

# The Use of Administrative Data for Official Statistics in Korea

2012.11

STATISTICS KOREA 

# Contents



**I** Background

**II** Major Progress and Limitations

**III** Future Strategy for Use  
of Administrative Data

**IV** Expected Outcomes

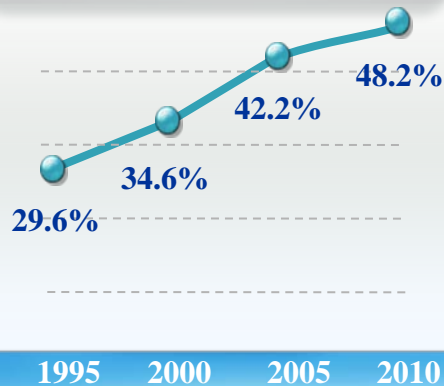
By utilizing administrative data, Kostat aims to overcome deteriorating condition of field survey to change high cost structure

➔ **“Improvement of Statistical System”**

## Deteriorating condition of Field Survey

- With growing number of 1 or 2-person households, and expanding awareness on privacy, increasing refusal to respond makes field survey more difficult

Rate of 1 or 2- person households



Non response rate in Household Income and Expenditure Survey



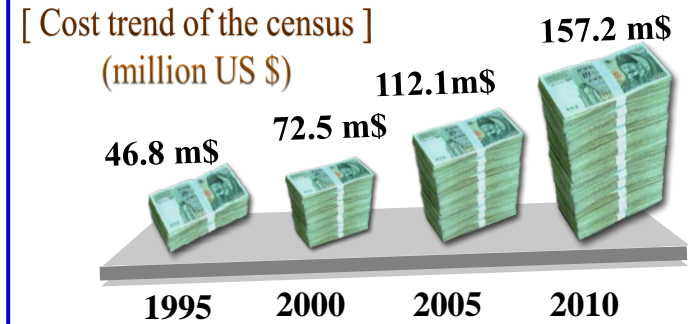
- Response burdens of businesses are increasing

Response burdens of businesses (Average annual)



## Survey costs are ever increasing

- Cost of taking surveys has rapidly escalated with growing labor cost



## Use of Administrative Data is a global trend

- OECD member countries are using tax data in compiling business establishment data
- European countries like Denmark and Finland introduced register-based Census

Strategic  
Direction

- To meet user needs and to improve national statistical system,  
**Kostat plans to expand the use of administrative data**
- **A comprehensive strategy needs to be laid out**

# Contents



**I** Background

**II** Major Progress and Limitations

**III** Future Strategy for Use  
of Administrative Data

**IV** Expected Outcomes

## Major progress

### 1 Establishment of Legal Basis

- Added a clause in the Statistics Act for the use of administrative data
- Amended Basic Law on National Taxation to allow to provide with taxation data
  - **KOSTAT may request taxation data from the National Tax Service for the purpose of producing official statistics.**

### 2 Establishment of Administrative Data DB

- So far 133 types of administrative data have been obtained, and databases of 18 administrative data have been established
  - **Used in producing 8 official statistics through substitution of survey variables, data editing or use as register information.**
  - \* **18 administrative data DB : 5 DB in economic area (e.g. business income reports), 9 DB in agricultural, forestry and fisheries area (e.g. crop production survey data), and 4 other DB (e.g. sample data from 2010 Population and Housing Census)**

### Major progress

#### 3 Expansion of Organizational Resources

- 3 divisions which are in charge of the use of administrative data have newly been established.

→ to enlarge the use of administrative data in statistical production and to acquire taxation data from NTS

\* 3 divisions : Administrative Data Division, Administrative Data Management Division, Register-Based Census Division

#### 4 Awareness of Private Information Protection

- Intranet system was designed to protect private information

→ To prevent administrative data from external leakages, the system is separated from the Internet

- Within the Intranet, statistical data and administrative data are closely linked and utilized

### Major progress

#### Still in Early Stage

- (Early Stage of Using Administrative Data) Progress has been concentrated on data collection and database development
  - 2 surveys use administrative data for substituting variables, 24 surveys for verification and supplementation of field survey
- Use for developing new official statistics or for substituting field surveys, which reduces cost greatly and has major visible results, has been limited

#### Strategic Direction

- **Paradigm shift on using administrative data is needed**

(Development phase) data collection / database development



(Utilization phase) new statistics development, substitution of surveys

- Create interactive relationship between statistical surveys and administrative data to complement each other

expand use of administrative data



(when necessary) strengthen field surveys



# Contents



**I** Background

**II** Major Progress and Limitations

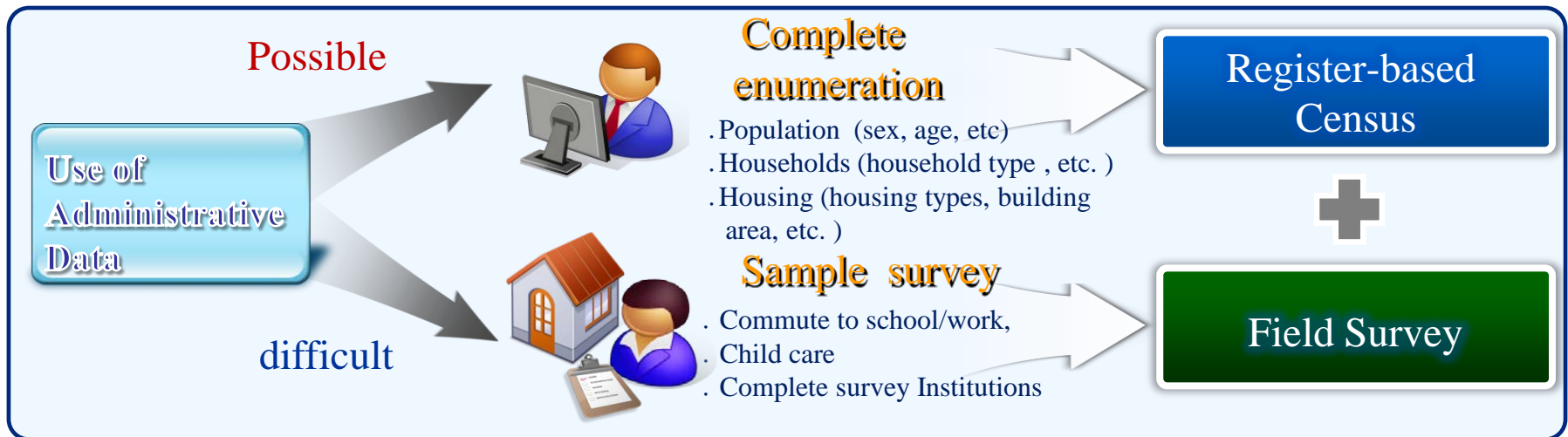
**III** Future Strategy for Use  
of Administrative Data

**IV** Expected Outcomes

## 1 Fulfillment of user needs and reduction of response burden

### 1-1 Introduction of Register-Based Census

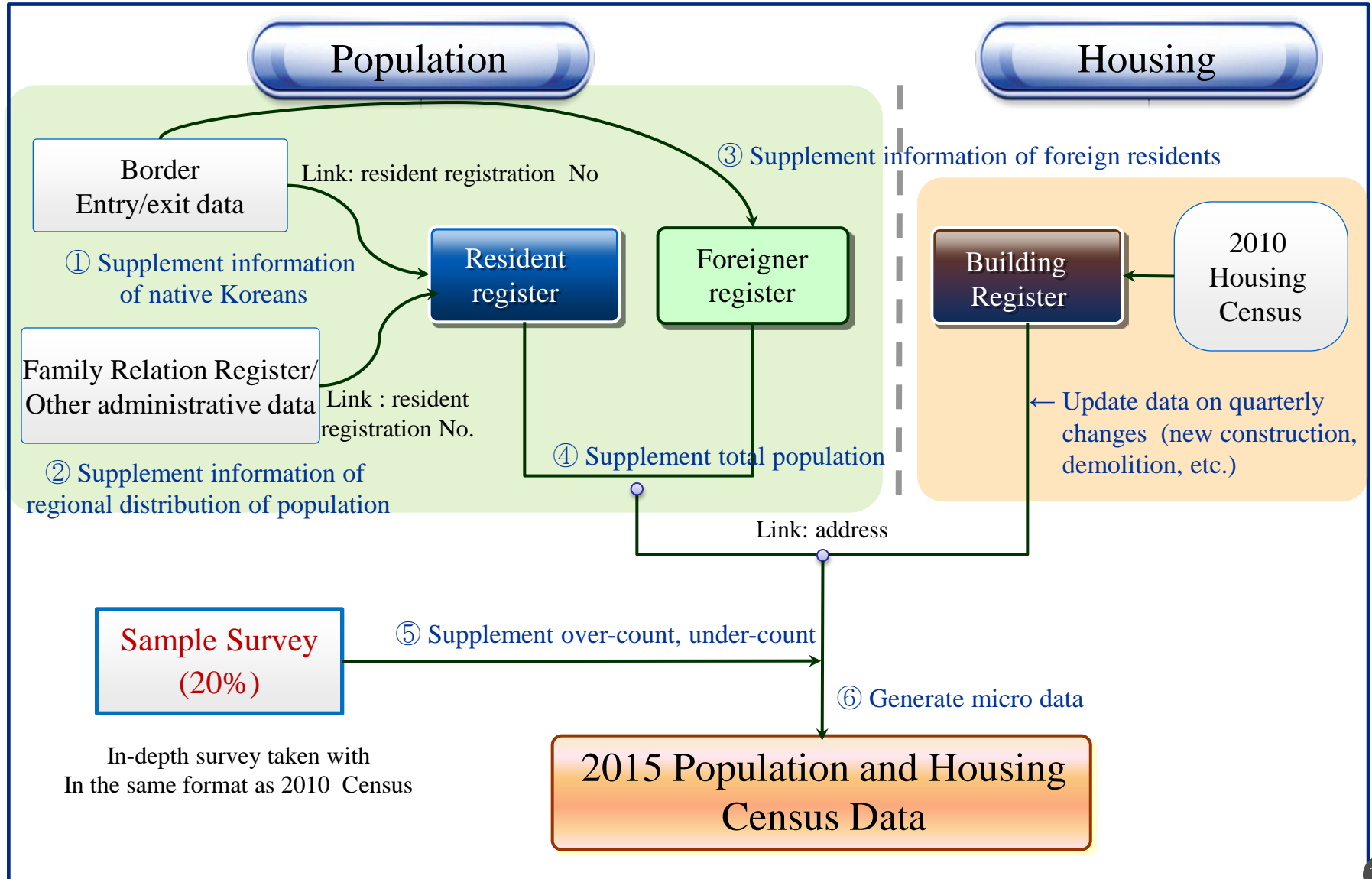
① Response burden on citizens ② Cost saving ③ Feasibility taken into consideration



For 2015 Census,

- **Basic fields (fields in complete enumeration) covering all citizens will be substituted using administrative data**
  - \* # of substituted fields : about 15 fields out of 19 total fields in 2010 census can be substituted
- **In-depth fields (fields in sample survey) covering sample population will be completed based on field survey**
  - \* Sample size : 2010(10%) → 2015 Register-based Census (about 20%)

## 2015 Census Model: Combination of traditional method and use of administrative data





4 major administrative data obtained ; other administrative data planned to be obtained

● **Data first obtained on 2010. 4.**

- Identification codes : resident registration no. & address
- Contents : name, resident registration no. (sex, date of birth), relationship to head of household, data of move, etc.

Resident Register

Obtained

● **Data first obtained on 2011. 3.**

- Identification code : resident registration no.
- Contents: name, family clan, place of birth, date of birth, marital status, date of marriage, etc.

Obtained

Family Relation register

Foreigner Register

Obtained

● **Data first obtained on 2010. 4.**

- Identification codes: foreigner registration no. & address
- Contents: name, foreigner registration no., date of birth, nationality, sojourn status, date of entry, etc.

Obtained

Building Register

● **Data first obtained on 2009. 4.**

- Identification code : address
- Contents : address, size of area, building purpose, structure, number of floors, approval issue date, name of owner, etc.



Fields that can be completed using administrative data have been selected.

Field	2010 Complete enumeration fields (traditional method)	2015 Register Based Census			
		Resident Register	Family Relation Register	Building Register	Foreigner Register
Population / Housing	Name	●	●		●
	Gender	●	●		●
	Age	●	●		●
	Relationship to Head of Household	●	●		
	Educational Attainment	●			
	Residence 1 year ago		●		
	Residence 5 years ago		●		
	Household Type	●	●		
	Period of Residence		●		
	Marital Status	●		●	
	Place of Birth			●	
	Family Clan			●	
	Nationality	●		●	●
	Date of Entry	●			●
	Number of Rooms in House	●			
	Type of Housing Facilities	●			
	Type of Housing Occupation	●			
	Floors of Building/Residence	●			
	Owner Household and Other Housing Ownership Status	●			●

Field	2010 Complete enumeration fields (traditional method)	2015 Register Based Census			
		Resident Register	Family Relation Register	Building Register	Foreigner Register
Housing	Type of residence	●		●	
	Total size of Living Area	●		●	
	Year Built	●		●	
	Total number of rooms**(tentative)	●		▲	
	Number of housing facilities**(tentative)	●		▲	
Total	19	8	4	4	5

### Number of fields

- Complete enumeration in 2010 Census taken by traditional method : 19 fields
- Complete enumeration in 2015 Register-based Census : about 15 fields

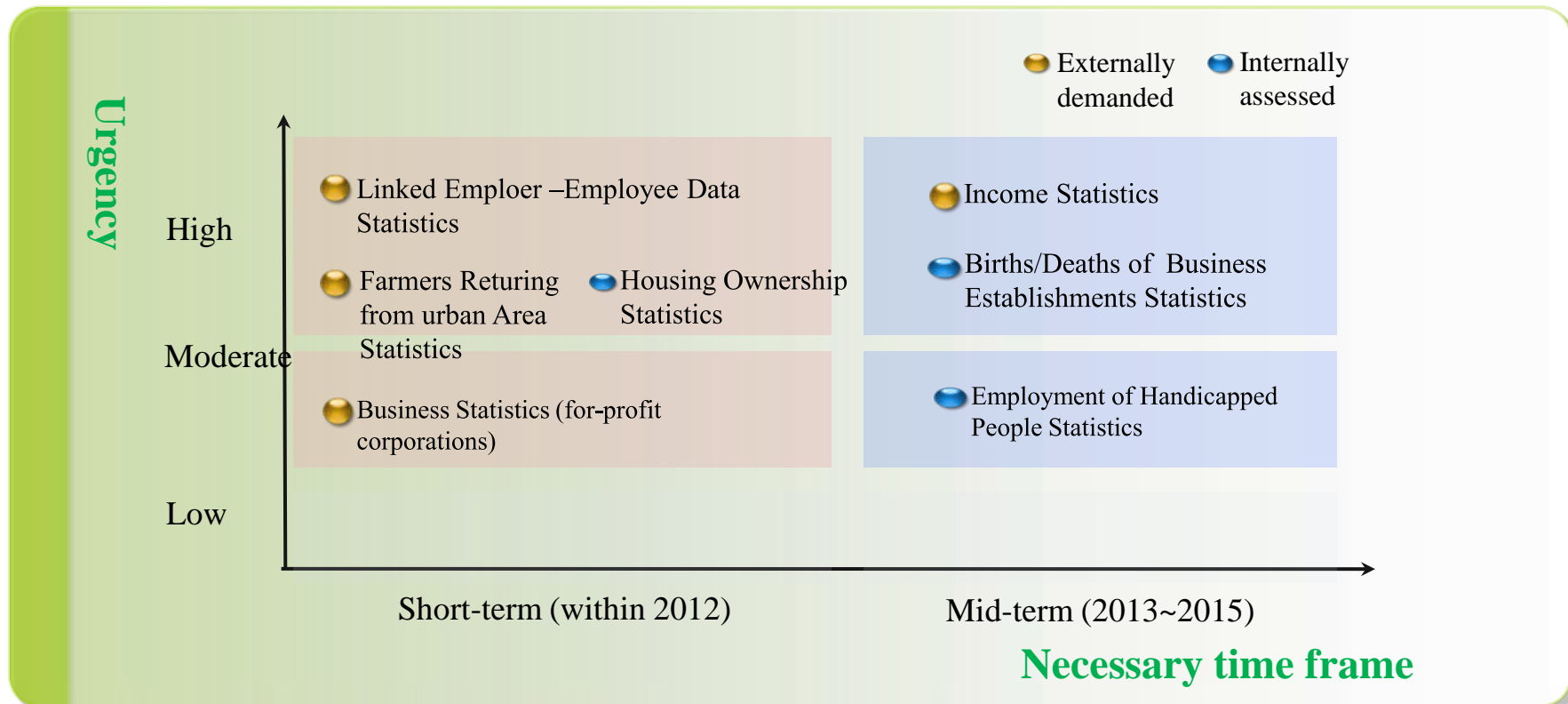
1-2

## Development of New official Statistics using Administrative Data

New Statistics are selected based on urgency and necessary time frame for development.

Criteria

- ① **urgency : existence of substitutable statistics is considered**
- ② **necessary time frame : type of necessary data(DB) is considered**



## ① Linked Employer-Employee Data Statistics

- Provide wage employee's working position statistics by business size and industry classification
- Provide information on total number of working position, job condition (stability, wage, etc.)

### A administrative data used

- 7 types of data are used
  - individuals/households : employment insurance, national pension, health insurance
  - establishments : occupational health & safety insurance

### Statistical Output

- (Fields Produced) 10 fields including gender, age, years of employment, educational attainment, type of business, industry classification, number of employees, years in business, revenue, capital stock
- (Major Indicators) Rate of changes in job creation, job disappearance, growth of jobs, Relocated jobs, continued jobs and change rates in the number of working positions



#### ② Farmers Returning from Urban Area Statistics

- Estimate the number of returning households and people, business type of returning farmers by city/province and district
- Provide basic data to support comprehensive policies on returning farmers and returns to rural areas

#### A administrative data used

- 4 types: Ministry of Food and Agriculture's data on returning farmers and returns to rural areas, register data on agricultural businesses, resident register data, family relation register

#### Statistical Output

- (Fields produced) 15 fields including basic items (i.e. name and gender of business owner) and relocation by region, type of business operation, farming products, type and number of livestock
- (Major indicators) regional statistics of returning farmer households and number of Household Members, type of farming, products, type of livestock, motives for starting farming, years in farming, participation in farming business organization, etc.

## ③ Business Statistics For-profit corporation

- Provide data on number of for-profit corporations (business corporations), Number of employed persons, revenue, business operation analysis, etc

### A administrative data used

- 5 types: business establishment register data, value added tax data, corporation register data, corporate tax data, etc.

\* Census on establishments and Business Activities Survey data can be used as reference data

### Statistical Output

- (Fields Produced) 8 fields including number of business establishments, number of employed persons, revenue, capital stock, and affiliated businesses
- (Major indicators) Business status by major industry classification, type of for-profit corporation, region and size, revenue, number of employees, business results, etc.

#### ④ Housing Ownership Statistics

- Provide data on housing ownership by individuals and by households by linking administrative data (building register, property ownership register, resident register) and Population and Housing Census data

##### A administrative data used

- 5 types: Building register, publicly announced housing prices (Ministry of Land, Transport and Maritime Affairs), property ownership register, resident register (Ministry of Public Administration and Safety)

##### Statistical Output

- (Fields Produced) 6 fields including housing owner, household members, number of household members, publicly announced housing prices, year of construction, number of multi and individual housing
- (Major indicators) housing ownership status (by individual, household, age and domiciles), publicly announced housing prices, area/year, household status, etc.

## ⑤ Income Statistics

- Provide income data of head of household, by region, industry/occupation, gender, age
- Provide statistical information to support policies on improving income distribution structure

### A administrative data used

- 3 types: Gross Income Tax (National Tax Service), Family Relations Register (Supreme Court), Resident Register (Ministry of Public Administration and Safety)

### Statistical Output

- (Fields Produced) 10 fields including socio-demographic characteristics (gender, age, etc.), economic activities status (position), housing, income type (interest, dividend), etc.
- (Major indicators) Income, Gini Coefficient, Quintile Income Distribution Ratio, Relative Poverty Ratio, etc.

#### ⑥ Employment of Handicapped people Statistics

- Compile job seeking handicapped people's list data and produce statistics on employment of handicapped people
- Alleviate difficulties faced by the handicapped in seeking jobs and utilize data in employment policies

#### A administrative data used

- 11 types: handicapped people register, employed handicapped people list, handicapped people's pension, family relations register, income reports, welfare recipients

#### Statistical Output

- (Fields Produced) 23 fields including disability information (type of disability/grade), establishments employing the handicapped (name of business, owner), employed handicapped people (position/wage), size of handicapped people having difficulty seeking jobs
- (Major Indicators) employment of the handicapped, employment rate of the handicapped by industry, handicapped people having difficulty seeking jobs, quality of employment of the handicapped, non-wage handicapped workers, etc.

#### ⑦ Births and Deaths of Business Establishments Statistics

- Identify births/deaths, life, survival rate of business establishment
- Assess changes in employment due to births/deaths of business establishments

#### Administrative data used

- 3 types : National Tax Service's business establishment register data, Business Information data managed by Korea Chamber of Commerce and NICE Credit Rating Information

#### Statistical Output

- (Fields Produced) Number of (new, birth/death, survival) establishments by year, survival rate of continuing business, birth/death rate and survival rate of establishments by industry and by region
- (Major Indicators) births/deaths of establishments by industry, births/deaths of establishments by region, survival rate by industry and by region, number of employees by industry and by region

## 2

## Strengthening infrastructure for the use of administrative data

### 2-1 Establish basic laws on the use of administrative data for statistical purposes

#### Background

- Administrative data can be obtained under Article 24 of Statistics Act, but an exclusive law should be established to obtain administrative data in a more effective manner
- Necessary for the successful implementation of register-based census in 2015

#### Major contents

- Mandating consideration of the use of administrative data when producing statistics
- Specify the statistics and the data to be used (specify the names of surveys and administrative data in the related law)
- Standardize administrative format and joint use of administrative data
- Matters pertaining to using administrative data in Population and Housing Census (register-based census)
- Protection of private and confidential information

\* Store access history of administrative data; fully guarantee confidentiality of private information

## 2-2 Systematic administrative data acquisition and establishment of database

### Progress

Total 133 sets of administrative data obtained, of which 18 databases established

### Administrative Data Needed

● Development of New Statistics or Field Substitution: 5 data sets

Statistics	Administrative Data
LEEP (develop)	Business Income Payment Itemized Report
Income Statistics (“)	Gross Income Tax
Business Statistics (“)	For-profit Corporation Data
Housing Ownership Statistics (“)	Individual Housing Publicly Announced Price Data
Livestock Statistics (substitute)	Cattle History Tracking System

● Data for Checking and Supplementation : 8 data sets

Surveys	Administrative Data
Estimated Future Population	Overseas Korean Registration
Household Finance	Basic Senior Pension Data
Construction	Real Estate PF Data
Construction Orders Received and Completed	National E-Procurement
Fishing Production Trend	Fishery Business Permit Register
Price Received and Paid by Farmers	Nonghyup Total Sales Price Data
Rice Price at Production Site	Nonghyup RPC Purchase/Sales Data

### Strategic Direction

- Obtain administrative data need to develop new statistics or for field substitution first
- For administrative data intended for checking and supplementing survey data, first review expected results and time frame related to using such data and then try to obtain it



2-3

## Development Plans for Database on Composite Population



### Development of database

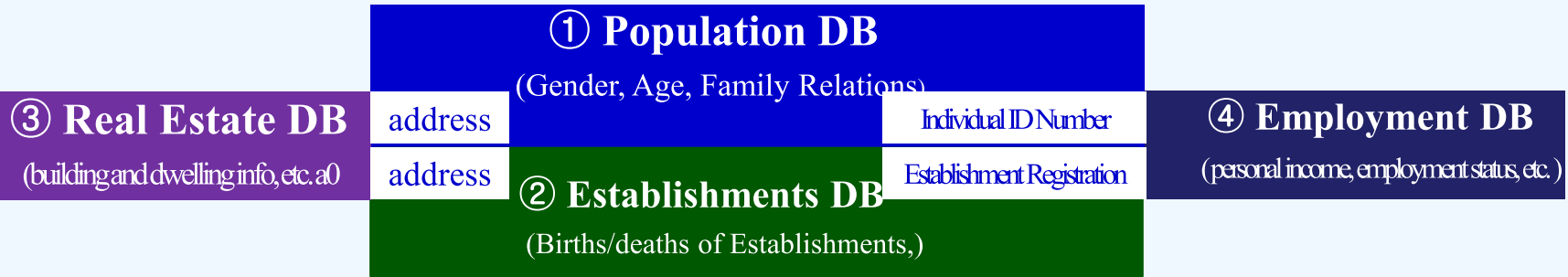
- Develop database of business establishment population and business enterprise population
- Develop database for Register-based Census by linking population and household DB with Housing DB

### Limitations

- Lack of systematic approach as database is established for each official statistic separately
- As separate databases are separately built for household and establishments, use of data is limited

### Future strategy for composite population

- Develop composite population database by linking ①Population - ②Establishments - ③Real Estate - ④ Employment database



\* Link Key : Address (link ①-②-③), Individual ID Number (link ①-④), Establishment Registration Number (link ②-④)

## 3

### Efficient management of administrative data

#### 3

#### Standardize the format of administrative data

#### Background

- Due to non-standardized format of administrative data, its use is limited.
  - \* 25,382 legal forms exist (as of 2011), about 4,000 legal forms revised each year
- Register data format needs to be standardized and unified code system is necessary.

#### Major Contents

- By participating in Ministry of Public Administration and Security’s **“Project for common use of Administrative Information.”**
  - Classification experts involved in standardization (terms, classification, etc. of various forms) initiative to allow common use of administrative data



- Standardize registration format in conjunction with revision of the forms for future use in statistics
  - (Example) National Tax Service business category classification (897) → Standardize as to be compatible with Korea Standard Industrial Classification (976)

4

## Enhance private information protection.



4

### Strengthening of data protection in production of statistics and data dissemination

Background

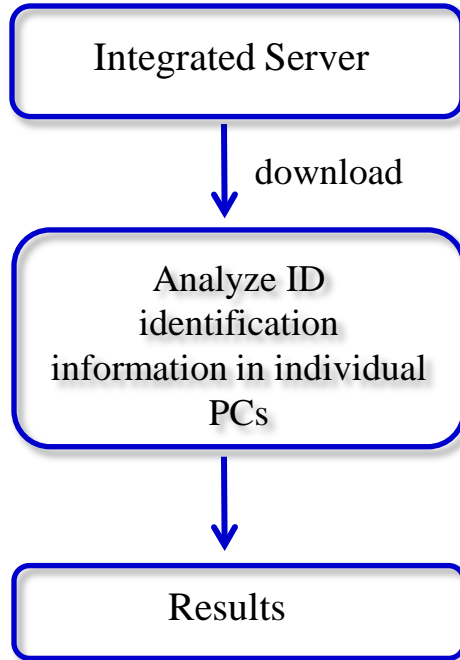
- Huge loss and distrust can be resulted if breach of administrative data occurs.
- Alleviate public's concern on **“Big Brother Nation”** by complete protection of private information



Status and  
Limitations

- **(Legal protection)** Criminal charge if privacy protection under the Statistics Act is violated.
  - \* Rule on Information Security of Administrative Data (2009) is established by KOSTAT
- **(Intranet)** Only Intranet, separated from Internet, is used.
- **(Limitations)** Security issues may arise if administrative data is stored and used in individual PCs.

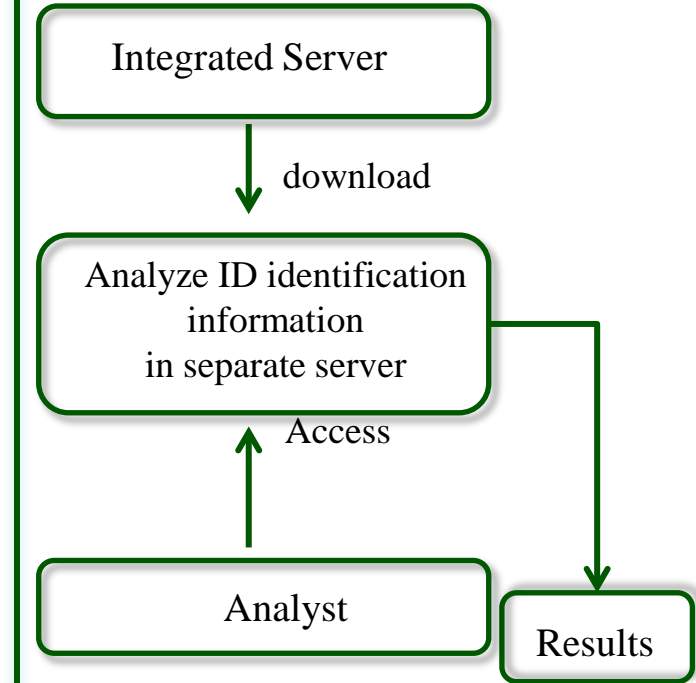
## Current : Individual PCs



\* Data with Personal identification is processed in individual PCs.



## Future System : Remote Analysis



\* Data with Personal identification is processed in the separated server. (screen capture and download are not possible)

# Contents



**I** Background

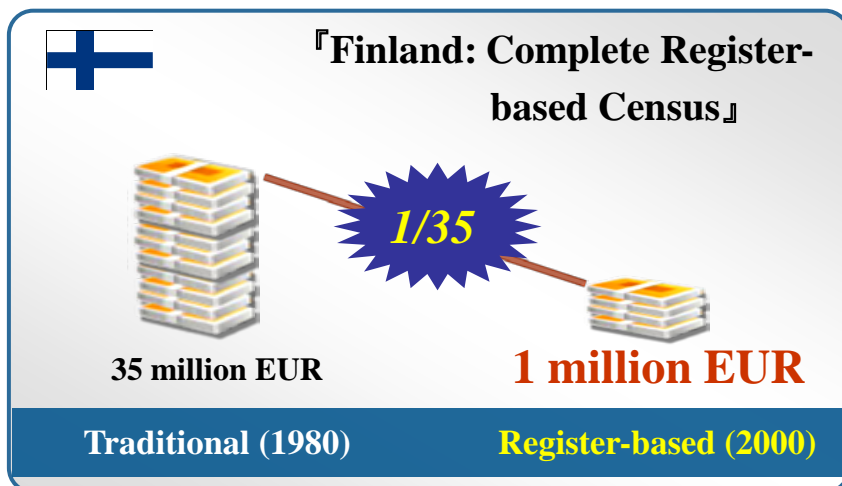
**II** Major Progress and Limitations

**III** Future Strategy for Use  
of Administrative Data

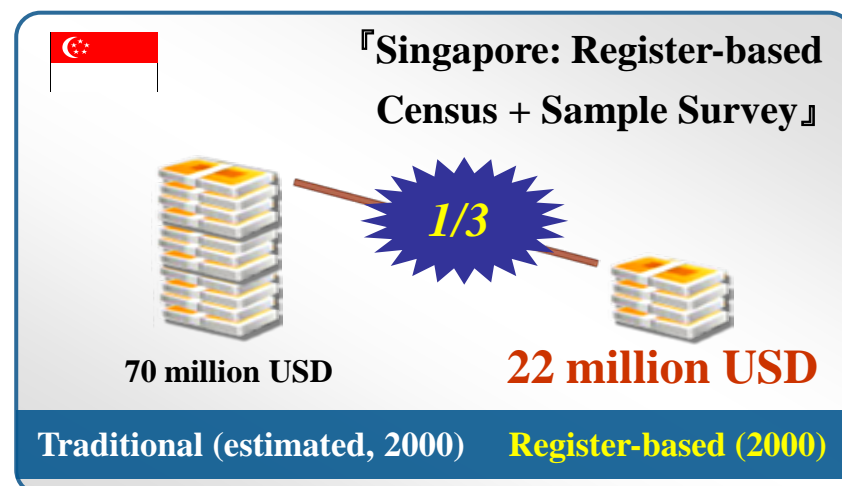
**IV** Expected Outcomes

- Fulfill various new statistical demands emerging from changes in social and economic conditions.
  - develop 7 new statistics (e.g. LEED statistics, housing ownership statistics, income statistics)
- Alleviate response burden of individuals and businesses
  - Substitute (about 15 to 19) complete enumeration fields in 2015 Population Housing Census using administrative data (no field survey)
  - (Census on Establishments) Of 13 fields, only 2 fields to be responded by businesses (substitute remaining 11 fields)
- Develop low cost, highly efficient statistical system
  - Approx. 120 million USD expected to be saved by using administrative data in 2015 Population and Housing Census (50% cost saving vs. traditional survey method)

\* Foreign cases of budget reduction



Source: Use of Registers and Administrative Data Sources for Statistical Purposes  
- Best Practices of Statistics Finland



Source: Combining Survey and Administrative Data for Singapore's Census of Population 2000  
- 53th ISI Seoul Conference (2001)

A background image of a bright blue sky with scattered white, fluffy clouds. The text 'Thank you' is centered in the middle of the image.

**Thank you**