

행정 데이터(Administrative data)

- (국가통계기관)데이터 수집에 거의 비용이 들지 않음
- 통계적 목적을 위한 데이터 수집 대신 개선된 현존 데이터 자원 사용

- 등록기반통계는 무효는 아니지만 일반적으로 표본조사 (특히 전통적인 센서스)보다 저렴함

- 응답부담 : 행정자료의 이용은 추가적인 응답부담이 없음
 - (기업) 당국에 보고하는 시간이 길음
 - (국민) 이미 준 정보에 대한 응답 거부
 - (국가통계기관) 증가하는 응답부담 문제

- 통계결과의 질
 - 타당성 : 조사대상, 범위(coverage), 변수, 조사기간 등
 - 정확성 : 등록자료(Registers)는 행정적 목적에 적합하므로 통계적 목적에 맞춰 정확성 향상
 - 통계작성 기간 및 시의성 : 행정 처리는 시간이 오래 걸리고 업데이트 및 데이터 추출이 오래 걸림
 - 비교가능성과 일관성 : 일관된 등록기반 통계 시스템 구축, 다른 소스에 근간한 통계와 조화

- 통계적 조사(statistical survey)와 등록 기반 통계(register-based statistics) 비교
 - cost(++), 응답부담(++), 타당성(-), 정확성(0), 시기적절성(-)
 - 등록기반 자료는 적은 비용으로 더 작은 그룹 통계 제공가능 (표본조사와 비교한 경우)
 - 매년(또는 더 자주) 데이터 제공 가능
 - 등록 기반 통계 생산이 더 효율적임
 - 등록기반 통계는 표본조사에 의해 보완될 필요가 있음

2) 기본개념의 정의 Definition of basic concepts

발표자 : Eric Schulte Nordholt(Netherlands)

□ 등록부(Register)의 정의

- 단위수준의 데이터의 체계적인 수집, 업데이트 가능

□ 구성 단위(Units)

- 단위(Unit)는 유일하게 식별되어야 함(가급적 식별 부호에 의해)
 - 예) 중앙 인구 등록부(Central Population Register)의 규칙에 의해 식별된 '거주자'
- 구성단위 = 조사대상

□ 갱신(Updating)

- 추가된 새로운 구성단위(신생아, 이민자)
- 삭제된 현존 구성단위(사망자, 귀화자)
- 데이터 속성(attribute)의 변화
- 데이터 정정, 변화가 있었거나 계속적으로 발생하는 변화에 대한 갱신
- 전통적 센서스 파일은 등록부(register)가 아님 : 모든 데이터는 정시 또는 같은 기간 동안에 단위수준으로 수집됨
 - 다음 센서스에서 수집될 새로운 데이터
 - 등록부(register)로 업데이트 안됨

□ 행정적 등록부(Administrative registers)는

- 주로 행정상 정보 시스템에서 사용됨
- 공공 또는 민간 부문의 재화 또는 서비스 생산에서 사용됨
- (개인, 건물, 기업) 각각에 대한 결정에 사용됨
- 통계용으로 사용된 행정상 등록
- 행정 데이터 소스 : 행정 목적에 사용된 모든 데이터 소스
- 등록 자료의 보유자 : 행정 등록에 책임이 있는 당국(행정 관리자)

□ 통계적 등록부(Statistical registers)

- 행정적인 등록에 의한 데이터를 통계목적으로 처리함으로써 생성
- 일반적으로 몇 개의 행정자료와 결합하며, 단일 행정자료도 가능

□ 기본 등록부(Base Registers)

- 전문화된 등록부(specialized register)와 대조됨
- 행정상 기본 등록부(Administrative base registers)
 - 행정에 기본적이고 일반적인 자원
 - 인구 관련하여 비축하고 고유정보를 보유
- 통계적 기본 등록부(Statistical base registers)
 - 해당되는 행정적 기본 등록부에 기반을 둠
 - 전체 통계적 시스템에 가장 중요

□ 등록기반 나라들의 주요 통계적 기본 등록부(Statistical base registers)

- 인구 register(PR) : 주민등록번호
- 사업체 register(BR) : 사업자등록번호

- 부동산 register : 부동산, 건물, 거주지, 주소
 - 개인 및 기관과 연결을 위한 ID : 수치적 주소

□ 전문화된 등록부(Specialized registers)

- 특정목적에 전문화 된 등록부는 한 가지 특정 목적을 제공
- 정의된 대상 지역의 데이터 포함
- 한 가지 이상 기본 등록부(basic register)의 연계
- 예시
 - 개인에 관한 등록부 : 세금, 교육정도, 사회보장, 보건의료
 - 사업 등록부 : 세금, 부가가치세, 무역
 - 부차적 인구관련 등록부 : 농장, 병원, 학교

3) '개인' 통계와 등록부 개발 Individual based statistics and register development

발표자 : Harald Utne, Statistics Norway

□ 통합된 데이터에 기반한 통계적 시스템

- 많은 나라들은 지리적 지역 단위의 통계 생산을 위한 분권화된 체계
- 방법 : 각각 수준에서 통계를 생산하고 다음에 결과를 넘김
- 국가 기관에 대다수 통계전문가들은 통합된 데이터로 작업

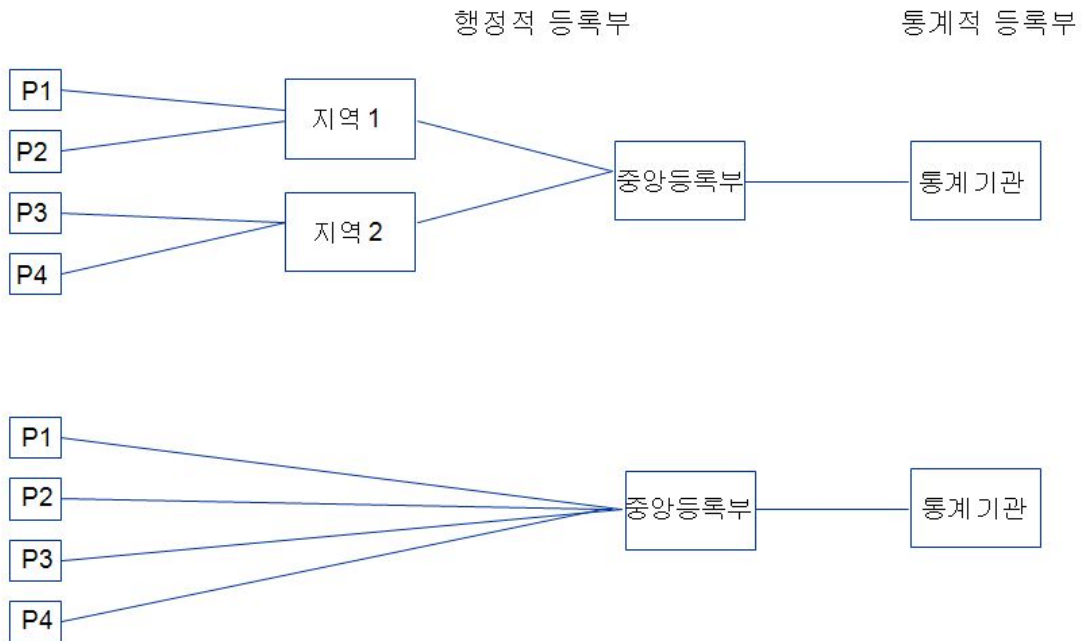
□ 체계적인 기본 구성단위(Unit) 데이터 수집

	snr	FOEDELSEL	FOEDE_KOMMNR_LAND	KJOENN	FOEDESTED	STATSBORGERSKAP	familiennr_snr	PERSONKODE	SIVILSTAND
1082	8856641	20030101	1201	2	HAUKELAND SYKEHUS	000	7966593	3	1
1083	8650067	20030101	0106	1	SH ØSTFOLD/FR.STAD	000	1891655	3	1
1084	5151514	20030101	1201	2	HAUKELAND SYKEHUS	000	6533864	3	1
1085	8284821	20030101	1201	2	HAUKELAND SYKEHUS	000	2121359	3	1
1086	7050510	20030101	1201	2	HAUKELAND SYKEHUS	000	2121359	3	1
1087	1272246	20030101	1201	1	HAUKELAND SYKEHUS	000	0620734	3	1
1088	7841320	20030101	1201	2	SYKEHUSET BUSKERUD	000	1804774	3	1
1089	8411646	20030101	0204	1	RIKSHOSPITALET	000	4968298	3	1
1090	7474984	20030101	0704	2	SSH VESTFOLD	000	3203430	3	1
1091	7487582	20030101	1443	1	NORDFJORD SYKEHUS	000	4209931	3	1
1092	8656910	20030101	1201	1	HAUKELAND SYKEHUS	000	0504011	3	1
1093	8314175	20030101	0704	2	SSH VESTFOLD	155	2012221	3	1
1094	8621854	20030101	0602	2	SYKEHUSET BUSKERUD	000	6278817	3	1
1095	8462337	20030101	0602	2	SYKEHUSET BUSKERUD	000	4961844	3	1
1096	8618451	20030101	1101	1	HAUKELAND SYKEHUS	000	3600814	3	1
1097	7919364	20030101	0602	2	SYKEHUSET BUSKERUD	000	1070361	3	1
1098	5947026	20030101	1201	1	HAUKELAND SYKEHUS	000	5994434	3	1
1099	8857951	20030101	1201	2	HAUKELAND SYKEHUS	000	4938466	3	1
1100	4649011	20030101	1201	1	HAUKELAND SYKEHUS	000	4888867	3	1
1101	7627675	20030101	1201	2	HAUKELAND SYKEHUS	000	2705598	3	1
1102	7312191	20030101	1201	1	HAUKELAND SYKEHUS	000	1523506	3	1
1103	7141149	20030101	1601	2	ST.OLAVS HOSPITAL	000	6611570	3	1
1104	6940163	20030101	0602	1	BUSKERUD SSH	000	7656806	3	1
1105	8505854	20030101	1443	2	NORDFJORD SYKEHUS	000	6258532	3	1
1106	8636880	20030101	0106	1	SH ØSTFOLD/FR.STAD	000	3387885	3	1
1107	0860657	20030101	0604	2	KONGSBERG SYKEHUS	000	6967551	3	1
1108	8158843	20030101	1201	2	HAUKELAND SYKEHUS	000	6147061	3	1
1109	8142985	20030101	1221	1	STORD SYKEHUS	000	2321811	3	1
1110	8426364	20030101	1719	2	LEVANGER SYKEHUS	000	5845937	3	1
1111	8636628	20030101	0301	2	ULLEVAL SYKEHUS	000	0540100	3	1
1112	6649788	20030101	0906	1	AUST AGDER SYKEHUS	000	0175748	3	1
1113	3437798	20030101	0106	2	SH ØSTFOLD/FR.STAD	105	7978676	3	1
1114	5975987	20030101	0230	1	SSH AKERSHUS	000	7337487	3	1
1115	6274581	20030101	1106	2	HAUGESUND SJUKEHUS	000	1698100	3	1

고유 식별자

□ 중앙집권 또는 지방 분권화된 등록부(개인)

- 분권화 : 개인은 지역등록부에 보고하고 해당 지역은 중앙등록부에 보고 만약에 중앙등록부가 없다면 직접 통계기관에 보고
- 중앙집권화 : 개인이 직접 중앙 등록부에 보고(지역등록부가 있을 수 있음)



- 중앙집권화된 등록 시스템의 이점
 - ID 숫자의 중앙 집권화된 체계
 - 중복 없음(고유 ID 숫자)
 - 더 효율적인 데이터 수집
 - 여러 등록부에 있는 데이터 연계가 쉬움

4) 일반적인 전제조건 General preconditions – Requirements to data from administrative sources

발표자 : Harald Utne, Statistics Norway

통계생산을 위한 행정자료 이용 전제조건

- 법적근거, 공증, 고유 식별 시스템, 신뢰할만한 등록 시스템, 당국 간의 협력

법적 근거

- 통계적 목적을 위해 행정자료 재사용함을 반영
- 입법은 국가통계기관에 부여하고 개정 가능해야 함
 - 구성기본단위(개인 식별 정보 포함)인 행정자료에 접근
 - 서로 다른 행정 자료 간 데이터 연계할 권리
- 정보보호에 관한 법률 제정
- 통계적 목적을 위한 행정자료, 통계목적 및 연구목적을 위한 통계적 데이터라는 일방통행 원칙(one way traffic principal) 적용(노르웨이)

통일된 식별 시스템

- 기본 구성단위(Unit)에 대한 고유 식별 번호
 - 개인, 기업, 사업체, 숫자로 나타낸 주소
- 고유한 ID 숫자는 절대적으로 필요하진 않지만 매우 가치있음

구성단위와 식별자

- 구성단위(Unit)의 정의

- 자연적 구성단위는 문제없음(예: 개인)
- 사업체 단위 : 매우 정교한 규칙에 의해 정의 되어야함
- 고유한 식별자
 - 고유 코드는 구성단위(unit)의 생애동안 변경되면 안됨
 - 아무런 정보를 포함하지 않는 코드를 사용하는 것이 최고의 방법
- 기준 시간과 실제 사건 데이터
 - 변화 및 사건 날짜가 존재하는 것이 중요
 - 시간상 실제 시점(출생, 사망), 점진적 변화(기업산업)
 - 실제 변화 날짜와 등록된 날짜
 - 실제 사건 발생과 등록부에서 정정한 것을 구별

5) 행정 등록부를 이용한 센서스 개관 Overview of approaches to censuses using registers

발표자 : Paolo Valente, UNECE Statistical Division

□ 완전한 행정 등록부 기반 센서스

- 다양한 행정 등록부 상 데이터를 통합함으로써 센서스 시행
 - 오랜 단계를 걸쳐 결정
 - 센서스는 등록기반 정보 시스템의 부분
 - 주요요건
 - 인구 등록부와 거주 등록부 이용가능
 - 가구 확인과 개인과 거주지 연결할 수 있는 능력
 - 인구등록부 사용에 대한 대중과 법적 지지
 - 장점
 - 개인 부담 없음
 - 일반적으로 매년 데이터 이용 가능
 - 고품질 행정등록부가 구축되면 전수조사보다 더 저렴한 비용
 - 한계
 - 상당한 자원 구축과 유지
 - 수집된 특성은 등록부 상 있는 것에 제한됨
 - 센서스의 품질은 등록부의 포괄범위와 품질에 의존
 - 등록부에 데이터는 일반적으로 통계적 목적이 아님
 - 통계작성기관은 또한 등록 자료 보유 당국에 의존적
- ※ 실시국가 : 핀란드, 덴마크, 오스트리아, 벨기에, 노르웨이, 슬로베니아, 스웨덴

□ 등록부 기반 데이터 + 기존 표본 조사

- 몇 개의 변수가 등록부에 이용 불가할 때 기존 표본조사 데이터가 사용가능한 경우(예 : 노동력조사)
 - 개인(individual) 수준으로 데이터 연결
 - 현장조사 안함
 - 장점
 - 응답 부담 없음
 - 예산 절감 : 특정한 현장조사 없음
 - 센서스 결과와 조사(survey)결과의 일관성
 - 한계
 - 서로 다른 “조사와 행정등록부”를 개인수준으로 정보를 성공적으로 연결하는 절차는 복잡
 - 표본조사에 의한 세부정보 한계
- ※ 실시국가 : 네덜란드(가상센서스), 아이슬란드

□ 등록부 기반 데이터 + 특별 표본 조사(ad-hoc sample survey)

- 기존 표본조사 이용 대신에 특별 표본조사 실시
- 특별표본조사는 선택된 변수에 대한 데이터를 조사하거나 행정등록부의 정확성 평가에 이용
- 장점
 - 인구 등록부의 정확성 검사 가능
 - 인구수 조정 가능
 - 제한적 응답부담(표본조사)
 - 제한적 비용(표본조사)

○ 한계

- 방법은 복잡할수 있음
- 표본조사로 얻은 세부 정보의 한계

※ 실시국가 : 이스라엘, 스페인, 스위스, 터키

□ 등록부 기반 데이터 + 전수 조사

○ 일부 변수의 데이터가 등록부상 누락 된 경우 세부적인 데이터가 요구됨(표본으로는 충분치 않을 때)

○ 전수조사는 등록부 자료의 정확성과 포괄범위에 eogoks 평가 가능

○ 일부 등록부 데이터는 조사표에 견본인쇄 가능

○ 장점

- 데이터 수집을 함으로써 완전한 등록기반 센서스 관점에서 등록부를 구축, 갱신, 개선하는데 사용됨

○ 한계

- 고 비용(전수 현장조사)
- 응답부담(전수 현장조사)
- 방법은 적용하기 복잡

□ 등록부 기반 데이터 + 전수 조사 + 특별 조사

○ 중요한 변수나 업데이트와 개선이 필요한 등록부를 대상으로 전수조사

○ 세부사항이 많이 요구되지 않는 변수에 대해서 표본조사 실시

2. 행정자료를 이용한 센서스 관련 전문가 회의 expert Group Meeting on Censuses Using Registers

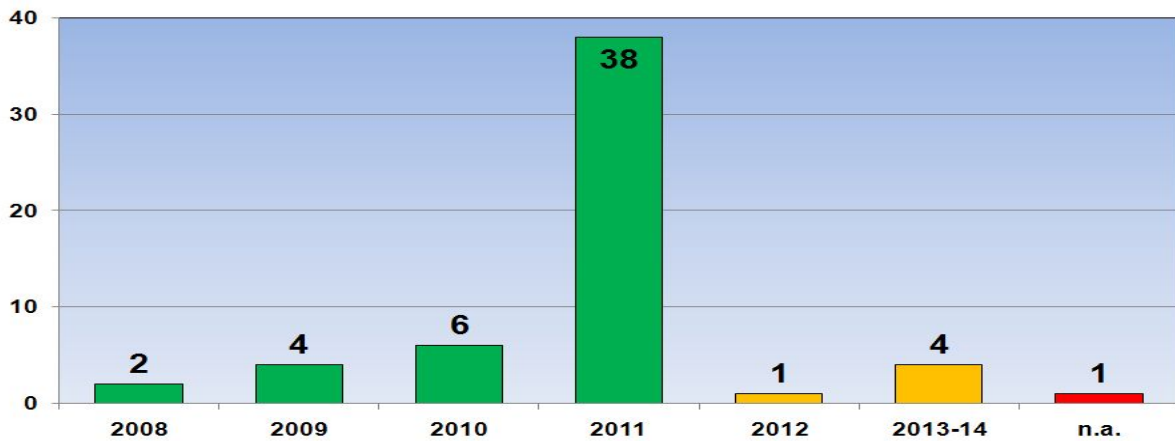
행정자료를 이용한 센서스(등록센서스)실시 국가 현황 및 등록센서스 실시 국가의 센서스 결과에 대한 발표

1) UNECE 지역 등록센서스 실시 현황(2010 라운드 센서스)

overview of 2010 round of censuses in the UNECE region with focus on the use of registers

발표자 : Paolo Valente, UNECE Statistical Division

□ 국가별 센서스 년도



□ 2000, 2010라운드 센서스 방법

		2010라운드 센서스 방법				
		전통적	결합된	등록기반	기타	전체
2000 라운드 센서스 방법	전통적	29	8	1	2	40
	결합된	-	4	2	-	6
	등록기반	-	-	4	-	4
	미실시	2	2	1	-	5
	전체	31	14	8	2	55

※ 참고 : <http://www.unece.org/stats/census.html> or using Google: census wiki

2) 등록 기반 센서스 실시 국가 비교 Comparing approaches of different (partly)register-based countries

발표자 : Eric Schulte Nordholt, Statistics Netherlands

□ 연구 주제

- 결측치 보정에 사용된 방법
- 등록부와 조사데이터 결합에 사용된 방법

□ 연구 참여국가

- 완전 등록기반 국가
 - 노르웨이, 덴마크, 스웨덴, 핀란드, 오스트리아, 슬로베니아
- 부분적 등록기반 국가
 - 스위스, 독일, 폴란드

□ 결과

- 센서스 정보를 얻기 위해 많은 접근법을 이용
- 결측치에 대해서는 Imputation 방법을 사용
- 등록부와 조사(Survey) 결합시에 전통적인 가중법 (Traditional weighting) 이용
 - 2011 네덜란드 센서스에서 교육수준과 일부 주택 변수에서 유일하게 반복 가중법(Repeated weighting)사용
- 네덜란드에서 이 접근법에 대해 가상센서스(virtual census)라 칭함
 - “완전한 등록기반(Fully register-based)” : 센서스 조사표없음
 - "부분적 등록기반(partly register-based)" : 몇가지 중요한 변수를 위해 조사(survey)를 실시

3) 행정 데이터의 통계적 데이터로 변환(2011 센서스)Model of transformation of administrative data to statistical data

발표자 : Janusz Dygaszewicz, Central Statistical office Poland

□ 데이터 품질 척도

○ 행정 등록부(administrative registers)의 품질 기준

- 데이터의 시의성, 방법론적 호환성, 완전성, 등록소에서 사용된 표준 식별자, 유용성, 행정자료와 조사(survey)에서 얻은 데이터의 호환성

○ 행정자료 등록부 처리 과정의 품질 기준

- 과도한 포괄 오류율(excessive coverage error rate), 불완전한 포괄 오류율(incomplete coverage error rate), 주관적 지표의 완전성, 객관적인 지표의 완전성, 임퓨테이션 비율, 데이터

□ 데이터 추출

○ 다양한 시스템의 다른 형식(format)의 데이터들 통합

○ SAS 소프트웨어 기반의 제작환경으로 데이터 추출

○ SAS 테이블 처리에 적합한 한 가지 포맷형식으로 데이터 전환

○ 가져온 데이터 구조의 검증이 이 과정에서 중요함

Register/System	Name	Central Y/N	Relational Y/N	Data format
PESEL	General electronic system of Population Register	Y	N	TXT
KEP	Register of the National Taxpayers	Y	N	TXT
GZM	Community Registers of Residence	N	Y	SQL Server
PIT	Personal income tax register	Y	N	TXT
SI MS	Ministry of Justice system	Y	N	XLS
POBYT	Foreigners evidence system	Y	Y	SQL Server
ZUS CRPS	Central Register of Contribution Payers	Y	N	TXT
ZUS CRU	Central Register of Insured Persons	Y	N	TXT
ZUS SER	Pension insurance system	Y	N	TXT
KRUS	Agricultural Social Insurance Fund System	Y	N	XML
CWU NFZ	Central Register of Insured	Y	N	TXT
PFRON	State Fund for Rehabilitation of Disabled Persons	Y	N	TXT
ARIMR	Agency for Restructuring and Modernisation of Agriculture system	Y	Y	XLS
EPN	Property Tax Records	N	Y	SQL Server

□ 데이터 변환

○ 자료수집(profiling)

Table: OSOBY_06_2011_OBM
 Schema: SAS
 Total records: 39056904

Field Name	F	Count	Null Count	Pattern Count	Unique Count	Uniqueness	Data Type	Data Length	Maximum Length
symt	1	39056904	0	1	2496	0.01	CHARACTER	6 chars	6
nr	2	39056904	0	1	2	0	CHARACTER	1 chars	1
OSOBA_Id	3	39056904	0	(not applicable)	4487373	11.49	DOUBLE	15 chars	(not applicable)
pesel_GZM	4	39056904	22339	1	38272319	98.05	CHARACTER	11 chars	11
nazwisko1_GZM	5	39056904	0	2885	385896	0.99	CHARACTER	120 chars	29
nazwisko2_GZM	6	39056904	38436269	659	105921	17.07	CHARACTER	120 chars	28
nazwisko_rod1_GZM	7	39056904	3563944	2359	431973	1.22	CHARACTER	120 chars	28
nazwisko_rod2_GZM	8	39056904	39003918	266	17100	32.27	CHARACTER	120 chars	28
imie1_GZM	9	39056904	121	782	40130	0.1	CHARACTER	120 chars	24
imie2_GZM	10	39056904	17400852	618	28559	0.13	CHARACTER	120 chars	42
stan_cywilny_os_GZM	11	39056904	5636558	1	8	0	CHARACTER	1 chars	1
data_ur_GZM	12	39056904	0	(not applicable)	40015	0.1	DOUBLE	15 chars	(not applicable)
msce_ur_GZM	13	39056904	5548	10527	381205	0.98	CHARACTER	255 chars	32
kraj_ur_kod	14	39056904	286227	35	180	0	CHARACTER	255 chars	32
kraj_ur_nazwa	15	39056904	286227	1	180	0	CHARACTER	255 chars	3
plec_GZM	16	39056904	0	(not applicable)	3	0	DOUBLE	15 chars	(not applicable)
adr_meld_kod_gmn_GZM	17	39056904	982142	2	3282	0.01	CHARACTER	7 chars	7
adr_meld_naz_gmn_GZM	18	39056904	983632	165	2438	0.01	CHARACTER	255 chars	34

○ 표준화와 구문분석

Incorrect data format	Format after standardization
1985-02-21	19850221
1985.02.21	19850221
1985 02 21	19850221

Voivodeship	City	Street	Place of birth
MAZOWIECKIE	WARZSWA	ul. DŁUGA	LONDYN - ANGLIA
MAZPWOECKIE	WARS-AWA	Ulica DŁUUGA	LONDYN - WLK BRYTANIA
ZAZOWIEVCKIE	AWRSZAWA	DLUGAA	LONDYN/CHELSEA
MZAOWIECIE	WARSZAAAWA	DŁUGA (ul.)	LONDYN BRIDGE



Voivodeship	City	Prefix	Street	Place of birth
MAZOWIECKIE	WARSZAWA	UL	DŁUGA	LONDYN

○ 데이터 정화(data cleaning)

: 데이터의 검출, 정정, 이동 또는 일관성 부여

Description		Before cleaning			After cleaning		
Group of variables	Variable	Total	Incorrect		Total	Incorrect	
			Total incorrect	In %		total incorrect	In %
Address of permanent residence	COMMUNITY	4320724	428469	9,92%	4316061	72797	1,69%
	CITY	4353209	207399	4,77%	4352983	43086	0,99%
	STREET	3514154	573899	16,34%	3440932	125392	3,65%
	PREFIX	0	0	-	108551	0	0,00%
Address of residence	COMMUNITY	739088	100282	13,57%	738717	11666	1,58%
	CITY	742388	30644	4,13%	742336	6344	0,86%
	STREET	607939	102725	16,90%	593370	21012	3,55%
	PREFIX	0	0	-	18416	0	0,00%
Corresponding addresses	COMMUNITY	2005	132	6,59%	2005	30	1,50%
	CITY	448791	21678	4,84%	448704	4796	1,07%
	STREET	377849	64871	17,17%	374220	20575	5,50%
	PREFIX	0	0	-	11192	0	0,00%
Personal Data	NAME	4355764	7208	0,17%	4355757	5534	0,13%

○ 데이터 변환

- 예 : 성별 변수인 Female, F -> 2 과 Male, M -> 1

○ 검증(validation)

- 데이터 검사
- 방법론 학자에 의해 준비된 알고리즘에 의해 비정상적 변수 수정
- 더 이상 개선 불가능한 데이터 제외

○ 중복 제거(deduplication)

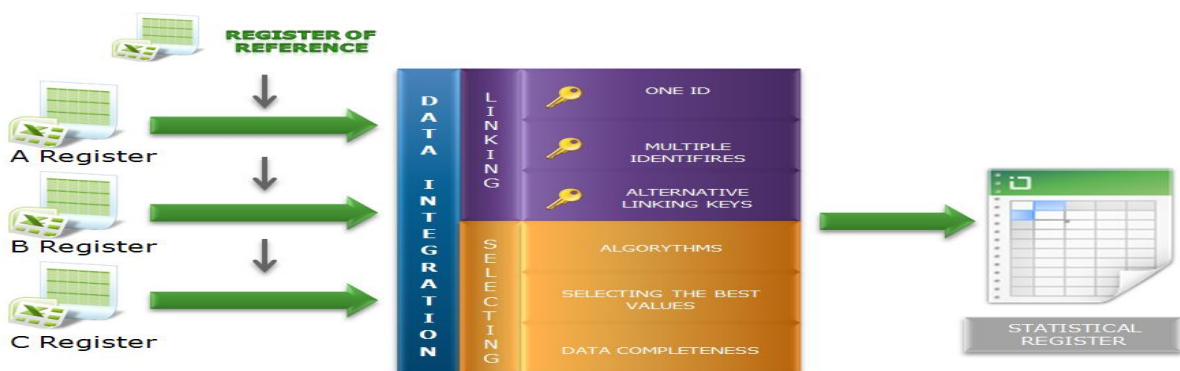
- 중복된 단위(Unit)의 제거
- 적법행위인지 분석을 포함한 세부 분석

imie1_GZM	nazwisko1_GZM	plec_GZM	pesel_GZM	adr_ulica_GZM	adr_nr_dom_GZM	adr_nr_lok_GZM	data_zam_od_GZM
JAN	KOWALSKI	M	00000000001	PULAWSKA	4	37	19840712
ANNA	MALINOWSKA	K	00000000002	ANDERSA	7		20010205
ADAM	PIOTROWSKI	M	00000000003	FILTROWA	2	45	20090101
JAN	KOWALSKI	M	00000000001	PULAWSKA	4	37	20070303

imie1_GZM	nazwisko1_GZM	plec_GZM	pesel_GZM	adr_ulica_GZM	adr_nr_dom_GZM	adr_nr_lok_GZM	data_zam_od_GZM
ANNA	MALINOWSKA	K	00000000002	ANDERSA	7		20010205
ADAM	PIOTROWSKI	M	00000000003	FILTROWA	2	45	20090101
JAN	KOWALSKI	M	00000000001	PULAWSKA	4	37	20070303

○ 데이터 통합(integration)

- 최적, 가장 최근, 정확한 값의 선택 과정
- 통계적 등록부(statistical register) 생성에 이용



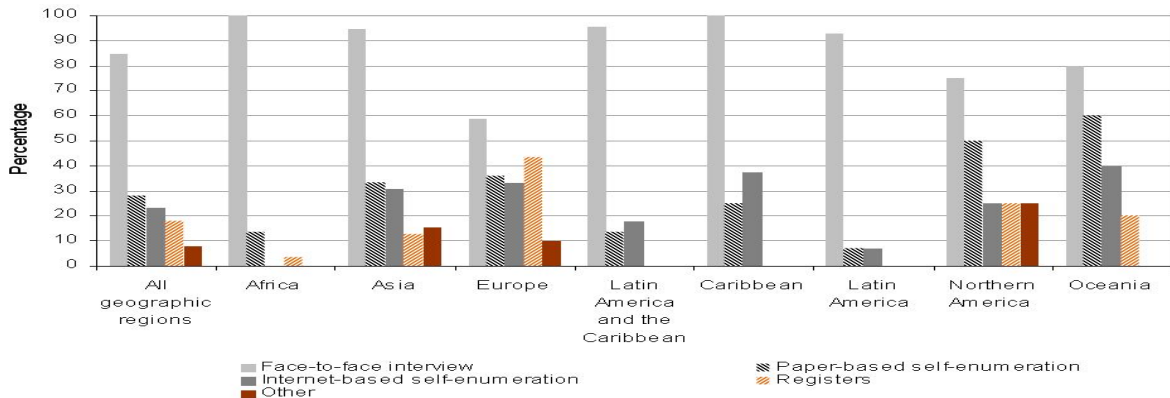
3. 인구주택센서스 전문가 그룹 회의 Group of Experts on Population and Housing Censuses

1) 전 세계의 2010 Round 센서스 현황 보고

발표자 : Keiko osaki-tomita, UN Statistics Division

□ 2010라운드 센서스 2005년에 183개국이 census 수행하고 2014년 46개국 계획 중에 있으며 2011년에 센서스를 가장 많이 수행했음

- 2010 라운드 센서스의 변화 : 데이터수집방법, 진보된 기술사용
- 조사방법 : 단일방법에서 추가적으로 또는 결합하여 데이터를 수집



(조사표 자기기입식, 인터넷조사, 전화면접, PDA를 이용한 면접조사등 기타)

- 센서스방법 : 현존하는 행정자료의 연결을 통한 센서스 실시 증가 (특히 유럽국)하고 많은 국가들이 센서스데이터를 보충하기위한 표본조사 실시(ACS)

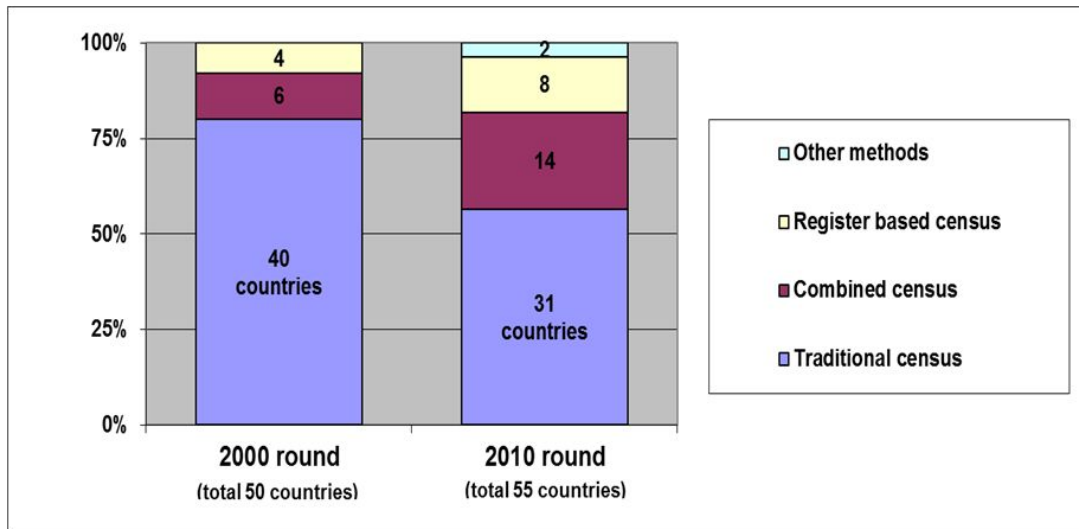
○ 현대 기술의 사용

- GIS*(Geographic Information System)과 Global Positioning System(GPI) 기술을 가장 많이 사용
 - * EA를 위한 지도 제작 기본 생성, 현장조사 감독, 매핑 도구를 가지고 센서스 결과 제공 등
- PDA, 태블릿 컴퓨터, 인터넷과 CATI를 이용한 조사
- 정제된 스캔기술 사용 : OMR, OCR, ICR
- 센서스 결과 제공방법 : CD/DVD, static web page, 온라인 데이터베이스, 웹 기반 매핑 GIS
- 결과 및 기대효과 : 자료수집과 결과제공 시기 간 격차가 줄어들었으며 데이터처리의 오차를 줄임으로써 데이터 질적 향상 기대
- 당면 문제 : 기술(GIS, 행정자료 연계기술) 발전과 함께 기술 전문가 부족, 응답거부 증가, 센서스에 소요되는 예산 증가, 정치적 사회적 불안전하여 센서스 연기 또는 취소된 국가가 있었음

2) UNECE 회원국 2010 Round 센서스 보고 Overview of 2010 round of censuses in the UNECE region

발표자 : Paolo Valente(UNECE)

□ 2000과 2010 라운드 센서스 방법



- 등록센서스, 결합센서스, 전통적인 센서스로 구분
- 전통적인 센서스는 감소, 결합하거나 등록센서스 증가

□ 전통적인 센서스에서 사용된 조사방법(31개국)

- 면접조사(20개국), 자기응답기입식(6개국), 면접과 자기응답(5개국), 인터넷조사(7개국)

□ 결합된 센서스에 사용된 조사방법(14개국)

- 면접조사(5개국), 자기응답기입식(5개국), 면접과 자기응답(4개국), 인터넷조사(10개국)

□ 인터넷조사(17개국)

○ 인터넷 조사율

- Estonia 65%, Canada 54%, Portugal 50%, Bulgaria 41%, Italy 33%, Czech Republic 27%

○ 인터넷조사 결과

- 데이터 질적 향상, 무응답 항목 감소, 데이터 처리과정 단순화, 종이절약(환경보호)
- 보고된 다른 주요 사건(보안문제) 없음
- 비용 상 제한된 정보만 가능

※ UNECE 회원국에 관련 정보 제공 : [Http://www.unece.org/stats/census.html](http://www.unece.org/stats/census.html)

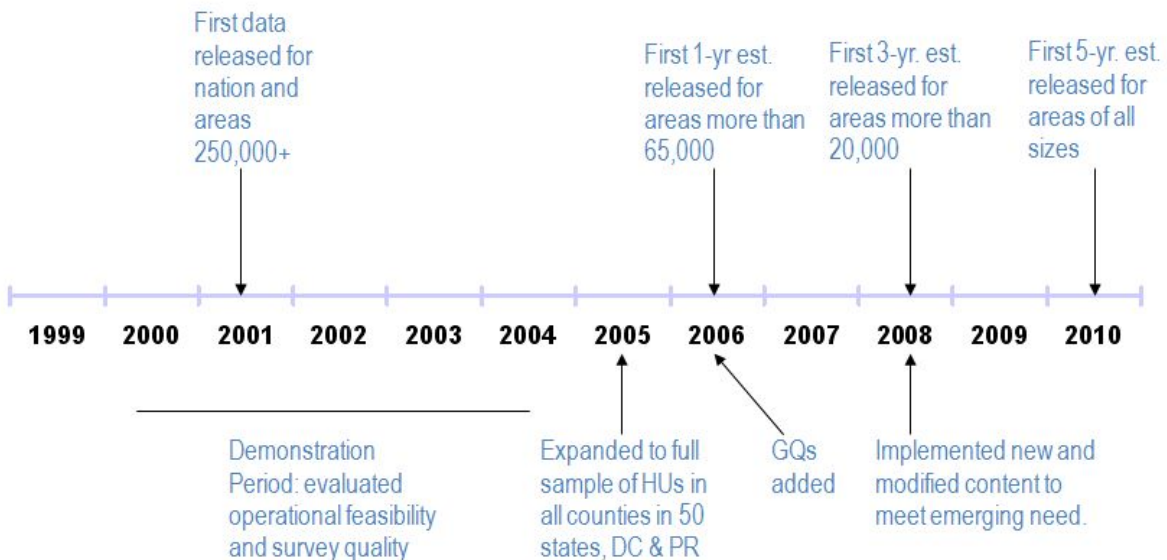
3) American Community Survey와 2010 센서스의 이점

발표자 : Frank A. Vitrano(US Census Bureau)

□ ACS 소개

- 1940~2000년 까지 미국 10년주기 센서스에는 전수항목 (Short form)과 표본항목(Long form)으로 구성
- 국회의원선거구 확정 및 재정분배에 사용
- 중앙정부의 필요에 의해 추진된 조사지만 주정부, 지역, 경제, 미디어, 교육 등 다른 기관에서도 이 데이터에 의존
- 시의성 있는 최근자료에 대한 요구를 충족하기위해 ACS 개발
- 2010센서스에는 기본항목만 조사하고 2005년에 세분화된 사회경제 데이터를 수집하는 ACS 시작

□ ACS 연혁



□ ACS 장점 : 최근데이터제공, 센서스의 단순화

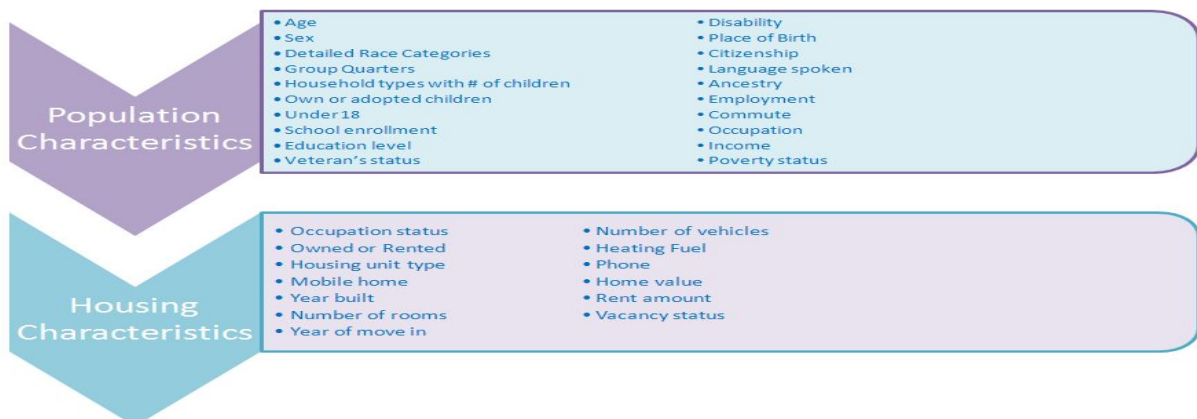
○ 표본항목(Long form)을 제거함에 따른 이점

- 응답부담 경감, 우편응답율 향상, 2개국어 제공, 인쇄 및 우편발송 감소, 조사의 단순화, 조사 업무량 및 자료처리 기계 수요 감소

□ ACS 데이터 활용

- 기본적인 5개 비영어권 언어와 45개 언어 선택
- 어떤 가구가 비영어권 언어로 직접적으로 우편 엽서를 받을 지 결정하여 전화조사도움센터에서 가능할수 있게 함
- 전화조사도움센터에 5개 외국어 담당마다 업무량 추정 및 채용인원 계산
- 2010센서스 결측치 추정
- ACS 처리시스템 인프라가 Island지역 데이터 처리에 지원
- 특정 인종과 Hispanic Origin 그룹이 충분히 과대표집 되도록 증화하는데 이용

□ ACS 데이터베이스



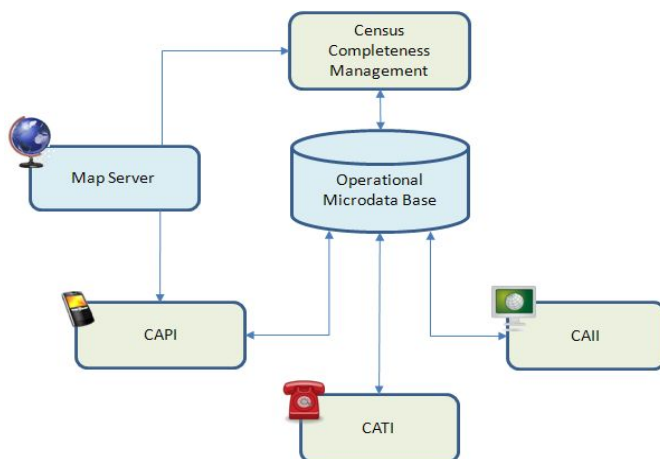
4) POLAND 현대 센서스 Modern census in poland

발표자 : Janusz Dygaszewicz(Poland)

- 센서스 방법 : 혼합 모형(Mixed Model)
 - 행정자료(기본 인구학적 변수를 포함한)와 20%특별 표본조사와 결합
 - 2010년 농업센서스, 2011년 인구센서스 실시

- 행정자료 획득(공간 데이터 참조 등록자료를 포함)
 - 행정자료 보유기관 : 행정내무부, 법무부(Ministry of Justice), 재정부(Ministry of Finance), 외국인 사무소(Office for foreigners), 사회보험기관(Social Insurance Institution) 등 포함 17개 기관
 - 센서스에서 행정자료 활용
 - 조사 데이터의 직접적인 소스, 센서스 프레임(거처주소 조사)에 포함된 개체들의 목록 생성을 위한 정보의 소스 그리고 imputation, 데이터추정, 데이터의 질 비교 등을 위한 정보 출처

- 데이터 수집의 온라인 경로



- .CAII : Computer Assisted Internet Interview
- .CAPI : Computer Assisted Personal Interview
 - GPS 기능 탑재한 HTC 이용
- .CATI : Computer Assisted Telephone Interview

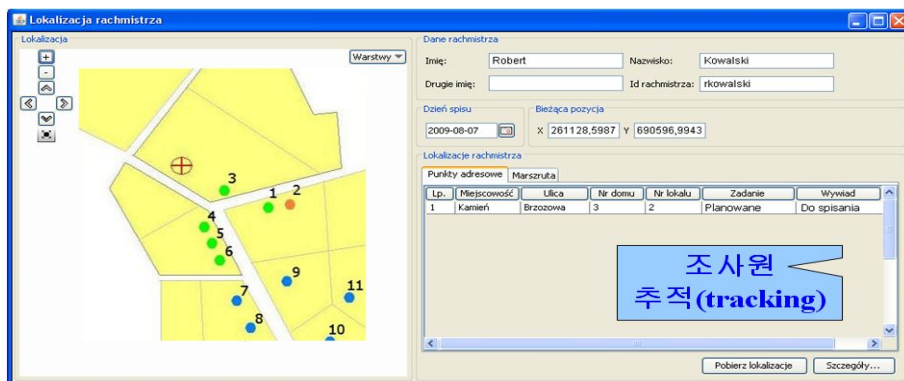
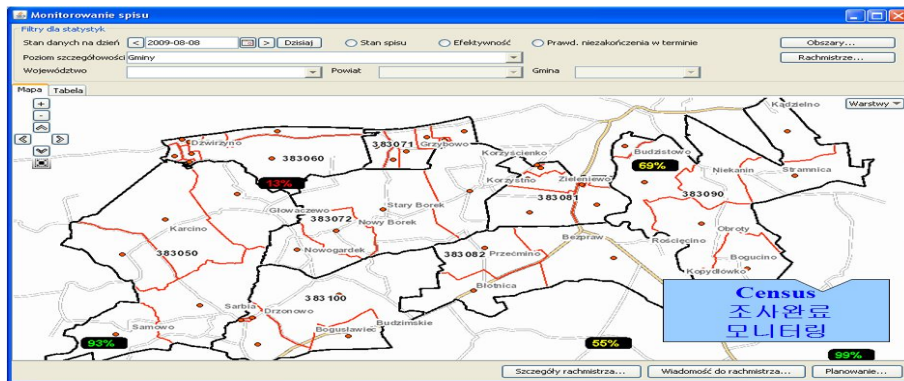
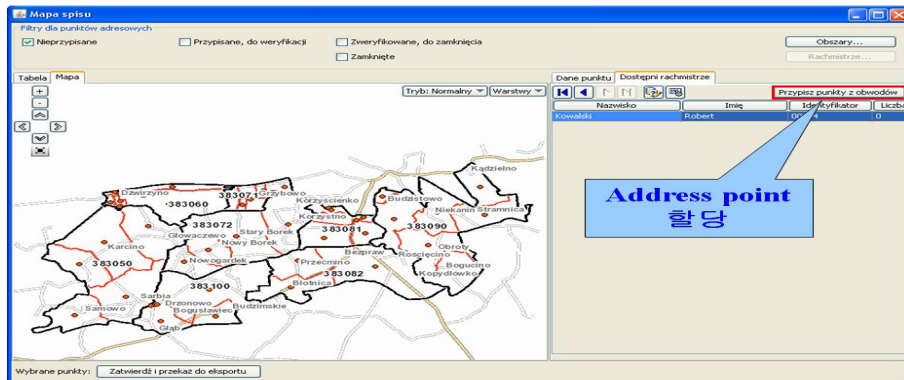
※ 자기기입식 조사를 실시한 일반적인 부류 : 남자, 약 24살, 도시 거주자, 중고등 교육수준

□ 현장조사 운영

○ 조사관리자 업무

- 주소 위치(Address Point, 조사원에의해 확인)와 센서스 지역 관리
- 조사원 관리 감독, 조사원에게 발생한 긴급상황 관리, 필요한 정보제공

<조사관리 시스템>

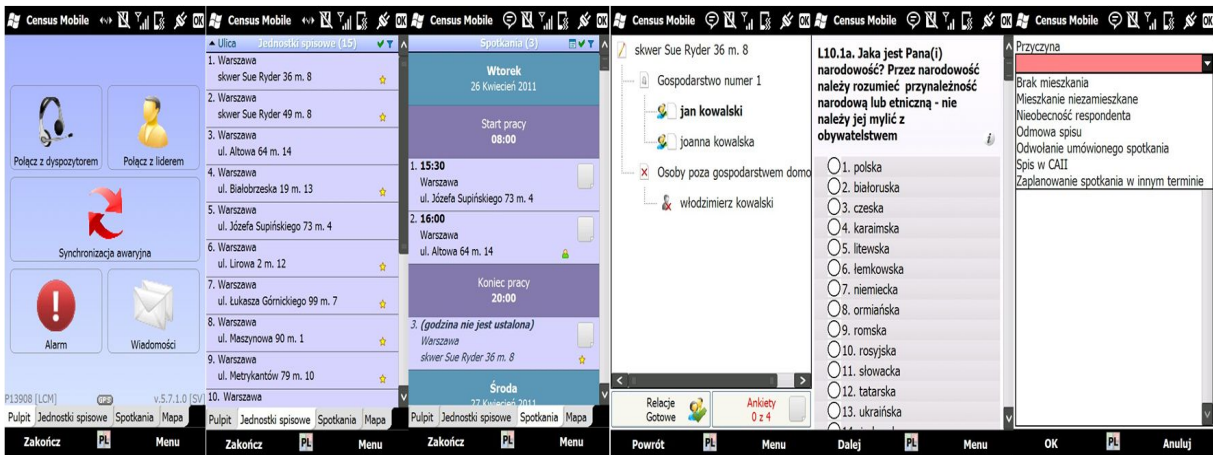


○ 조사원 업무

- 할당된 가구(거처) 방문 : 전자 조사표 작성
- 매일 동기화 작업(synchronization)
- 업무 스케줄 관련하여 관리자와 연락
- 새롭게 확인된 가구(거처) 추가

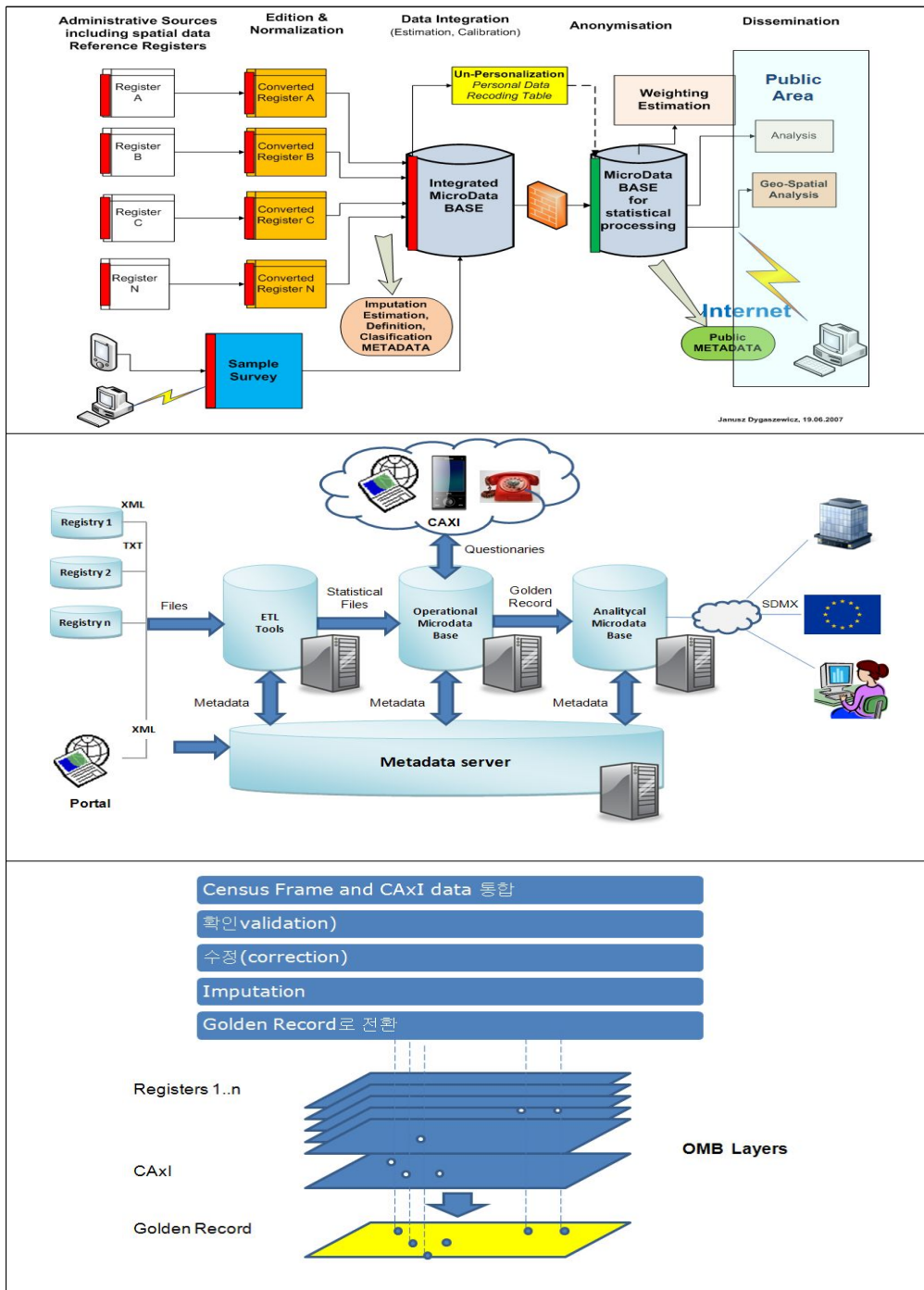
		<p><HTC 휴대용 단말기 기능></p> <ul style="list-style-type: none"> .긴급상황 알람 <ul style="list-style-type: none"> - 조사관리자에게 알람 신호 전송, 문자발송 .지도 모듈(GIS) <ul style="list-style-type: none"> -점사사진지도, 지적 데이터, 할당된 업무, 시작된 업무, 완료된 업무
--	--	--

<휴대용 단말기(HTC)의 스케줄 관리 및 조사표 지원>



○ 행정자료와 표본조사 결과를 통합하여 확인 수정, 임putation을 거쳐 Golden Record 생성

<센서스 데이터 처리 기반시설>



□ 다양한 경로를 통해 수집한 데이터 연결 방법

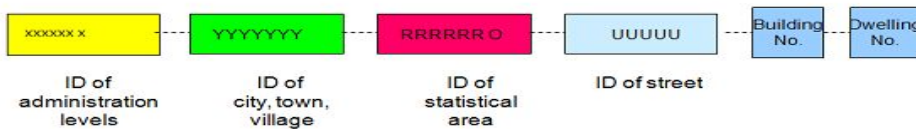
○ 데이터 연결

- 인구 등록부(Population Register)는 사람에 대한 시스템의 근간
- 개인 식별 숫자(PIN)은 연결 식별자의 기본
- 모든 파일은 인구 등록부에 정확히 매칭됨
- 연결 가능한 다른 요소들 : 이름, 성별, 주소

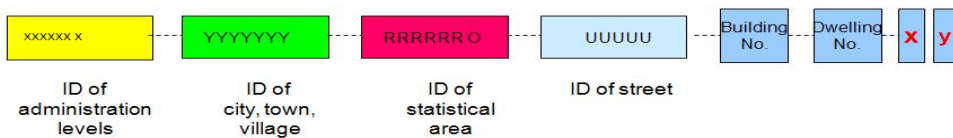
○ x, y 좌표부여

- 관련 건물의 주소 포인트로써 모든 통계적 단위 위치 데이터 제공
- 건물 위치는 지리학상 x, y 좌표를 사용하여 표시
- 행정경계가 변화 될 때 데이터를 재검토할 필요를 제거
- 포인트 할당은 선택된 어떤 지역에 통계적 데이터 처리를 쉽게 만들
- x, y 좌표 부여는 센서스의 세가지 단계에서 중요
- 준비작업, 조사요원관리, 센서스결과 다차원 공간분석

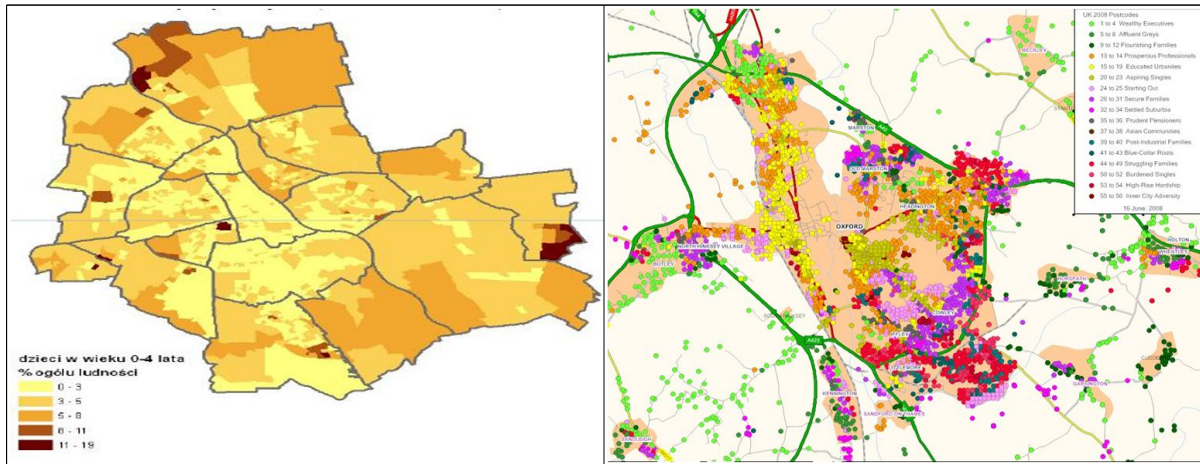
<주소 포인트 식별 시스템>



<공간 주소 포인트 식별 시스템>



<면적 할당 VS 포인트 할당 비교>



- 주소 포인트 할당은 지리적 통계데이터를 제공

5) 2011 UK 센서스 마이크로데이터 접근권에 대한 계획 Plans for Access to UK Microdata from 2011 Census

발표자 : Emma White(UK statistics office)

□ 제시된 파일(6개)

File
1. Secure: Individual Controlled Access Microdata Sample (CAMS)
2. Secure: Household CAMS
3. Safeguarded: Individual Samples of Anonymised Records (SAR)
4. Safeguarded: Household SAR
5. Public Use: Individual SAR
6. Public Use: Individual Test Dataset

- Secure 수준 : 접근 통제 마이크로데이터(개인 샘플, 가구샘플)
- Safeguarded 수준 : 익명화된 레코드(개인 샘플, 가구샘플) 제공
- Public Use 수준 : 익명화된 레코드(개인샘플), 테스트 데이터셋(개인) 제공

□ Level 3 접근권한 : Secure

File	Sample size	Geography	Licence	Availability
Individual Controlled Access Microdata Sample (CAMS)	3%	지방자치/ Northern Ireland 선거구	승인된 연구자	영국통계청 가상 마이크로데이터 연구실 (VML)
Household CAMS	3%	상동	상동	상동

□ Level 2 접근권한 : Safeguarded

File	Sample size	Geography	Licence	Availability
Individual Sample of Anonymised Records (SAR)	5%	지방자치	최종 실제 사용자 (End user licence)	특정 사용자 동의/ 다운로드
Household SAR	1%	지역 또는 국가	상동	상동

□ Level 1 접근권한 : Public Use

File	Sample size	Geography	Licence	Availability
Public Use Individual Sample of Anonymised Records	3%	지역 또는 국가	Open Government Licence	다운로드
Public Use Individual Test Dataset	상동	상동	상동	상동

* Open Government Licence : 영국왕실에서 만든 라이선스로써 제한없이 저작권자 이외 사람도 무료 이용 가능(Crown Copyright)

6) 2011 센서스 주소 등록부 구축(England & Wales) building the address register for the 2011 census

발표자 : Alistair Calder(UK)

□ 2011센서스에서 주소(addresses)의 역할

○ 조사 체계

- 2011년 센서스에서 조사표는 모든 거주자들의 주소로 우편 발송되고 집단시설가구에 대해서는 별도로 직접 방문 조사를 실시
- 주소는 조사표 우편발송 및 회수를 추적하는데 중요(각각 조사표는 바코드와 스캔)

- 주소 등록부는 이 후속작업(follow-up)*의 우선순위와 대상을 정하는데 사용되며 행방불명된 조사표를 추정 가능하게 함

* 회수되지 않은 가구는 조사원이 조사표를 작성하도록 독려

□ 주소등록부 구축 과정

○ 결합(Matching)

- 주소등록부의 핵심은 주요 국가적 데이터셋들을 결합하는 것
- 영국우정공사 우편번호 주소파일(PAF), 지방정부에 의해 작성된 토지부동산지명(NLPG), 기준격자를 일부 포함 (Address Layer2)해서 자동화된 주소 매칭 수행
- 주소목록은 주소 구조와 분류를 일관성있게 표준화
- AL2와 PAF는 고유한 주소 식별자로 병합
- 정밀과 퍼지(exact and fuzzy) 매치방법을 포함한 알고리즘을 이용하여 원문 그대로 결합
- 원문 그대로 결합 불가능한 나머지 주소들은 공간적으로 결합

- 매치되지 않은 주소들은 자료 제공자인 지방정부와 우정공사에서
검사 및 해결

○ 현장점검

- 2010년 여름 360만개 주소(England and Wales의 주소
15% 정도) 240명을 고용하여 현장점검 수행
- 매치되지 않는 주소들과 두개 이상의 다중 점유(multi-occupancy)
주소를 주로 대상으로 점검

○ Evidence base 와 규칙

- 여러 가지 자료를 사용함으로써 어떤 주소가 포함 또는 제거
되어야 하는지 선택의 문제 발생
- Business rules를 적용하지만 조심스럽게 접근

7) 캐나다 센서스의 인터넷 조사 Internet data collection in the canadian census of population

발표자 : Marc Hamel(canada)

□ 2011년 목표

- 인터넷조사 응답률 : 40%(2006년 18%)
- 80% 메일발송(2006년 거쳐 중 73% 메일발송)
- 무응답 없이 인터넷응답을 최대화하는 조사방법 적용

□ Wave 방법론

- 다양한 접근 방법으로 조사 응답을 독려
 - * 진행 순서 : Wave1 → Wave2 → Wave3 → 무응답 대상 조사 (follow-up)
 - Wave1 : 조사표 대신에 편지를 발송
 - Wave2 : Wave 1을 통해 응답이 감소하기 시작할 때 응답하지 않은 사람들에게 2차 메일발송
 - Wave3 : Wave 2를 통한 응답이 감소할 때 쯤 여전히 응답하지 않는 사람들에게 조사표를 발송
 - 조사표 발송을 한 후 감소할 때 무응답 후속절차를 시작

□ 조사 방법(Collection Methodology)

Wave 단계	메일 발송 지역(거주지의 79%)	나머지 지역(19%)
1	센서스 시작 7일 전(5.3) 75% 편지 발송(그룹1) 25% 조사표 발송(그룹2)	센서스 시작 8일 전(5.2) 전통적인 조사표 발송
2	센서스 시작(5.10) 편지 발송 거주지 중 무응답 지역에 편지 발송	센서스 시작(5.10) (수신자 주소/성명이 없는) 거주 지에 조사요청 카드 배송
3	센서스 시작 8일 이후 Wave 1 편지(그룹1)를 받은 무응답 거주지에 조사표 발송 Wave 1(그룹 2)는 음성 방송 메시지 전송(MCS 상 전화번호 보유)	-
4	센서스 시작 22일 이후(6.1) 무응답 후속조치(조사원) 실시 - 어떤 연락도 안되면 방문알 림 카드 또는 전화자동응답에 메시지 남김	센서스 시작 10일 이후(5.20) 무응답 후속조치(조사원) 실시

- 조사표 요청 시스템(QRS) : 보안접근코드(SAC)를 이용하여 종이조사표 요청할 수 있는 전화 서비스
- 마스터 컨트롤 시스템(MCS) : 메일발송 이나 배달 할 수 있는 거주지 리스트

□ 인터넷 조사결과

응답률	계획	결과
조사 응답률	98%	98.1%
인터넷 조사 응답률	44%	54.4%
자기기입	80%	84.1%

□ 조사방법별 응답 모드

Collection method	Response Mode				Non-response	Total
	Mail	Internet	CHL	NRFU		
Letter	16.3%	71.6%	0.7%	9.1%	2.3%	100.0%
Questionnaire	50.1%	25.8%	0.8%	20.0%	3.4%	100.0%

* CHL : Census Help Line(민원전화서비스)

**NRFU : Non-Response Follow-Up(무응답 대상 후속 절차)

□ 항목별 무응답 비율

2011 QUESTION number	QUESTIONS	2006 non-response rates	2011 non-response rates	Difference (2011 - 2006)
CENSUS				
2	Sex	1.4	0.8	-0.6
3	Date of Birth	0.7	0.5	-0.2
4	Marital Status	2.9	2.5	-0.4
5	Common Law Status	6.7	4.4	-2.3
6	Relationship to Person 1	1.7	1.4	-0.3
7	Official Languages	N/A	1.6	
8A	Home Language	N/A	1.8	
8B	Other Languages	N/A	2.0	
9	Mother Tongue	2.0	1.9	-0.1
10	92 Years Consent	8.1	5.5	-2.6

□ 교훈

- Wave 방법론은 효과적이었음(높은 인터넷 조사율)
- Wave 간 시간이 촉박
- 메일 발송 파일의 정확도가 관건
- 일부 캐나다는 전통적인 메일 서비스를 이용하지 않음

IV. 참고자료

1. UNECE–UNFPA Training Workshop on Censuses Using Registers

(2012년 5월 21일)

Welcome and introduction(UNECE, UNFPA, Harald Utne)
Why register-based statistics?(Statistics Netherlands)
Definition of basic concepts(Statistics Netherlands)
Individual based statistics and register development(Statistics Norway)
General preconditions–Requirements to data from administrative sources(Statistics Norway)
overview of approaches to censuses using registers(UNECE)
National presentations 1. Belarus 2, Rep. of Moldova, 3. Croatia, 4. Serbia
Population and housing censuses in the Nordic countries
The Dutch Virtual Census based on registers and existing sample surveys(Statistics Netherlands)

※ 참고자료 : Register-based statistics in the Nordic countries(Review of best practices with focus on population and social statistics)

2. Expert Group Meeting on Censuses Using Registers

(2012년 5월 22-23일)

Information session on the 2010 Round of Population and Housing Censuses
overview of 2010 round of censuses in the UNECE region with focus on the use of registers(UNECE)
Session 1: Experiences with the use of registers in censuses
Comparing approaches of different(partly) register-based countries (Netherlands)
A general approach to the importance and use of registers in the Spanish Census(Spain)
Model of transformation administrative data to statistical data(Poland)
Using registers for improving the quality of census data and estimated population size(Estonia)
The 2011 population and housing census in Turkey(Turkey)
Session 2: Should we use registers for our census?
Beyond 2011-The future of population statistics?(United Kingdom)
The canadian census of population : A review in preparation for 2016(Canada)
A census is a census is a census? (Eurostat)
Session 3 : Availability, completeness and quality of data from registers and other sources
Quality of registers used for the Dutch census(Netherlands)
The use of population registers in the 15th Italian census : challenges and preliminary evidence(Italy)
Register-based census 2011 in Slovenia - some quality aspects(Slovenia)
Use of Survey data(LFS) to evaluate the quality of register-based census in Finland(Finland)
The register of addresses and buildings - a combination of different registers(Germany)
The population registers in Germany - the main data source in the 2011 Census (Germany)
Quality assessment of register-based census data in Austria(Austria)
Session 4: Methodology for estimating information missing in registers
Methodology used for estimating Census tables based on incomplete information(Netherlands)

3. Group of Experts on Population and Housing Censuses

(2012년 5월 24-25일)

item 1b. Information session on the 2010 Round of Population and Housing Censuses
Overview of 2010 round of censuses in the UNECE region with focus on the use of registers(UNECE)
New developments in the 2010 round of censuses in terms of census methods and use of modern technologies, from a global perspective(UNSD)
item 2. Full enumeration versus sample survey
The introduction of the American Community Survey and its benefits to the 2010 Census(USA)
Full enumeration or sample survey?(CIS-STAT)
Modern census in Poland(Poland)
Transition from traditional census to sample survey?(Slovakia)
item3. Access to microdata
Plans for access to United Kingdom microdata from 2011 census(UK)
Place of Work Census Anonymised Records(Ireland)
IPUMS-International and Integrated European Census' Microdata Projects reduce risks of managing trans-border access and add significant value
item4. Address/dwelling listing
Experience with address lists and other preliminary work(Italy)
Building the address register for the 2011 Census-England and Wales(UK)
Dwelling and Building Register Based on Municipal Taxation List-Quality and Distinctiveness(Israel)
2010 Population and Housing census in Mexico. Embedding the building list as a basis for the generation of the dwelling inventory(Mexico)
item5. Internet data collection
Internet data collection in the Canadian census of population(Canada)
The web-based information system of the Italian Population(Italy)
2011 Census in Portugal - Implementation of the e-Census system(Portugal)
2011 Census for England and Wales; Internet Data Capture(United Kingdom)
Internet data collection in the 2011 Population and Housing Census in the Czech Republic (Czech Republic)
Internet data collection in 2011. Population and Housing census in Lithuania (Lithuania)
Development of internet survey method in Japan(Japan)