
종합모집단 DB 구축 관련 핀란드, 스위스 사례 수집 결과보고

2014. 12

조사관리국
행정자료관리과

목 차

I. 개요	1
II. 주요 내용	1
- 핀란드, 스위스 통계청 수집결과	
III. 세부 내용	2
<핀란드 통계청>	2
1. 핀란드 통계청의 등록(행정자료)기반 통계생산 체계	
2. 인구·가구 센서스의 행정(등록)자료 활용방법	
3. 통계 생산방식 변화에 따른 품질평가	
4. 사회분야 통계의 Data Warehouse	
5. 행정자료를 활용한 주요 고용형태 등록부 추정	
<스위스 통계청>	13
1. 스위스의 센서스 시스템(가구와 개인 통계의 포괄적인 시스템)	
2. 행정자료의 통합과 경제 구조통계 소개	
3. 사회보험(OASI)과 사업자등록(BR) 연계 및 이슈사항	
4. 부가가치세 자료의 연계 및 이슈사항	
VI. 시사점 및 향후 추진계획	19

□ 개 요

- (목적) 행정자료의 체계적이고 효율적인 관리 및 활용을 위한 종합모집단 DB구축 사업이 추진 중에 있어 오랜 기간 행정자료를 활용한 선진국(핀란드, 스위스) 사례를 벤치마킹할 필요
- (일시 및 장소) 2014. 10. 26 ~ 11. 1. / 핀란드, 스위스 통계청
- (출장자) 김태준 사무관(행정통계과), 이재진 주무관(행정자료관리과)

□ 주요 내용

<핀란드 통계청>

1. 핀란드 통계청의 등록기반 통계생산 체계
2. 인구·가구 센서스의 행정(등록)자료 활용방법
3. 통계 생산방식 변화(조사 → 행정(등록)기반)에 따른 품질평가
4. 사회분야 통계의 Data Warehouse
5. 행정자료를 활용한 주요 고용형태 등록부 추정(설명자료)

<스위스 통계청>

1. 스위스의 센서스 시스템(가구와 개인 통계의 포괄적인 시스템)
2. 행정자료의 통합과 경제 구조통계 소개
3. 사회보험(OASI)과 사업자등록(BR) 연계 및 이슈사항
4. 부가가치세 자료의 연계 및 이슈사항

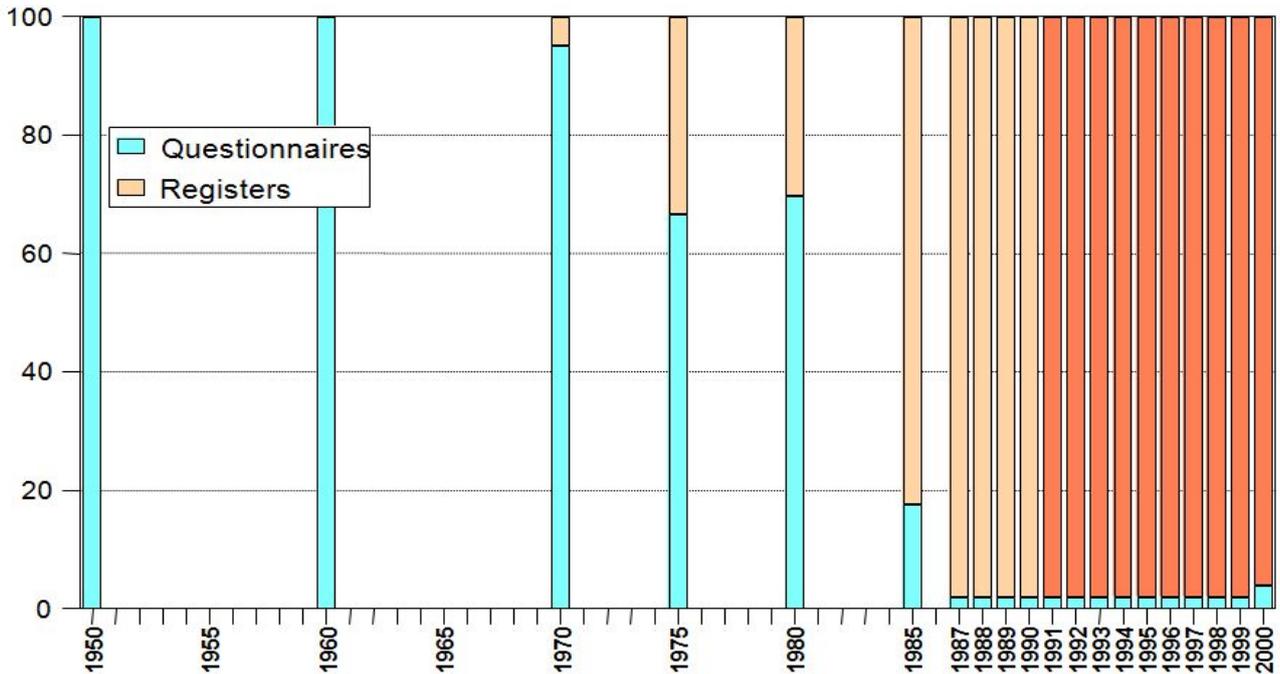
□ 세부 내용

<핀란드 통계청>

1. 핀란드 통계청의 등록(행정자료)기반 통계생산 체계

- (연혁 및 현황) 핀란드는 1970년 인구센서스 이후 사회.인구통계부터 경제통계까지 활용하고 전체자료*의 약96%는 행정자료에서 수집되며 수집 곤란한 부분은 조사자료와 상호 보완적으로 활용중임
 - * 통계 조사단위(unit) × 조사단위별 항목수(variables)
- (배경) 핀란드가 행정자료 활용이 가능하게 된 원인들은 하나의 개인식별번호 사용, 정부차원의 전체 건축물 DB에 대한 관심 및 구축, 조사비용 절감, 국민 응답부담 경감, 법적 근거 등임
 - 행정(등록)자료를 사용하는 주요 이유는 비용절감, 응답부담 경감, 전체 모집단들을 갖기 위한 것임
 - 핀란드 통계법은 존재하는 모든 자료를 사용하게 강제하였고, 연방정부와 사회 보장 관련 위원회들이 통계청에 자료를 제공하도록 의무화
- (활용방법) 행정자료 활용법은 등록자료 직접사용, 등록부 추정(Register estimation), 조사자료와 등록자료를 통합한 추가정보 생성, 표본틀로 사용, 무응답 처리, 항목대체, 무응답 구조 파악 등
- (예시_인구 관련 통계) 핀란드는 중앙인구등록센터 설립(1969), 인구센서스에 행정(등록)자료 활용 시작(1970), 건물.거주지 행정(등록)자료 활용(1980), 행정(등록)기반 고용통계 시스템 활용 통계공표(1989), 세계에서 2번째로 행정(등록)기반 센서스 실시(1990)
 - 모집단 센서스들의 행정(등록)자료 활용 발전은 20년 동안의 단계별 업무 추진과 재정부 등 관계기관의 강력한 지원으로 가능

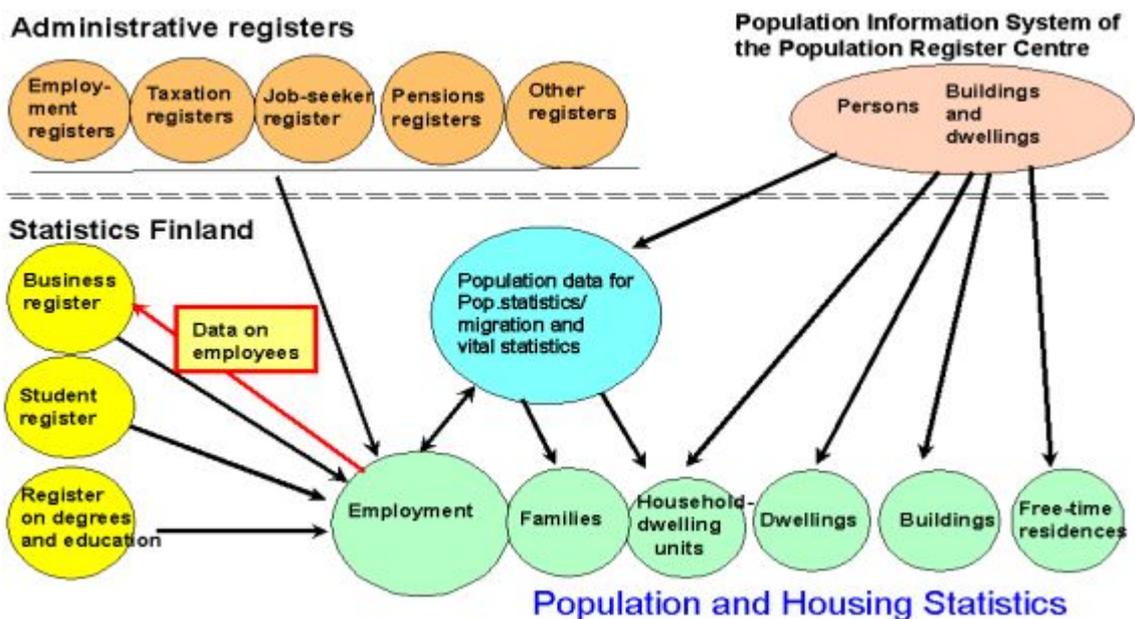
Census data 1950-2000 according to the data source



〈그림 1. 센서스 데이터의 원천자료 1950 ~ 2000〉

2. 인구·가구 센서스의 행정(등록)자료 활용방법

- (자료 입수) 인구·주택센서스는 행정자료를 활용한 대표적인 사례이고, 다른 기관(5~6개)에서 행정(등록)자료를 입수하고 인구등록센터(별도기관)에서 개인, 건물, 거주지 행정(등록)자료를 입수

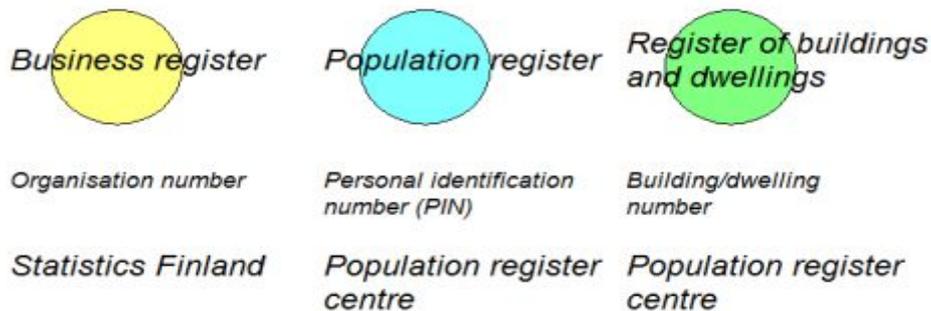


〈표 2. 인구·주택센서스의 행정(등록)자료, 통계자료 연계도〉

○ (등록기반 통계시스템) 통계 기본 등록부(그림 3)를 다음과 같은 절차에 따라 연계, 코드를 부여(그림 4)

- ① 건물번호 부여 및 거주지와 연계
- ② 거주지번호* 부여하고 거주지와 거주지-세대, 가족, 개인을 연계
* 거주지별로 부여하나 시계열 유지는 안되고, 통계 생산할 때마다 생성
- ③ 기업체번호(국세청 부여)를 개인, 사업체에 부여 및 연계
- ④ 사업체번호(통계청 부여)를 기업체, 개인에 부여 및 연계
- ⑤ 각각의 단위에 주소정보를 부여

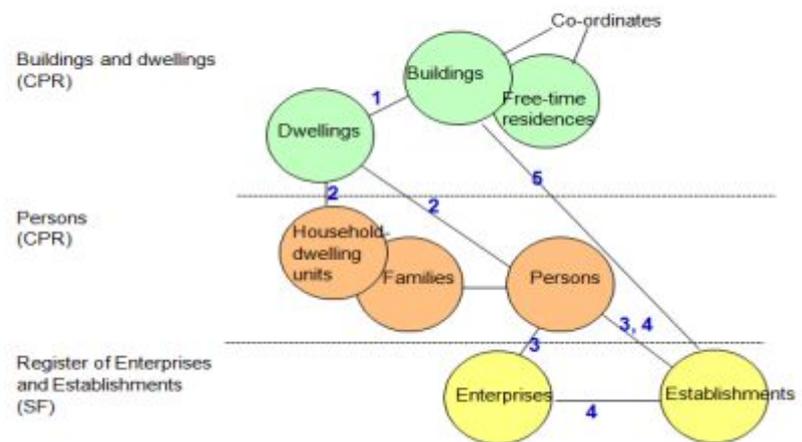
※ 경제분야 통계에서는 기업체의 사업체 구조를 조사하면서 실제 개인이 어느 사업체에서 일하고 있는지 조사(일정규모 이상 기업체)



〈그림 3. 핀란드 인구.사회통계 기본 등록부〉

The basic units of register-based statistical system

- 1 Building code
- 2 Domicile code
- 3 Enterprise number
- 4 Establishment number
- 5 Address



〈그림 4. 핀란드 인구.사회 등록기반 통계 시스템 연계〉

3. 통계 생산방식 변화(조사 → 행정(등록)기반)에 따른 품질평가

- (기본방향) 핀란드 통계청은 행정(등록)기반 센서스 시스템 적용을 결정하기 전에 행정(등록)기반 자료들의 신뢰성을 관찰하고 분석하는 연구를 실시
- (품질평가기준) 통계 생산방식 변경에 따른 품질평가의 기준
 - ① 타당성 : 현재와 미래 사용자들의 요구를 충족 여부
 - ② 정확성 : 통계 추정량이 실제 값에 근접 여부
 - ③ 신속성 : 실제 사건(현상)과 자료가 실제 활용되는 시점의 차이
 - ④ 시간업수 : 자료 공표 시점과 공표 목표시점 사이의 시간 차이
 - ⑤ 기타 : 사전 품질점검, 메타데이터, 다른 행정(등록)자료와 일치 등
- (품질측정방법) 입수되는 자료의 품질은 아래의 기준으로 판단
 - 자료를 읽을 수 있는지, 식별정보를 가지고 있는지, 요청 항목이 있는지, 자료는 정확한지, 다른 행정(등록)자료와 비교 가능한지, 전년 또는 지난 자료와 비교 가능한지 등

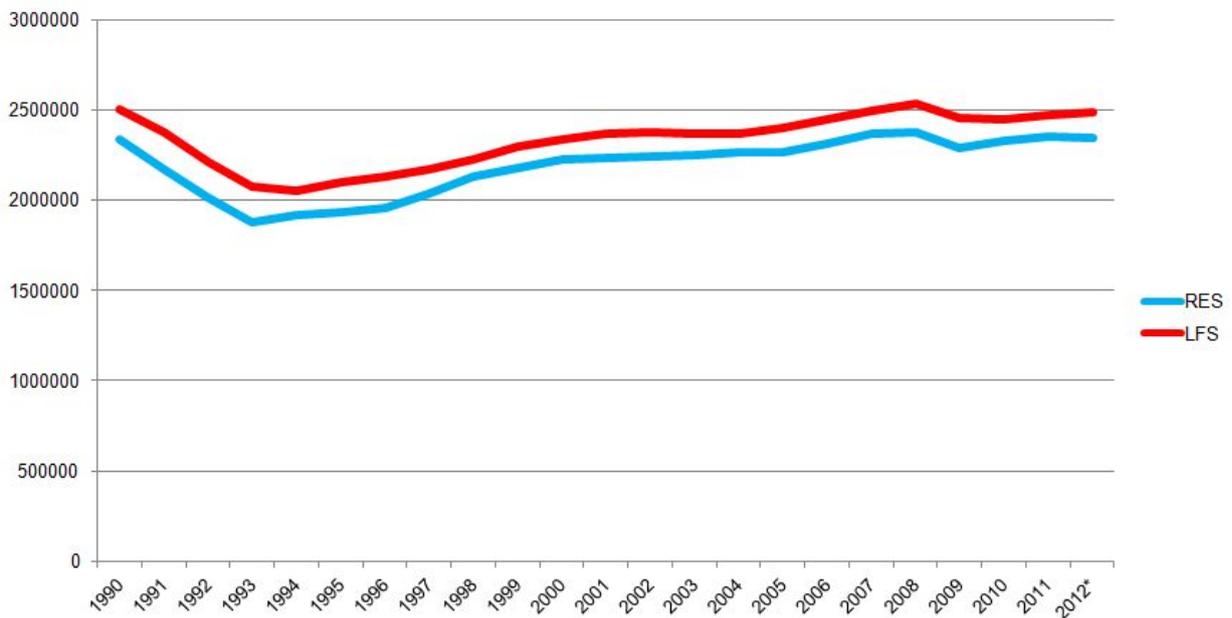
⇒ 자료에 큰 문제가 있으면 원 행정소유 기관에 재요청
- (조사에서 등록기반으로 변화) 다른 현장조사의 같은 조사대상이 응답한 값들은 다르게 나타나기도 하는데 그 차이는 현장조사와 행정(등록)기반 자료의 차이와 유사
 - 고용통계 중 조사기반 LFS(Labour Force Survey)와 행정(등록)기반 RES(Register-based Employment Statistics)를 같이 공표
 - 핀란드 통계청은 개인들에 대한 자료와 주요 활용 항목 자료들은 정기적으로 자료 품질을 연구

Main type of activity of the population according to the registers and questionnaire in 1985 census (percentages)

Register	Questionnaire					
	Total	Em- ployed	Unem- ployed	Stu- dents	Pensio- ners	Others
Total	100,0	57,5	3,5	9,4	23,7	5,9
Employed	100,0	93,7	0,6	2,1	1,0	2,6
Unemployed	100,0	17,7	70,3	1,3	2,6	8,1
Students	100,0	3,5	1,1	90,7	0,2	4,5
Pensioners	100,0	2,4	0,2	0,2	94,5	2,9
Others	100,0	24,6	9,7	5,1	3,3	57,2

〈그림 5. 고용통계 관련 행정자료와 조사자료의 주요활동별 일치율〉

Employed according to the Labour Force Survey (LFS) and Register-based Employment Statistics (RES)



〈그림 6. 조사기반 LFS와 행정(등록)기반 RES의 종사자수 추이〉

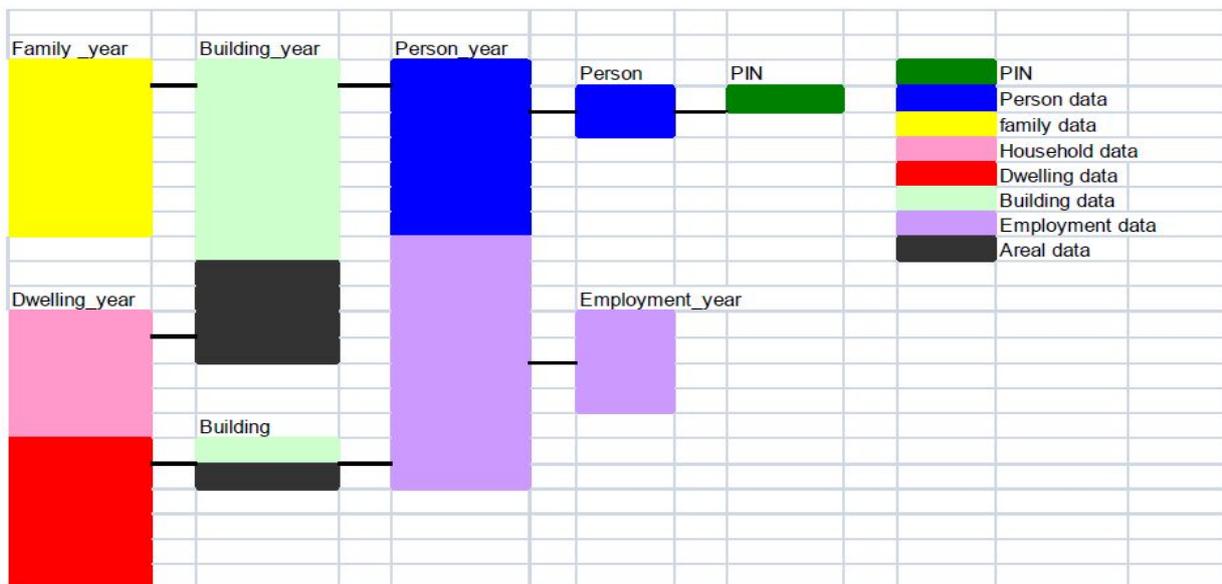
4. 사회분야 통계의 Data Warehouse

- (개선사항) DW에 통계 data files을 통합하여 동일 환경, 산출물의 쉽고 빠른 생산, 동일시점에 여러 단위(개인, 가족, 가구 등) 자료통합, 동일 자료의 장기간 파악, 표본추출, 새로운 자료생성이 가능
- (DW) File 형태의 자료는 각 자료별로 여러 특성항목을 포함하고 있으나 DW는 특성별로 Database를 구성하고 Key로 연계

Data files - Before

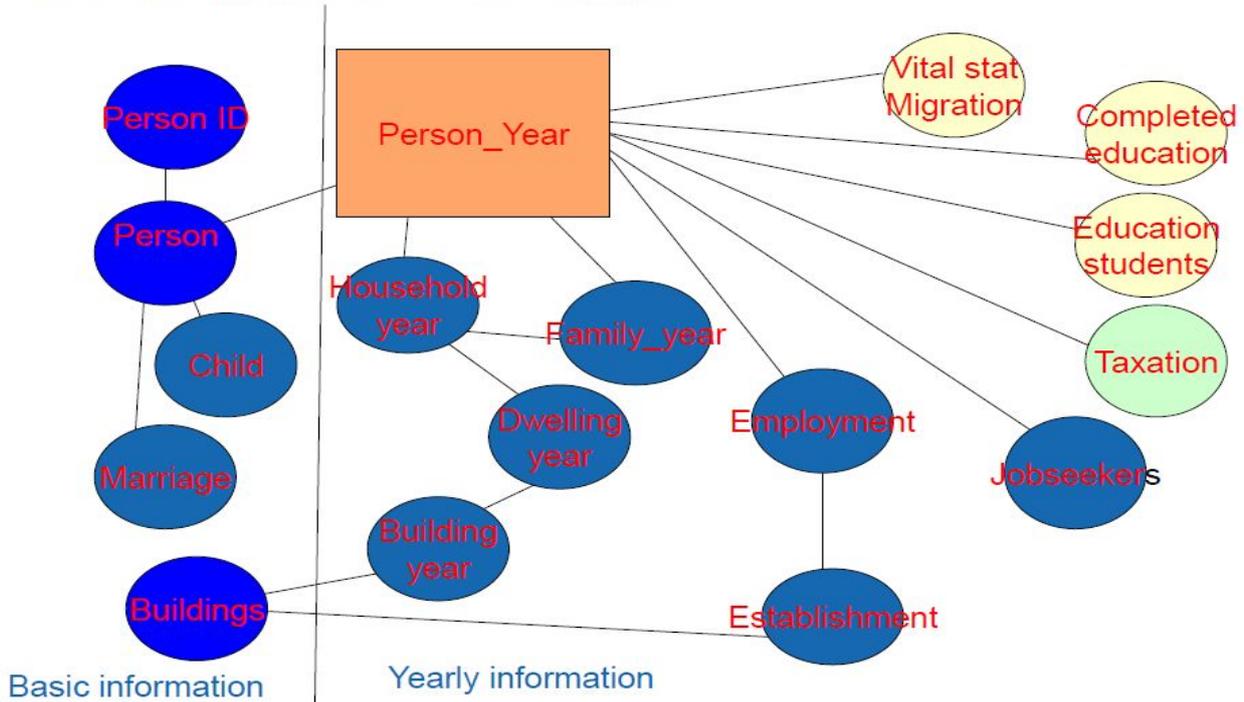


Data Warehouse



<그림 7. Data files에서 Data Warehouse로 변화>

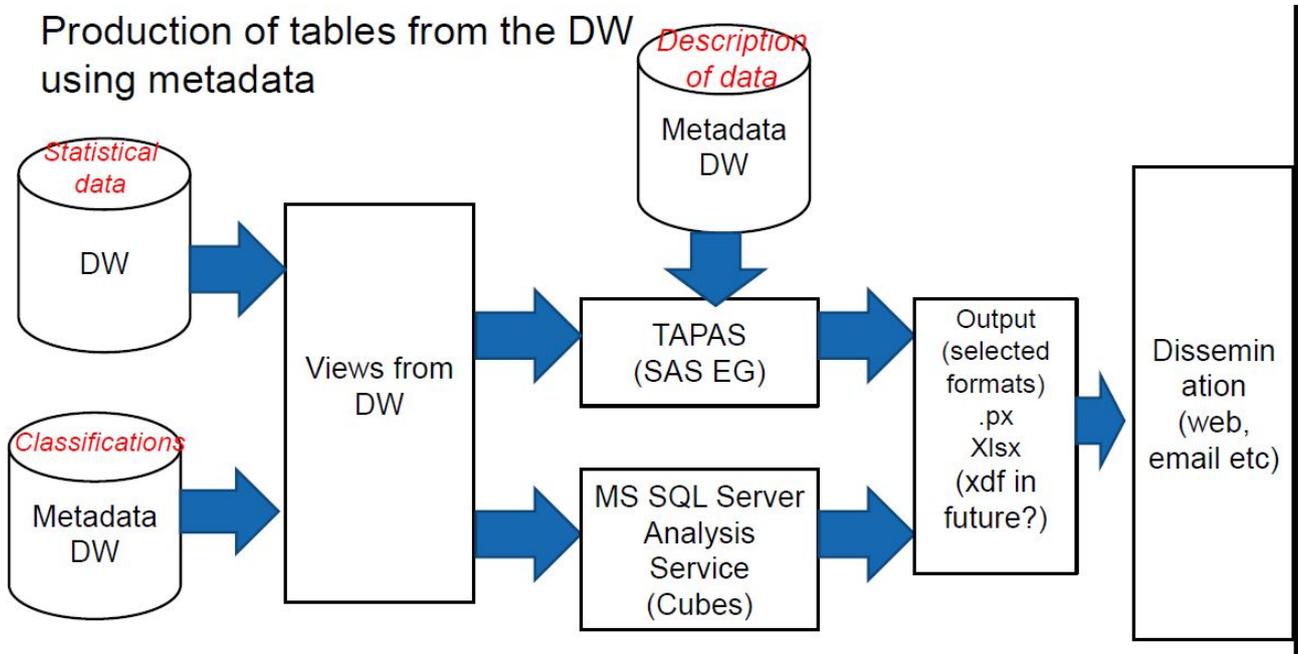
Datawarehouse - Structure



<그림 8. Data Warehouse 구성도>

- (통계작성) DW에서 메타데이터의 자료설명*과 분류기준**을 활용하여 통계 테이블을 만드는 구성도(그림 9)

Production of tables from the DW using metadata

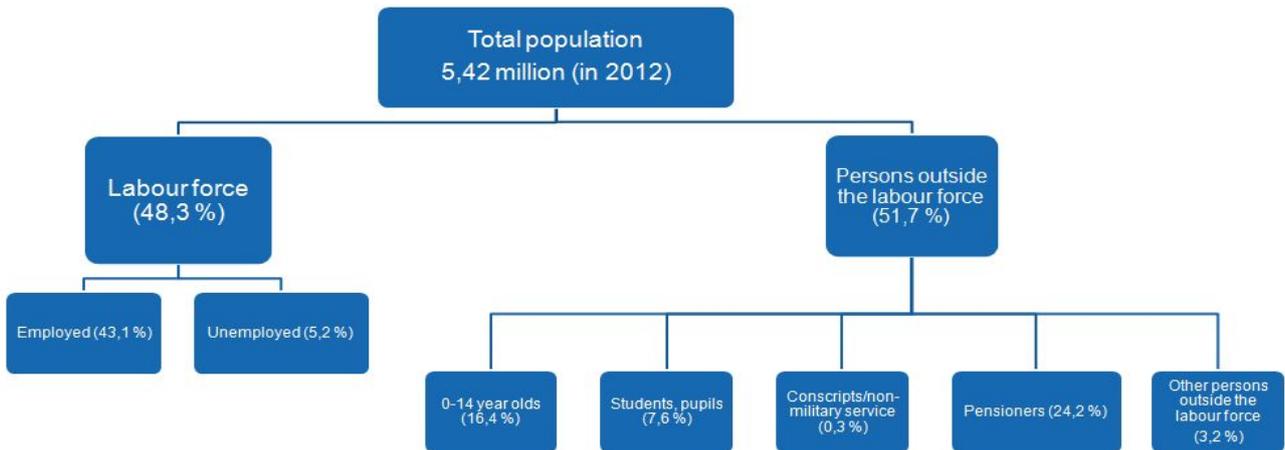


<그림 9. Data Warehouse 통계생산 과정>

- * Description of data : 각 통계조사별로 항목에 대한 기준을 설정
- ** Classifications : 각 통계조사의 항목별 기준에 따라 자료를 생성

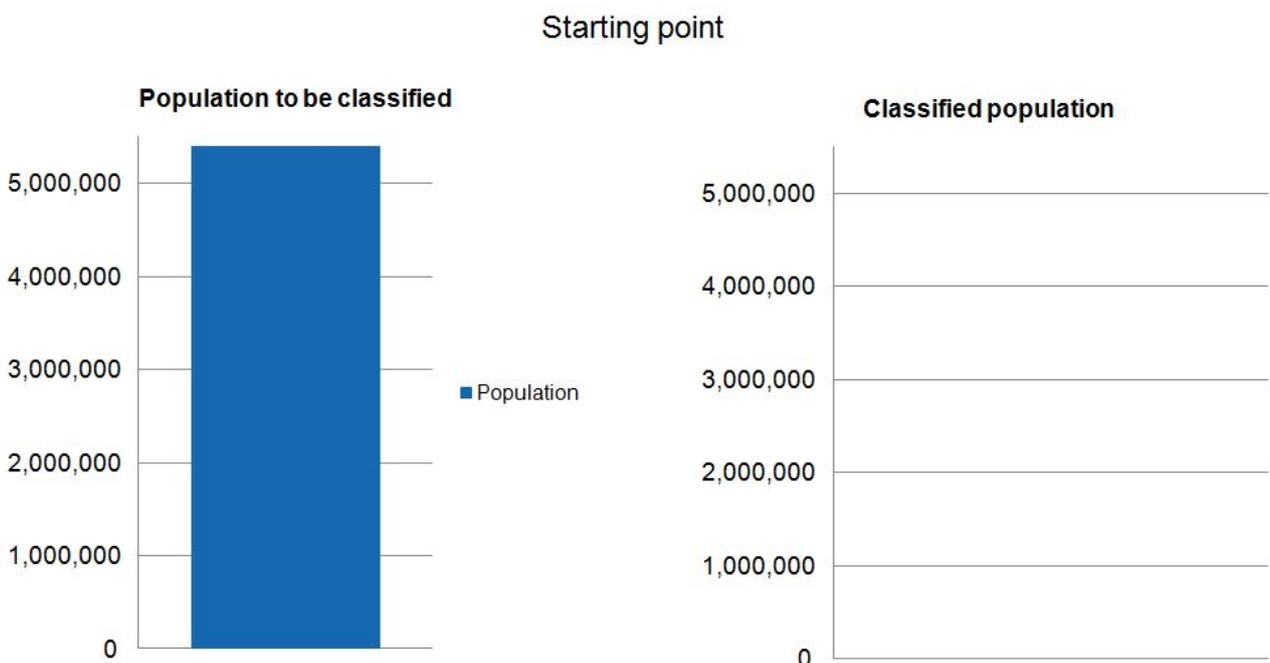
5. 행정자료를 활용한 주요 고용형태 등록부 추정(설명자료)

- (등록부 추정) 등록부 추정은 행정(등록)자료의 변수로부터 새로운 변수를 생성해 내는 작업이고 행정(등록) 정보만으로 각 통계단위의 새로운 변수의 값을 추정하는게 목적

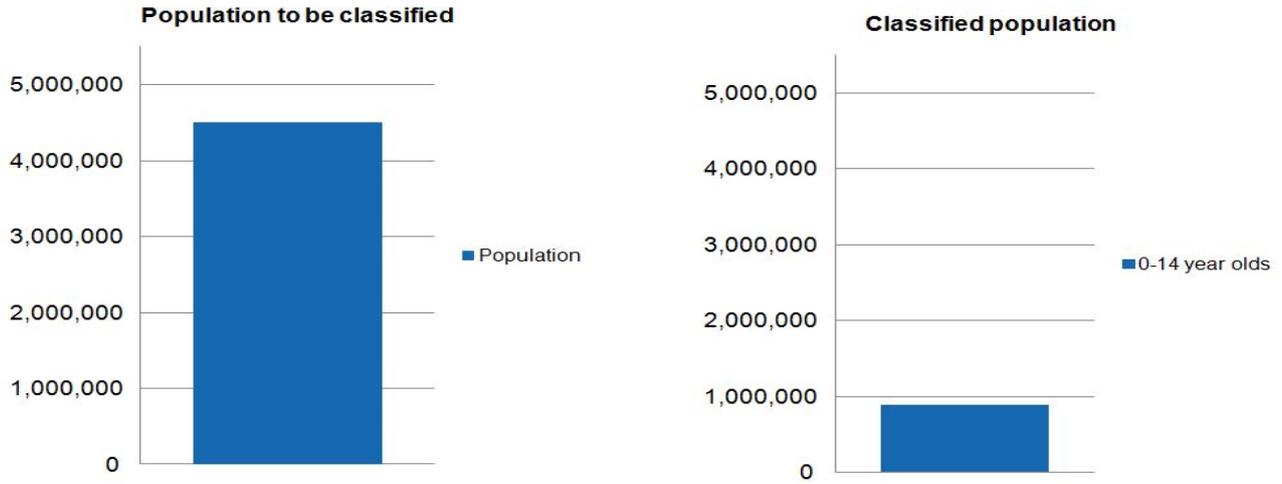


〈그림 10. 행정자료를 활용한 고용형태별 비율 생성 결과표〉

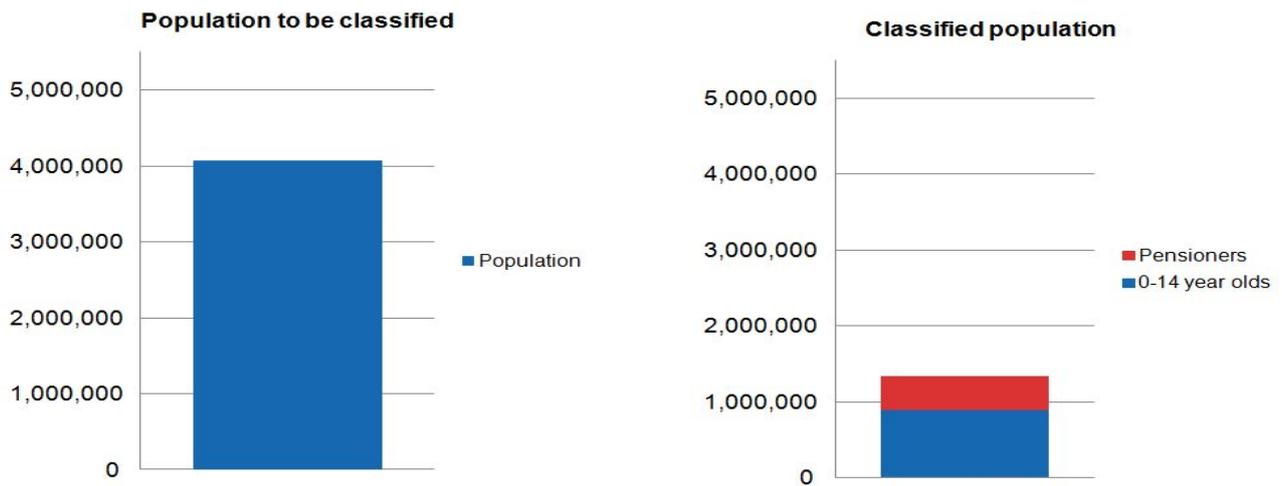
- (단계별 추정) 전체 인구를 행정자료를 통해 파악할 수 있는 고용 기준에 따라 인구분포를 작성하는 과정을 그래프로 설명



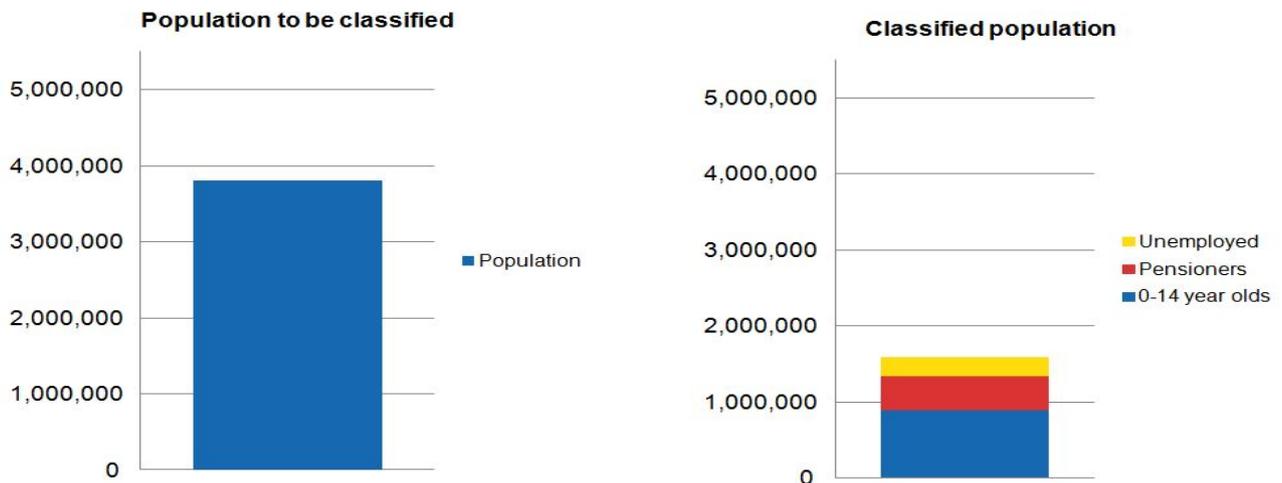
Step 1. All persons under fifteen years of age are categorized as '0-14 year olds'



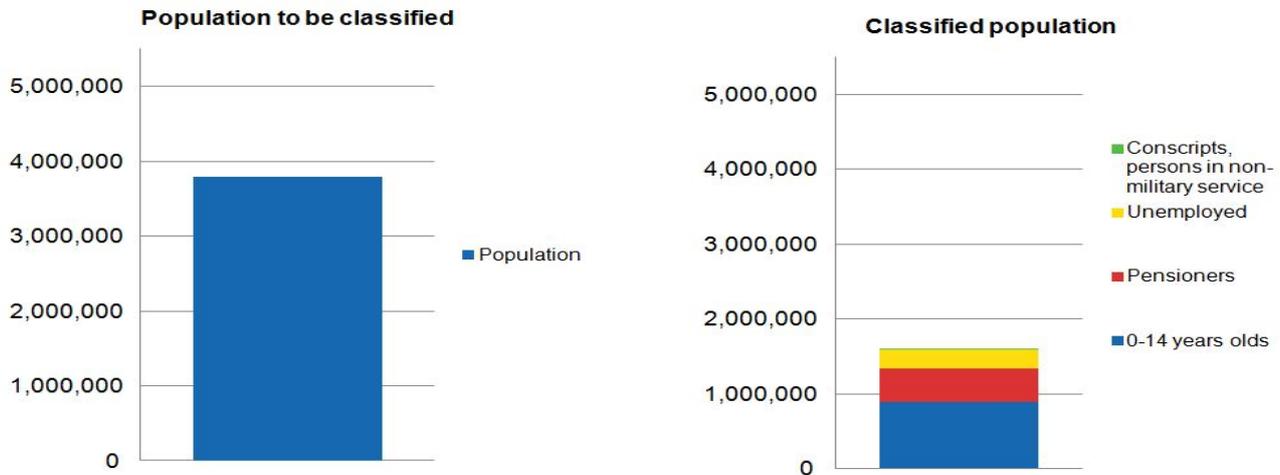
Step 2. All persons over the age of 74 are categorized as 'pensioners'



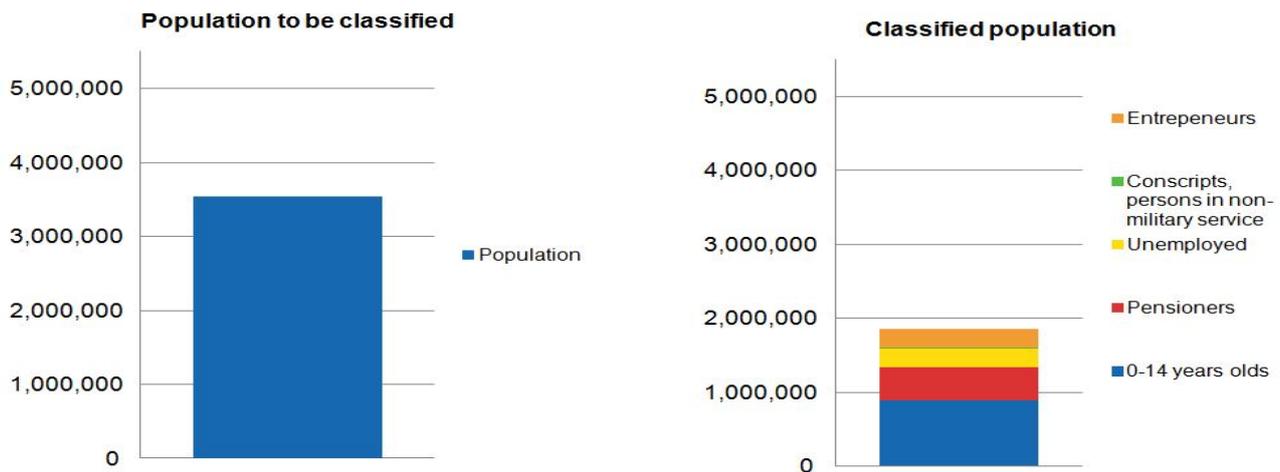
Step 3. Persons who unemployed jobseekers according to the Ministry of Employment and Economy's register are categorized as 'unemployed'



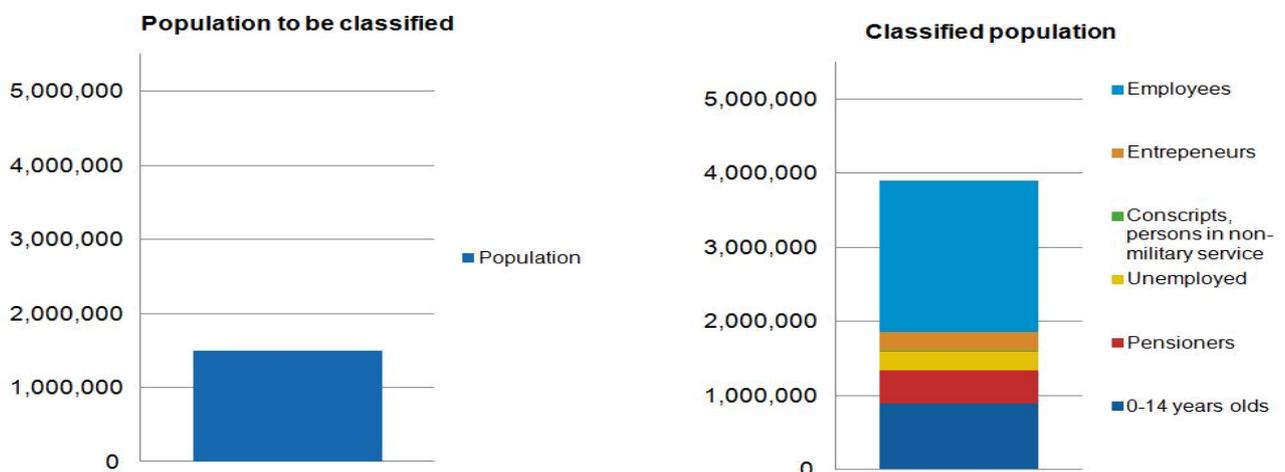
Step 4. Persons in military or non-military service are categorized as 'conscripts or persons in non-military service'



Steps 5-7. A person is defined as 'entrepreneur' if one of following rules apply:
 5. Valid self-employment pension's insurance and no other form of pension insurance
 6. Valid self-employment **and** employment pension insurance, but higher entrepreneurial income
 7. Entrepreneurial income is higher than 1017 EUR and no valid employment relationship

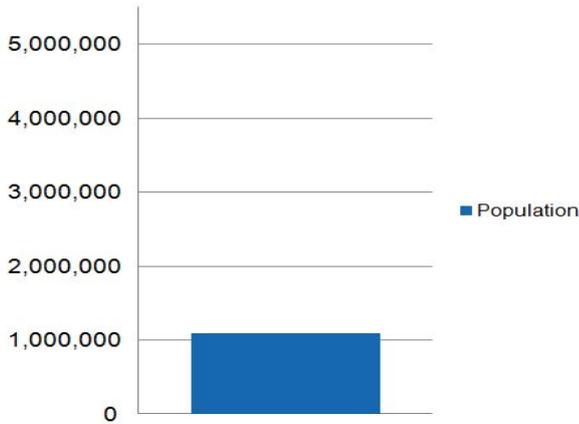


Steps 8-9: A person is defined as 'employee' if one of following rules apply:
 8. Valid employment pension insurance and earned income (>0)
 9. No employment relationship but is employed with labour policy measures

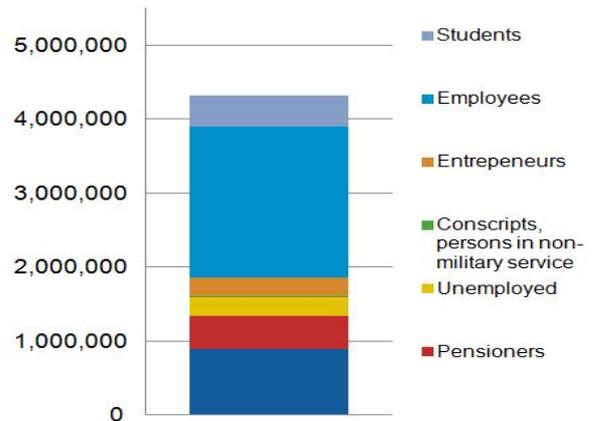


- Steps 10-13. Persons are categorized as 'students' if one of the following rules apply:
- 10. Persons who according to the student register study full-time in the autumn term
 - 11. Those 15-year-olds who were not classified into any other group at earlier stages
 - 12. Persons in labour market training in the last week of the year
 - 13. Persons who have received study aid in the autumn term

Population to be classified



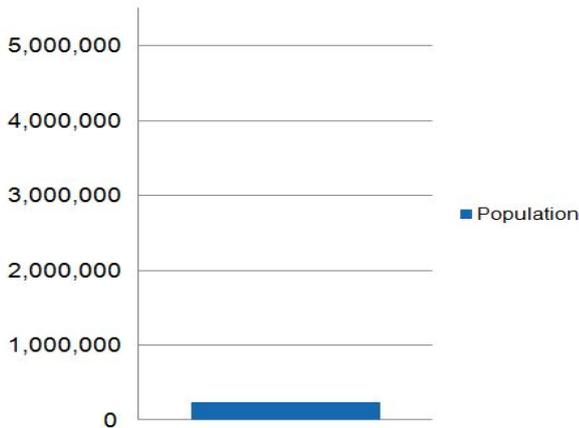
Classified population



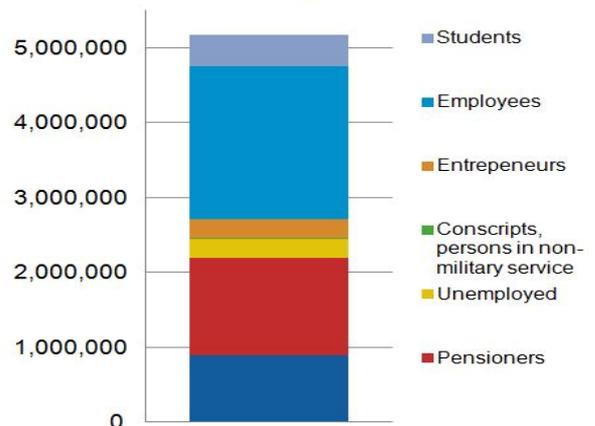
- Steps 14-15. Persons are classified as 'pensioners' if either of the following rules apply:

- 14. Persons who receive old-age, disability, unemployment or special farmers' pension
- 15. All those whose pension is higher than their income from work and above EUR 665 and the pension is not a survivor's pension or part-time pension

Population to be classified

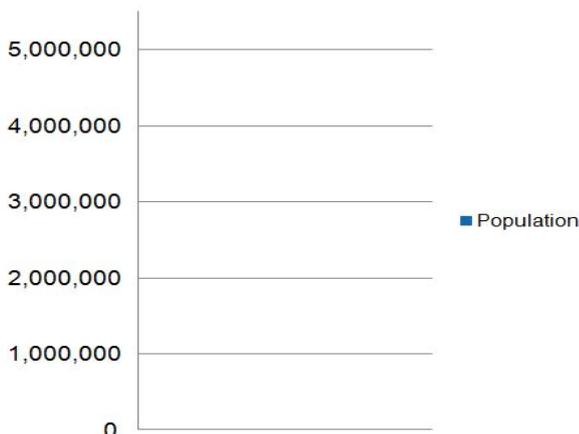


Classified population

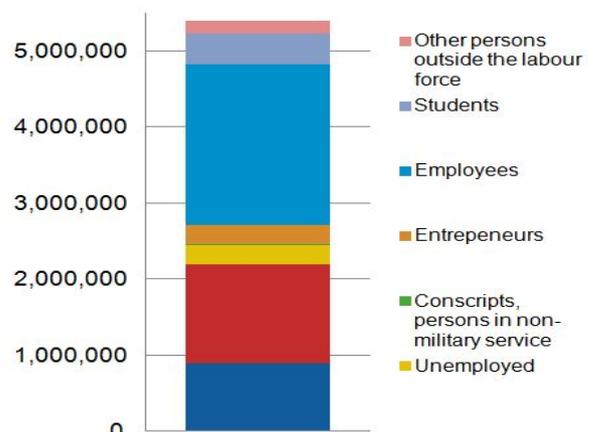


Step 16. Persons who do not meet any of the above criteria

Population to be classified



Classified population

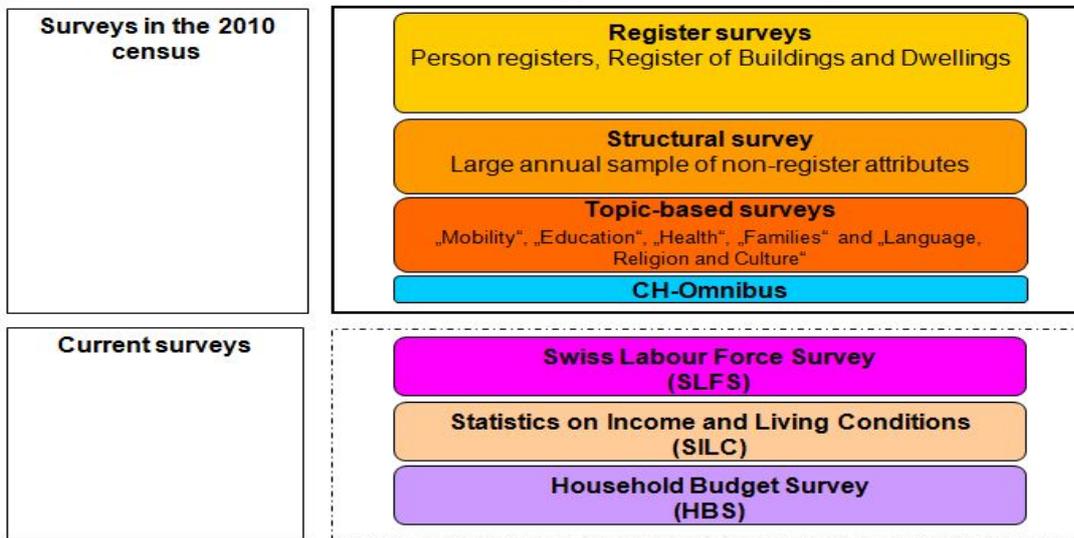


<스위스 통계청>

1. 스위스의 센서스 시스템(가구와 개인 통계의 포괄적인 시스템)

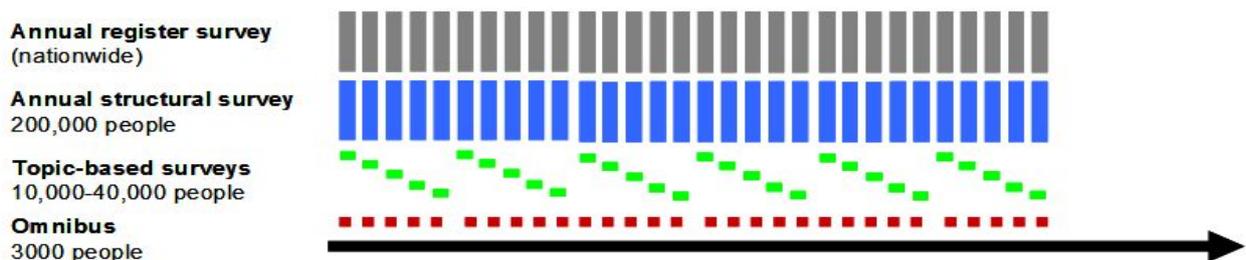
- (기본내용) 새로운 센서스는 행정자료와 등록자료를 기반으로 하고, 추가적으로 필요한 정보들은 보완적인 표본조사에 의해 채워지며, 자료수집부터 결과공표까지의 모든 작업은 포괄적인 시스템에서 가구, 개인 통계들의 통합과 연계로 이루어짐

Integrated system for household and person statistics (SHAPE)



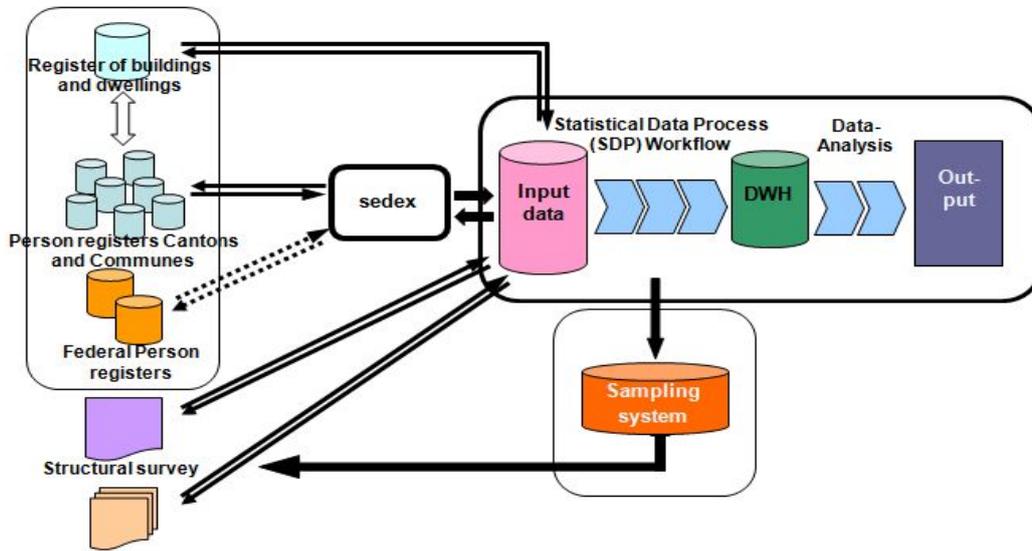
<그림 11. 스위스 가구.개인 통계 포괄시스템 개념도>

Survey time line



<그림 12. 행정(등록)기반과 조사와의 일정>

The integrated production system



〈그림 13. 통합생산시스템 개념도〉

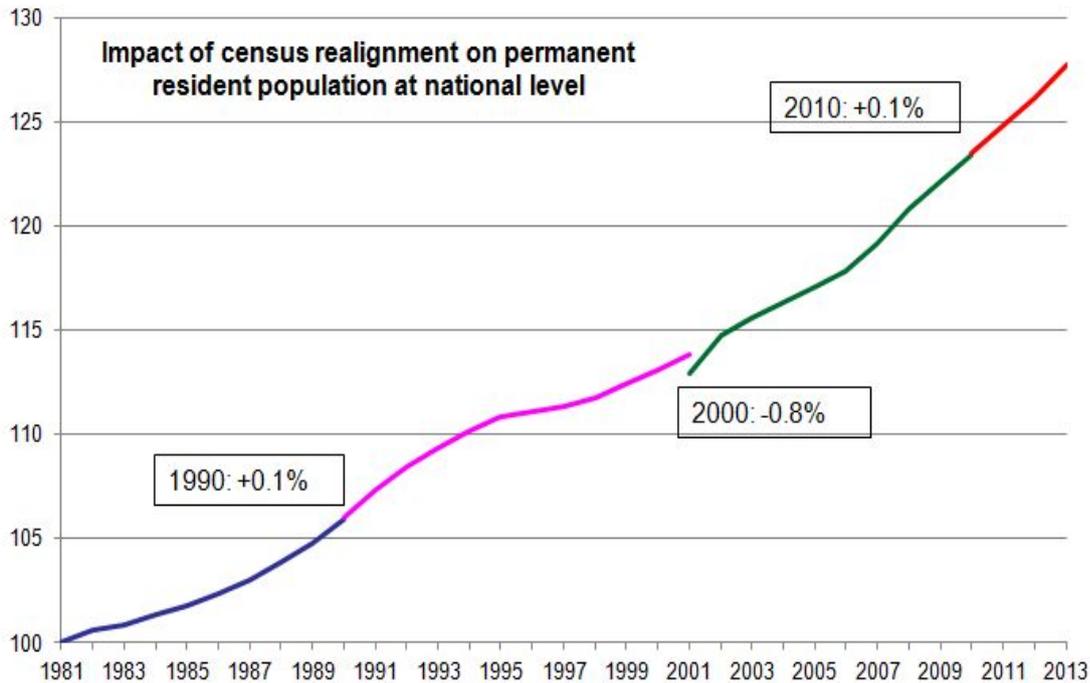
- (시계열) 조사에서 행정(등록)자료 기반으로 전환에 따른 시계열 단절이 발생을 고민했으나 실제로 그 차이가 크지 않아 자료를 조정하였고 다만, 연구자들이 요구할 경우 단절된 자료를 제공

Time series (I)

	stock data	stock and flow data	annual population statistics
1989		population estimates	realignment demographic accounting equation demographic accounting equation demographic accounting equation realignment demographic accounting equation demographic accounting equation demographic accounting equation realignment
1990	census	population estimates	
1991		population estimates	
1992		population estimates	
...		population estimates	
2000	census	population estimates	
2001		population estimates	
2002		population estimates	
...		population estimates	
2010	register survey	population estimates	
	stock and flow data		NEW POPULATION DEFINITION
2011	register survey (population registers) + civil registers		
2012	register survey (population registers) + civil registers		
2013	register survey (population registers) + civil registers		
....			

〈그림 14. 조사에서 행정(등록)기반으로 연도별 변화〉

Time series (IV)



〈그림 15. 조사에서 행정(등록)기반으로 시계열 단절〉

2. 행정자료의 통합과 경제 구조통계 소개

- (경제센서스) 스위스 통계청은 조사기반의 경제센서스를 2008년에 마지막 실시 이후 사업자등록(BR), 부가가치세(VAT), 사회보험(OASD), Profiling of enterprise('08년 ~), 특별조사(신생기업, 경제활동코드 품질, 고용통계 등)를 활용 2013년 행정자료를 활용한 경제센서스 실시
- (주의사항) 행정자료는 통계생산 목적이 아니므로 통계기준에 맞게 생성.변환을 해야 하고, 전체자료(전국 임시근로자수)는 있으나 세부적인 자료(사업체별 임시근로자수)는 없거나 통계단위가 변화(사업체 → 기업체)되는 경우도 있어 세심한 접근 및 검토가 필요

3. 사회보험(OASI)과 사업자등록(BR) 연계 및 이슈사항

- (연계결과) 사회보험은 성명과 주소를 기초로 개별단위별로 사업자등록과 결합하고, 사회보험 개별단위 전체자료는 2,036,935개('05년~), 이 중 '14년 활동단위 자료는 1,226,993개
 - '14년 사업자등록 활동단위 자료는 810,772개로 반드시 1:1관계가 아님

How to match OASI data with the Business Register BR/1

OASI register units

2	123456	Peter Buchs	Bern	347583	
2	123457	Claude Macchi	Biel	347583	→ BR
2	123458	Buchs & Macchi	Neuchâtel	347583	

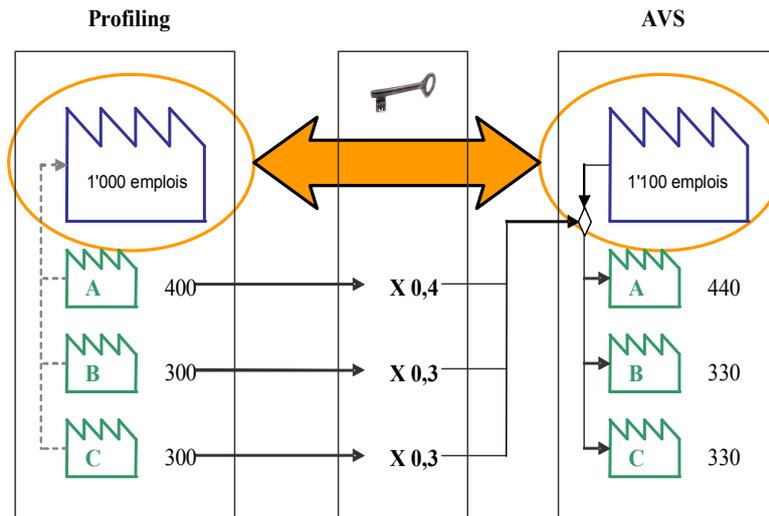
2	123456 → 123458
2	123457 → 123458

OASI register links

2	123456	756456789	1	12	2013	50'000	OASI salary data
2	123457	756999876	1	12	2013	50'000	
2	123458	756112354	5	12	2013	35'000	

〈그림 16. 사회보험과 사업자등록 연계방법 예시〉

- (이슈사항) 사회보험(OASI)과 사업자등록(BR) 연계시 추가적으로 검토하거나 조사자료로 보완해야 하는 경우가 발생
 - ① 어떤 고용주들은 사회보험에는 있으나 사업자등록에는 없는 경우가 있어 별도의 연계 규칙이 필요
 - ② 사회보험은 사업자등록(기업단위)으로 연계되나 지점(지역별)단위 자료가 없어 조사자료를 참고하여 지점(지역별)단위 자료를 생성

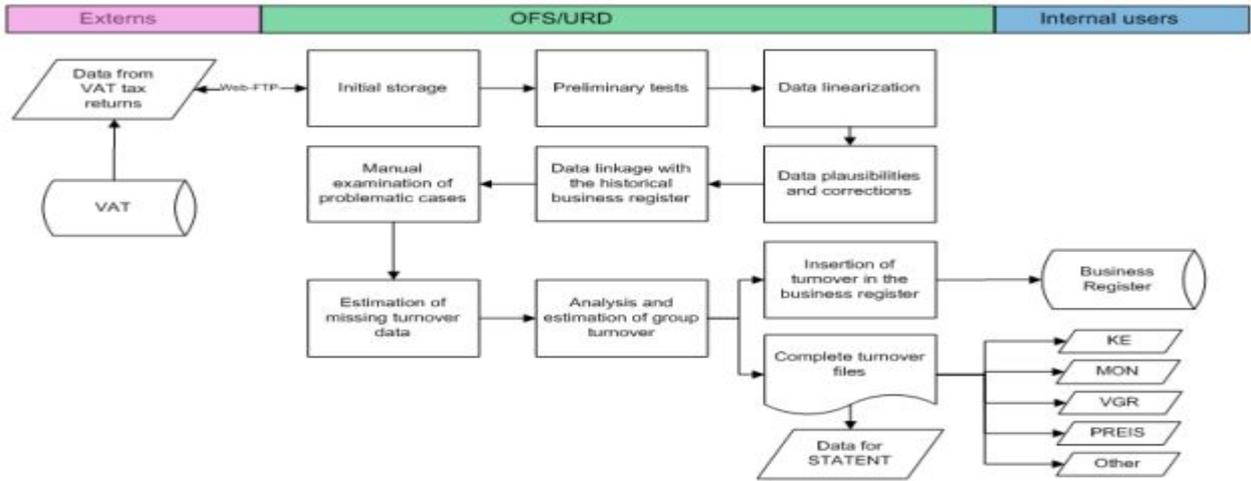


〈그림 17. 조사자료를 참고하여 지점(지역별)단위 자료를 생성 예시〉

- ③ 초대형 기업들과 중앙·지방정부들은 근로자들을 기업 또는 사무소 단위가 아닌 하나의 사회보험단위로 신고하는데 스위스 통계청은 정보가 없어 사회보험의 급여자료와 통계자료를 비교하여 알아내야 하고, 어떤 경우 스위스 통계청에 통보 없이 변경하므로 이런 초대형 기업들의 기업단위 세부자료는 조사에 기반한 통계적 절차화가 필요
- ④ 기타 2명이상의 자영업체, 공공분야 등에 대한 검토 필요

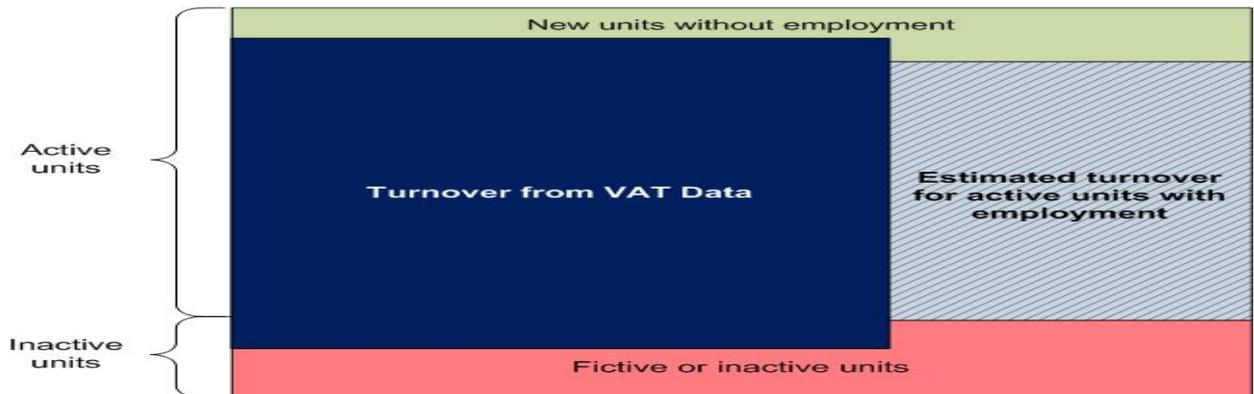
4. 부가가치세 자료의 연계 및 이슈사항

- (장점) 스위스 국세청은 중요자료인 매출액자료를 제공하는데 부가가치세의 장점은 하나의 원천자료 사용, 등록 단위들의 품질 제고, 기업의 응답부담 감소, 경제 분야 전체 구조정보 제공, 통계적 목적을 위한 총화·가중치 변수의 사용과 특히, 목표모집단의 범위를 확대
- (이슈사항) 부가가치세 자료를 활용하는데 있어 통계적으로 활용하기 곤란한 개념, 포괄범위의 문제, 자료를 활용할 수 있는 시점과 통계 공표시점 사이의 지체 등의 검토 필요



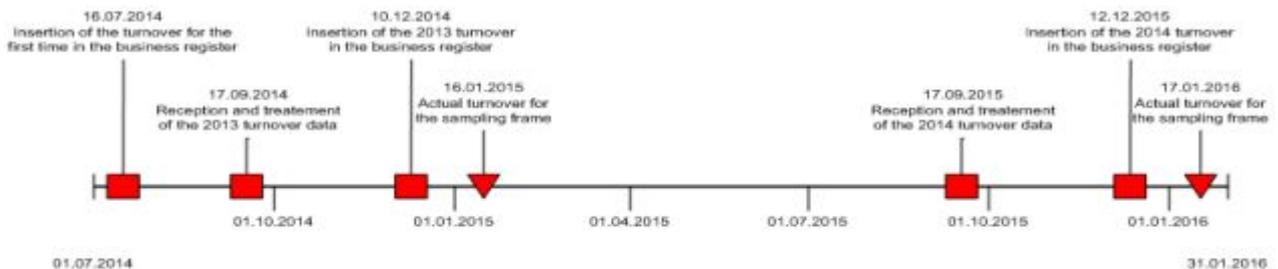
〈그림 18. 부가가치세의 통계생산 활용과정 체계도〉

- ① 부가가치세 그룹들은 오직 본사(가상)만이 국세청에 등록되어 있어 그룹전체에 대한 매출액을 활동단위(사업체 등)로 분배하기 위한 종사자규모.산업별 가중치에 기반을 둔 수학적 모델 사용
- ② 부가가치세 자료의 모집단은 경제통계 모집단과 같지 않음



〈그림 19. 모집단에서 부가가치세와 통계모집단 비교도〉

- ③ 조사기준 시점('13년말기준), 부가세 매출신고('14년 중), 부가세 매출자료 시스템 적재('14.12.10)사이에 시차가 발생 등



〈그림 20. 부가가치세 자료 입수 및 활용 가능 시점〉

□ 시사점 및 향후 추진방향(제안)

- (행정자료 기반의 통계생산) 내.외부의 조사환경 변화를 고려할 때 행정(등록)자료를 기반으로 한 통계생산 방식은 선택이 아닌 필수이나 행정자료만으로 수집하기 곤란한 항목은 현장조사와의 조화 필요
 - (스위스) 사업자등록(BR).부가가치세.사회보험을 활용하여 기업체별 매출액, 종사자수를 파악하나 사업체별.산업별 매출액, 종사자수는 현장조사를 통해 보완
- (Step by step) 행정(등록)자료를 활용한 통계 생산을 위해서는 법적근거, 이해관계자의 협조, 모집단.통계항목의 재정의, 행정(등록)자료 기반 구축 부서와 통계 생산부서의 협업, 행정(등록)자료 통계 생산에 활용할 수 있는 기본환경* 충족이 필요
 - * 인구.가족.건물.거주지.기업체.사업체 등의 고유번호 생성, 행정(등록)자료와 조사자료를 포괄하여 연계.통합할 수 있는 종합시스템
- (행정등록부) 현재 입수중인 행정자료 중 인구.건물.사업자 등 다른 행정자료를 연계.통합할 수 있는 기본 행정자료를 행정등록부라 하고, 전체 행정(등록)자료가 DB구축 되도록 행정자료관리시스템 보완
- (모집단 재정의) 인구.가구.사업체 등 통계생산의 기본이 되는 모집단을 행정(등록)자료 기반으로 구축할 수 있도록 재정의가 필요하고, 모집단 재정의에 따른 항목값, 통계량 등의 변화 연구
- (통계등록부/종합모집단) 모집단 재정의에 따라 i) 행정등록부와 기존 조사모집단을 연계한 기본 통계등록부 구축, ii) 각각의 기본 통계등록부를 연계.통합한 종합모집단을 구축, iii) 종합모집단을 기반으로 여러 행정자료, 조사자료를 연계한 통계생산시스템 구축