KOSTAT's Experience in Enabling an Institutional Environment for Data Integration

UNESCAP Regional Workshop 2020
November 24, 2020
Administrative Data Management Division, Statistics Korea

01 Background

02 Acquisition and Use of Administrative Data

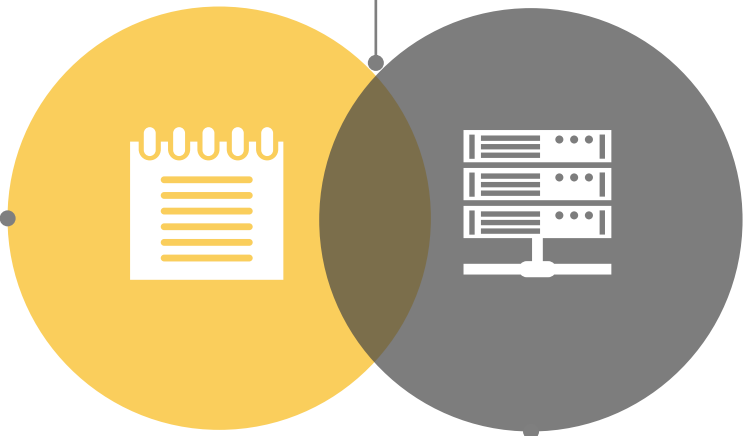
03 Protection of Privacy and Confidentiality

04 Collaboration with Data Provider

05 Future Plans and Issues

1. Background

New paradigm from surveys to administrative data



Integrated Data

Administrative data assist surveys and replace some part of the surveys (e.g., Statistical Business Registers to replace business surveys)

Survey Data

Traditional method for statistical data through F/F, CAPI, Internet, etc.

Administrative Data

Administrative data are directly used to produce statistics, replacing the existing programs (e.g., Register-based Census)

1. Background

New paradigm from surveys to administrative data

Survey Data

- Growing apathy / hostility to surveys
- Growing concern for response burden
- Budgetary constraints
- Growing need for protecting personal information

Administrative Data

- Already available and cheap
- Growing demands for statistical data
- Improving the statistical programs
- Checking the quality of statistical products

1. Background

Favorable environment for Integrated Data

Revision of 3 Laws Related to Data

- Personal Information Protection Act, Credit Information Use and Protection Act, Act on Information and Communications Network
- Possible to use pseudonymous data for compiling statistics, research, or public purpose without the approval of the owners

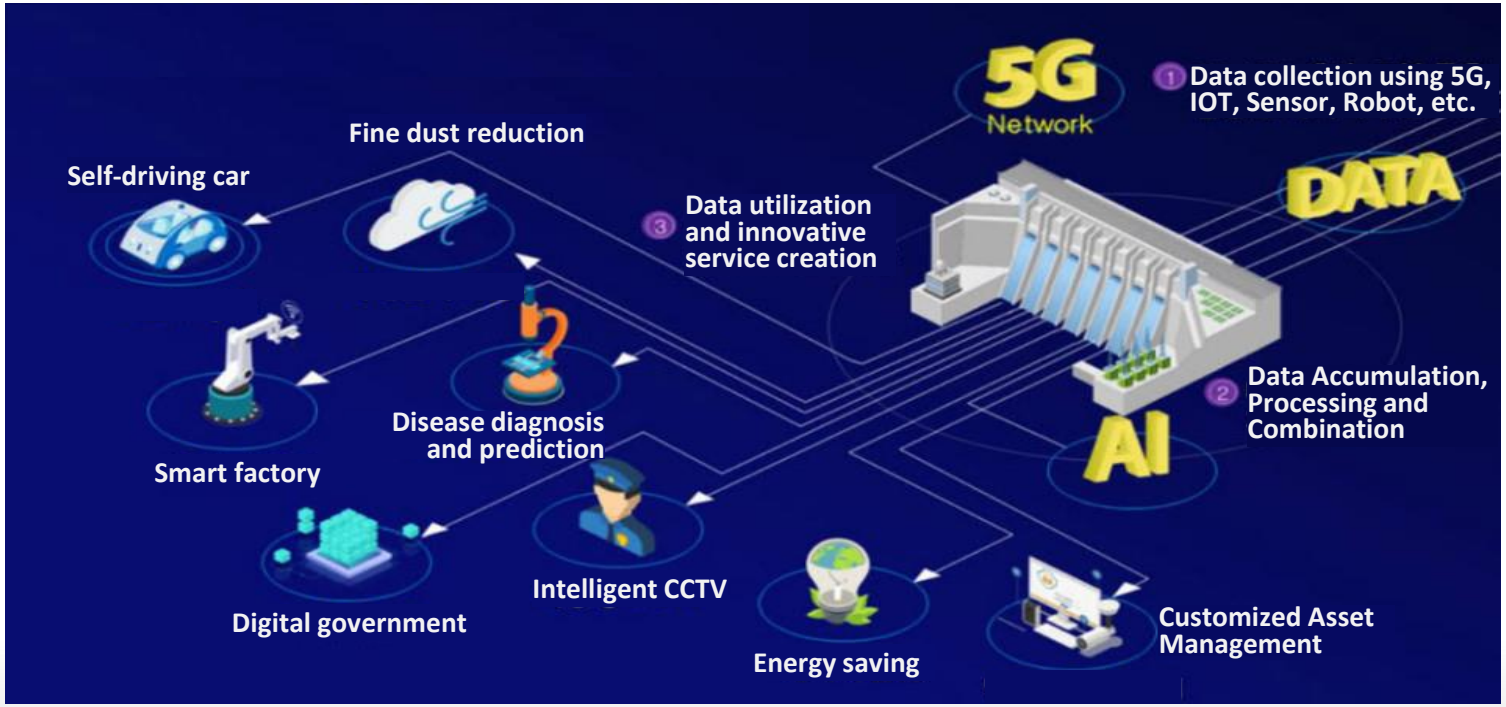
Personal Information Protection Act	Credit Information Use and Protection Act	Act on Information and Communications Network
Using Pseudonymous Data for Commercial Purpose	Using Pseudonymous Data for Big Data Analysis in Financial Sector	Transferred Privacy-related Matters to Personal Information Protection Act

1. Background

Favorable environment for Integrated Data

The Korean New Deal Policy (2020.7.)

- Building ‘Data Dam’ for Collecting, Sharing and Integrating Data
- Accumulating Data on ‘Data Dam’ and Using it with Artificial Intelligence or 5G Network



1. Background

Use of Administrative Data is a Global Trend

- (UN) FPOS specifies use of administrative data

> **UN Fundamental Principles of Official Statistics (5th Principle)**
>
> - Data for statistical purposes may be drawn from all types of sources including surveys and administrative data

- (EU) 'European Statistics Code of Practice' specifies right to access administrative data
- Most OECD member countries are using tax data in compiling business establishment data
- European countries like Denmark and Finland introduced register-based Census

2. Acquisition and Use of Administrative Data

The general precondition to use administrative data is **LEGAL BASIS**

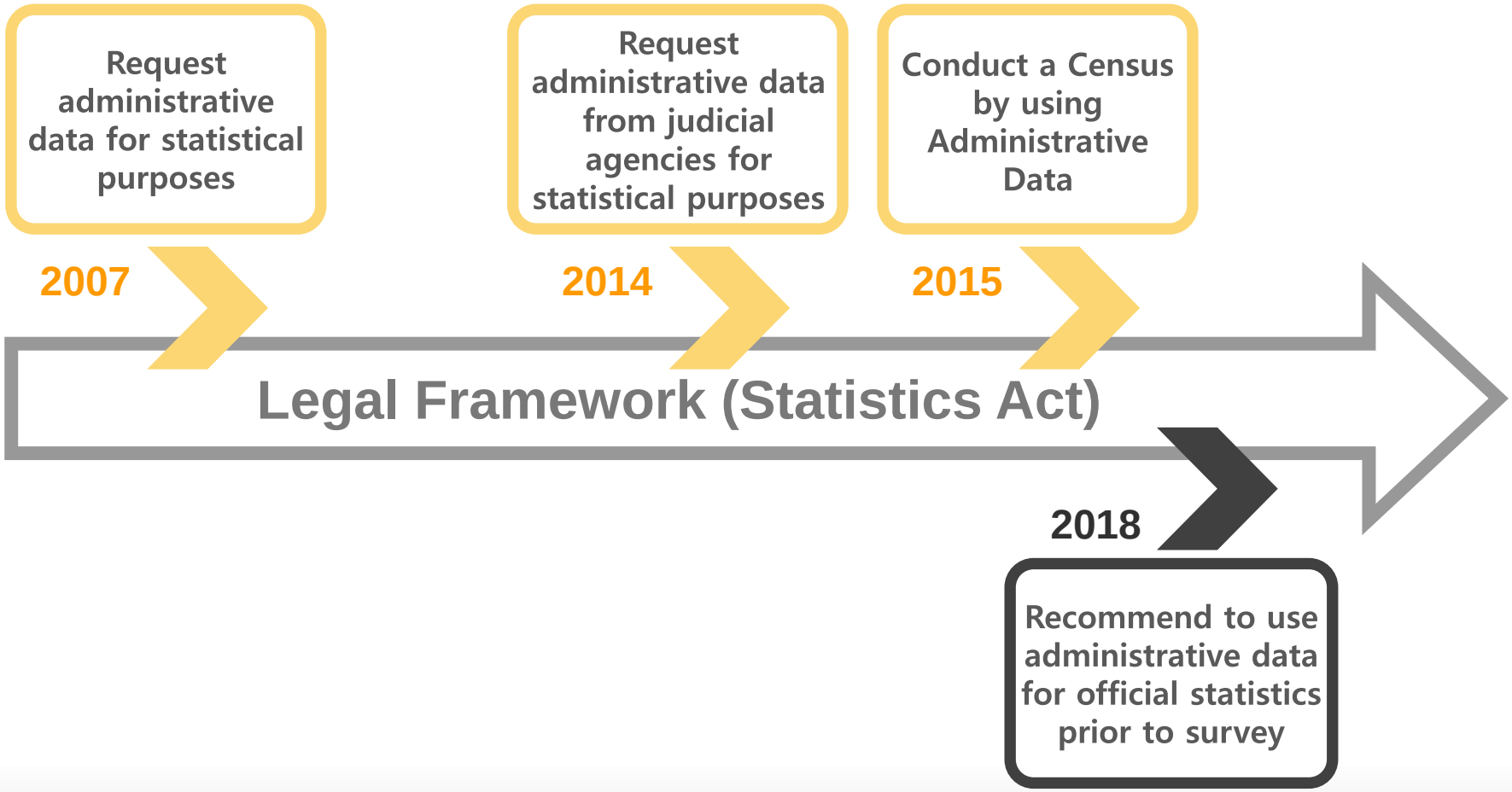
UNECE Guidelines on Statistical Business Registers)

The preferred approach is for the NSI's right of access to administrative data to be enshrined in a general Statistics Act.

In any case the NSI should consider initiating and introducing changes to its statistical legislation to ensure access is guaranteed by law.

2. Acquisition and Use of Administrative Data

The general precondition to use administrative data is **LEGAL BASIS**



2. Acquisition and Use of Administrative Data

Establishing legal framework for using administrative data by amending Statistics Act

Preparing the Legal Basis for Acquiring and Using Administrative Data

- Statistics Act Article 24 (2007)

→ Request for use of administrative data can be made for statistical purpose.

→ Head of public agency receiving such requests must supply administrative data unless there is a reasonable excuse.

→ The terms of provision of administrative data shall be decided based on consultation between two agencies.

→ The administrative data supplied by public agencies shall not be used for any purpose other than for the collection of statistics nor shall it be supplied to other persons.

→ In cases of violation of the provisions, supply of data may be suspended or limited.

- Statistics Act Article 24-2 (2014)

→ (Commissioner of Statistics Korea) is entitled to request access to Electronic Family Relations Register Data and Criminal Data for statistical use.

- Article 2-2 of the Enforcement Decree of Statistics Act (2015)

→ Can take a Census by using administrative data based on the Statistics Act Article 24

2. Acquisition and Use of Administrative Data

Establishing legal framework for using administrative data by amending Statistics Act

Implementation of the law which allows the use of administrative data prior to approval

- Statistics Act Article 18 (2018)

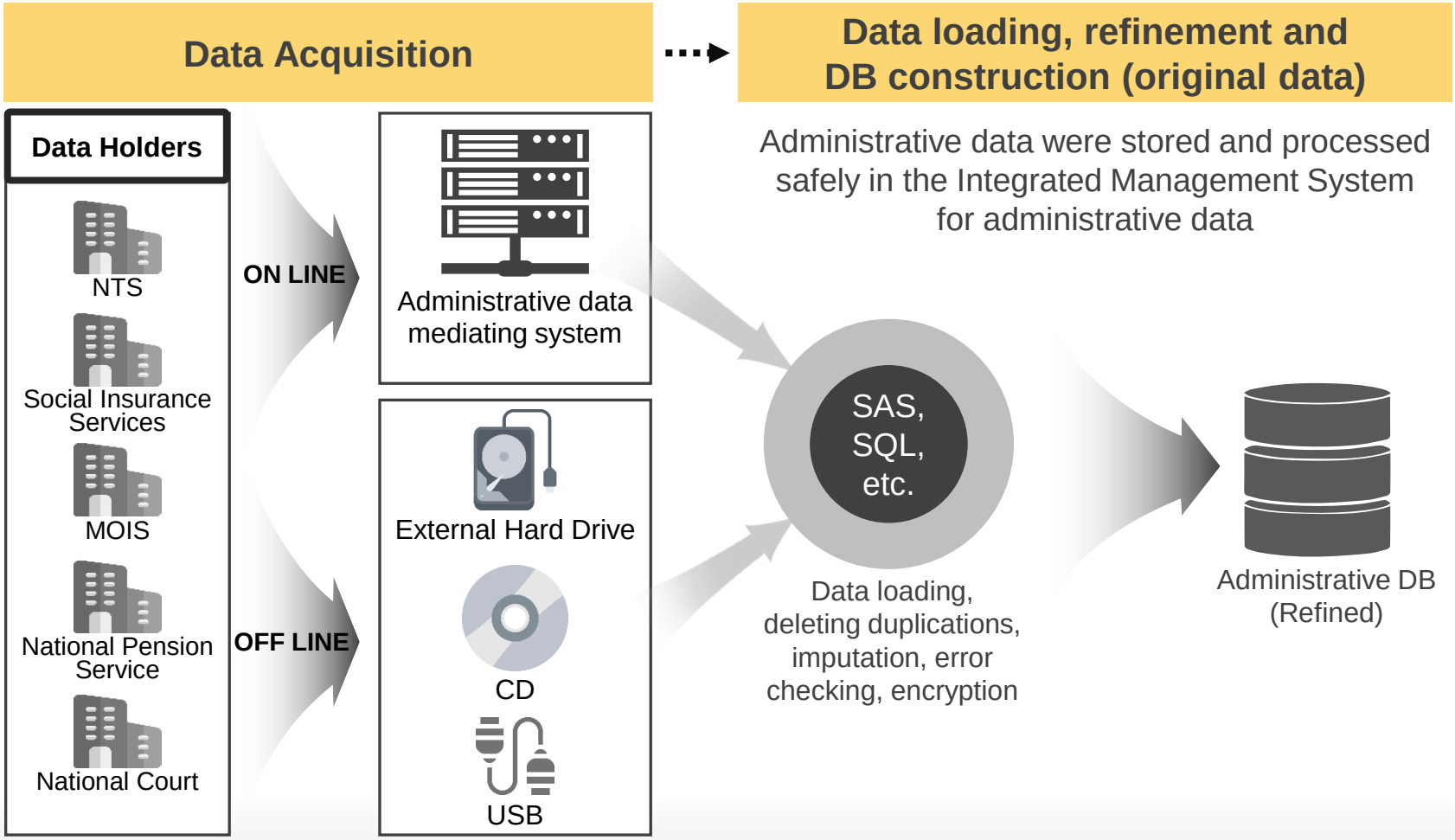
  → The head of a statistics service agency shall judge in advance whether it is possible to produce statistics by utilizing administrative data before he/she obtains approval for Production of Statistics.

- Statistics Act Article 25 (2018)

  → Where relevant data are necessary for the production of designated statistics, the head of a central administrative agency or local government shall judge in advance whether he/she may achieve the purpose of producing designated statistics by utilizing administrative data provided pursuant to Article 24.

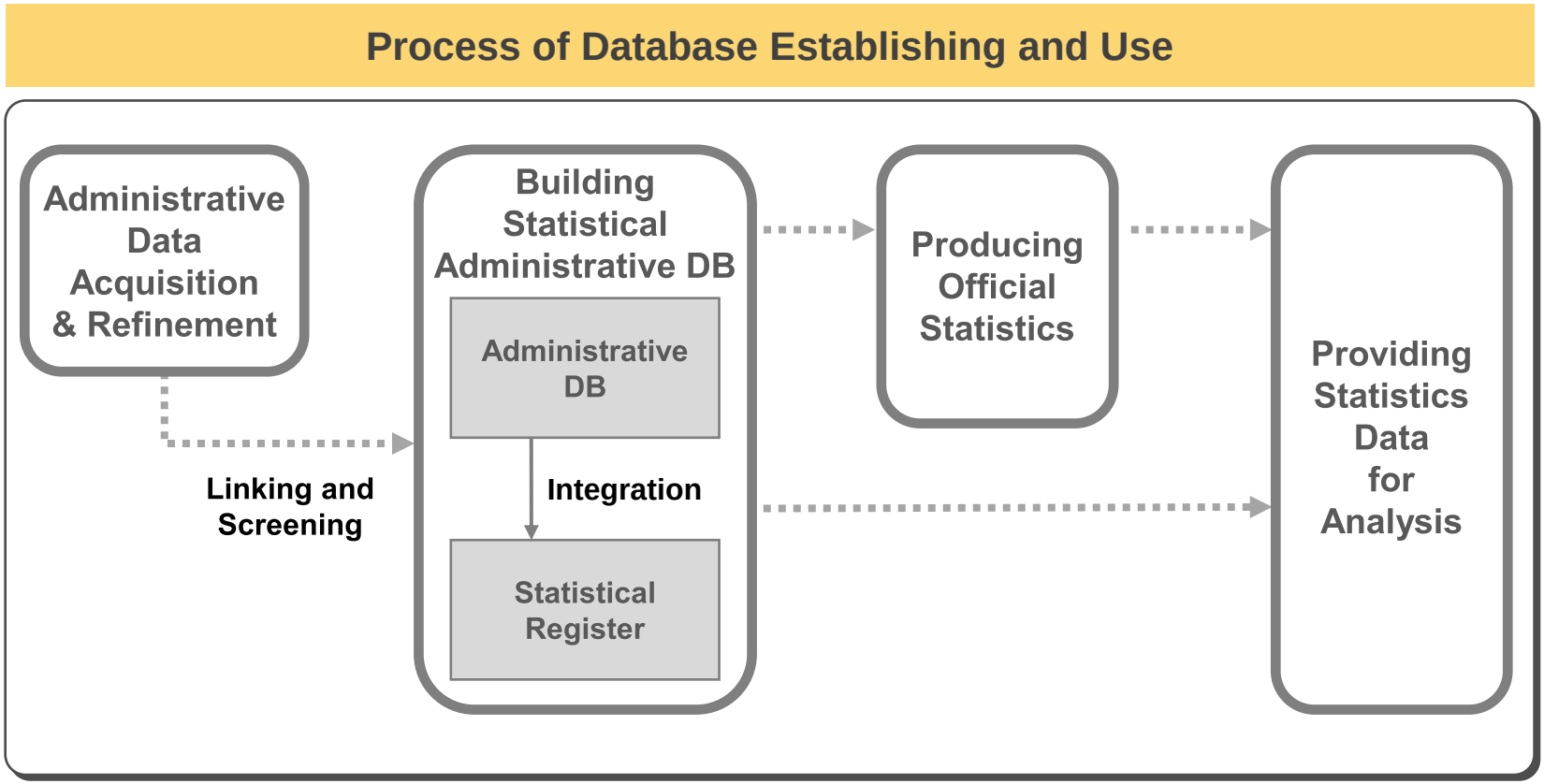
2. Acquisition and Use of Administrative Data

Building Administrative DB after Refining & Securing Process



2. Acquisition and Use of Administrative Data

Producing various statistics by using administrative data



2. Acquisition and Use of Administrative Data

Producing various statistics by using administrative data

Acquisition of Administrative Data

- 263 administrative data from 98 public organizations (e.g. Ministry of Government Administration, Ministry of Land, Infrastructure and Transport, National Tax Service, etc.)
  - Resident register, Building register, Social insurance, Tax data, etc.
- 131 Administrative datasets in 2011 → 263 Administrative datasets in Oct. 2020

Building Administrative DB

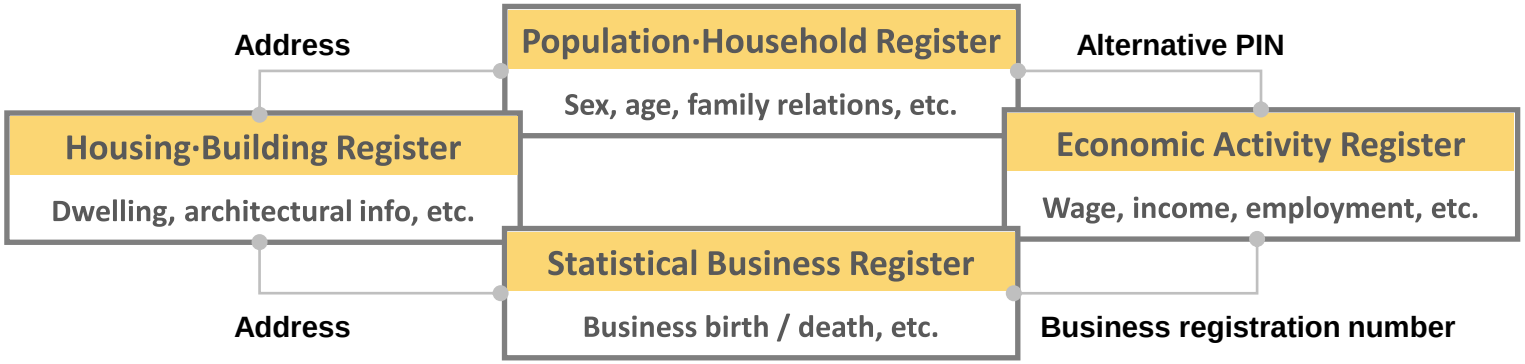
- Refining administrative data for building individual Administrative DB
- 69 DB in 2011 → 125 DB in 2019

2. Acquisition and Use of Administrative Data

Producing various statistics by using administrative data

Building Statistical Register

- Linking different administrative data for building 4 mail Sectoral Statistical Registers (since 2014)



- Producing new socio-economic statistics by using Sectoral Statistical Registers
  → Reinforcing collaboration with organizations for improving statistics
  → Providing Statistical Registers and customized DB at Statistics Data Center

2. Acquisition and Use of Administrative Data

Producing various statistics by using administrative data

Producing Official Statistics

- 48 Official Statistics are using administrative data
  - 15 Administrative Statistics
  - 26 Surveys replace their items with administrative data
  - 7 Surveys use administrative data for editing
- 35 Statistics in 2011 → 48 Statistics in 2020

Providing Statistical Registers

- Providing Microdata of Administrative Statistics and 14 Statistical Registers for researchers and external users

3. Protection of Privacy and Confidentiality

Physical measures for protection of privacy and confidentiality

Intranet system was designed and constructed to protect private information and confidentiality

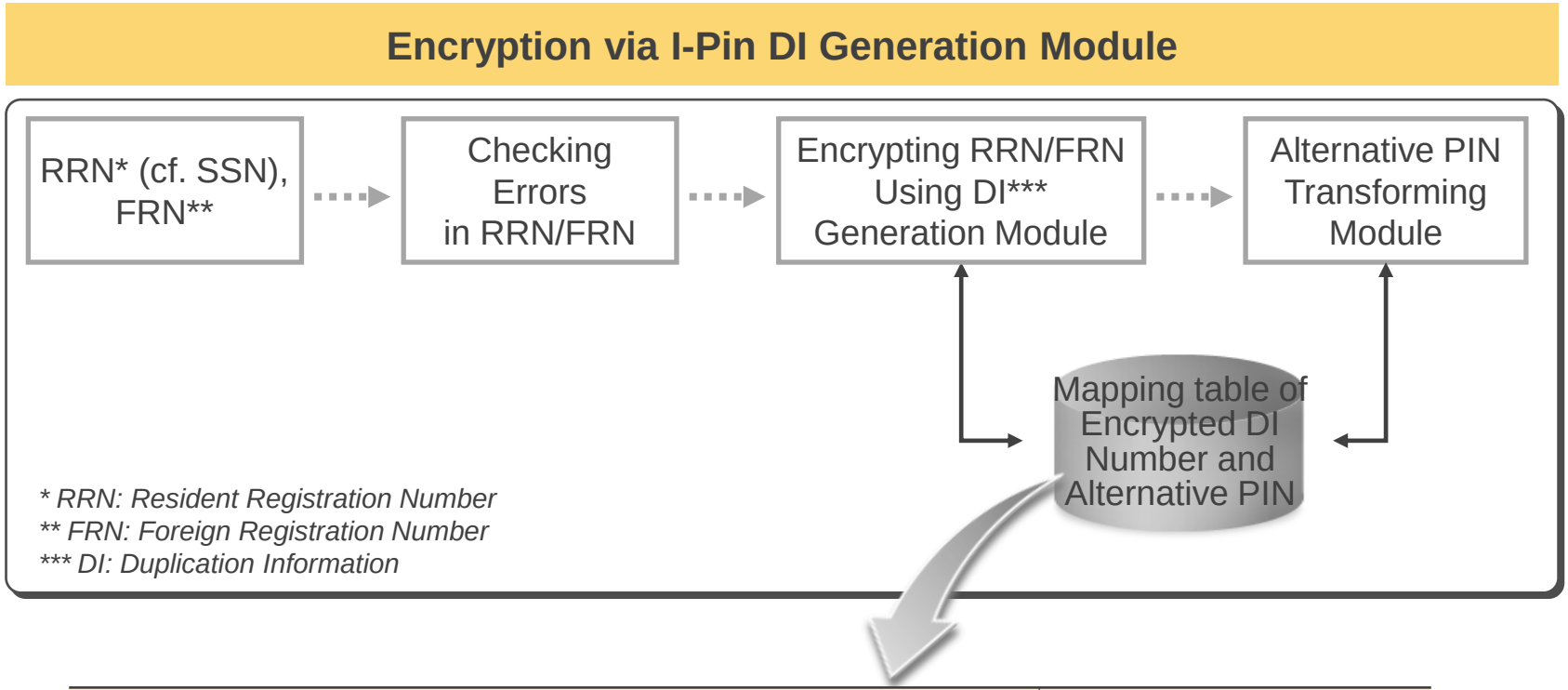
- Internal business systems and network (intranet) was separated from the Internet in 2000
- Fire walls and network control were strengthened

Administrative data were stored and processed safely in the Integrated Management System For Administrative Data

- Use and analysis of administrative data is only possible through access to that system in highly secured servers
- Downloading data or capturing the screen at the PC were forbidden

3. Protection of Privacy and Confidentiality

Encrypting personal information for data security



Encrypted DI Number	Alternative PIN
MC0GCCqGSIb3DQIJAyEAbDSAY…..Cw+3Li4ENiRKJhDHsZXXXXX=	123456789
MC0GCCqGSIb3DQIJAyEAu36eI/W…..DKJhI8pUqcN2SeRaXXXXX=	234567890
MC0GCCqGSIb3DQIJAyEAseIHV37…..ghViiONXgzIQdw8GXXXXX=	345678901

3. Protection of Privacy and Confidentiality

Masking data for protection of privacy and confidentiality

Personal identification numbers (PIN) were replaced to record identification numbers

- Converted PIN to 64-bit encrypted numbers, and then to Alternative PIN

Existence of variables with text is reduced

- Most of text information such as sex, address, academic achievement, etc. in administrative data have been coded

4. Collaboration with Data Provider

Convert Business category code to KSIC*

Building Tax Administrative DB by using Business Register and Tax Data

- Building 3 Tax datasets by tax data from National Tax Service
  - Business Register DB, Value Added Tax DB, Corporate Tax DB
- Used by various official statistics by KOSTAT

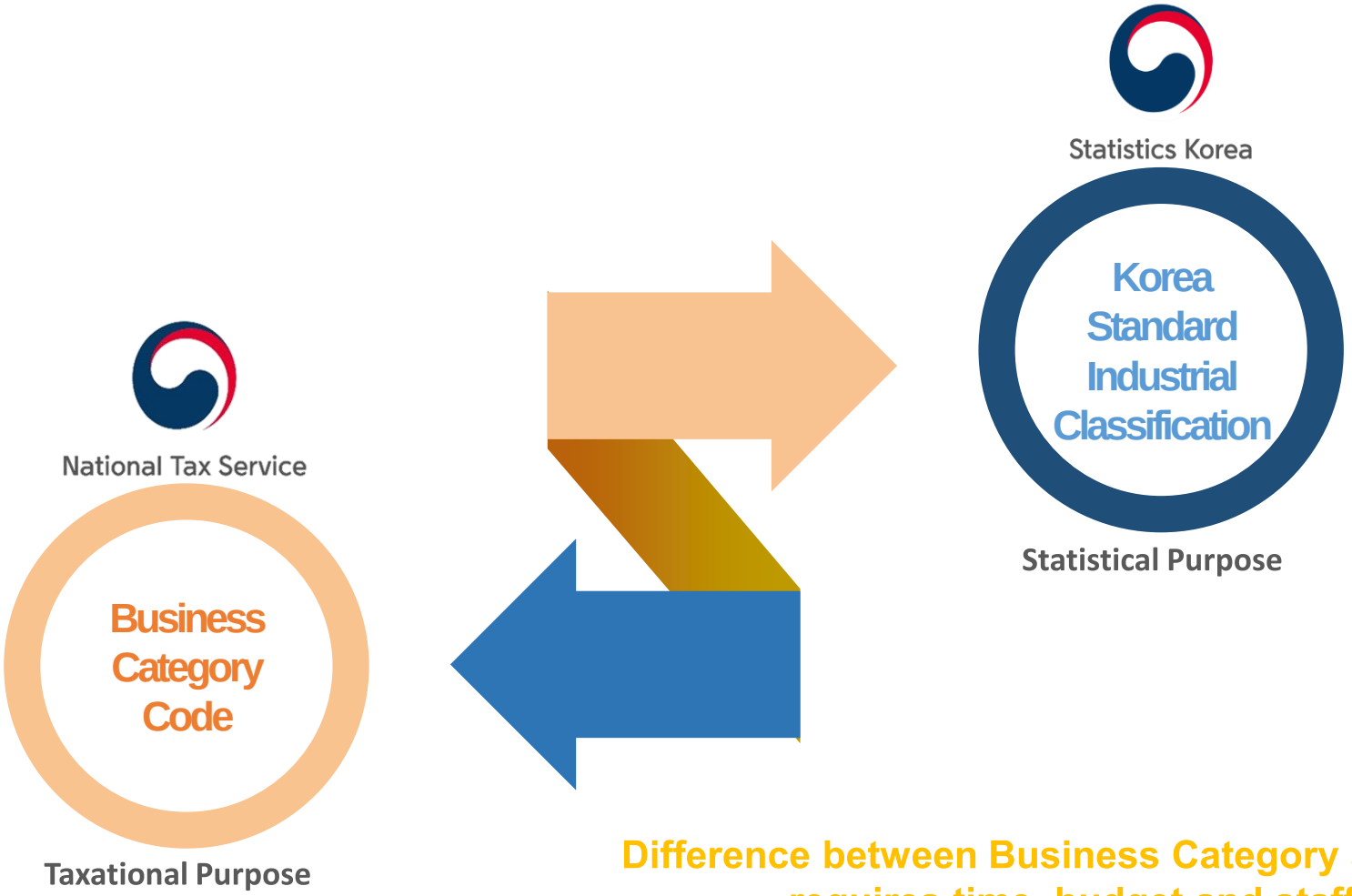
Converting Business Category Code to KSIC*

- Industrial classification is used for understanding the structure of businesses
- Industrial classification is fundamental element for official statistics
- Supporting the decision makers for making economic policies

** KSIC: Korea Standard Industrial Classification*

4. Collaboration with Data Provider

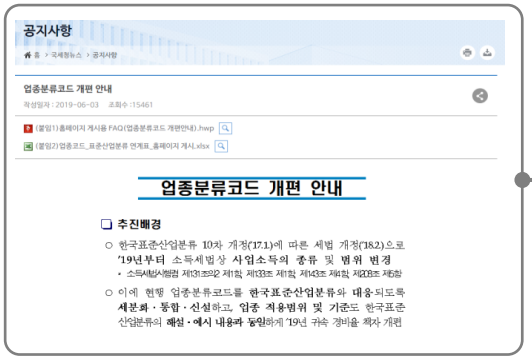
Automatic classification process is highly needed

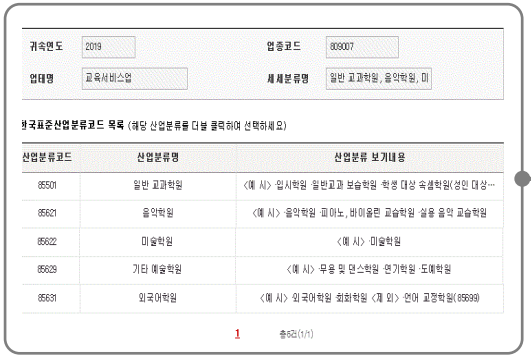


Difference between Business Category and KSIC requires time, budget and staff for transformation

4. Collaboration with Data Provider

Collaboration with NTS has improved Automatic Classification Process





Started to obtain administrative data

2009

Added KSIC selecting function on business register system

2017

Reformed Business Category System of NTS

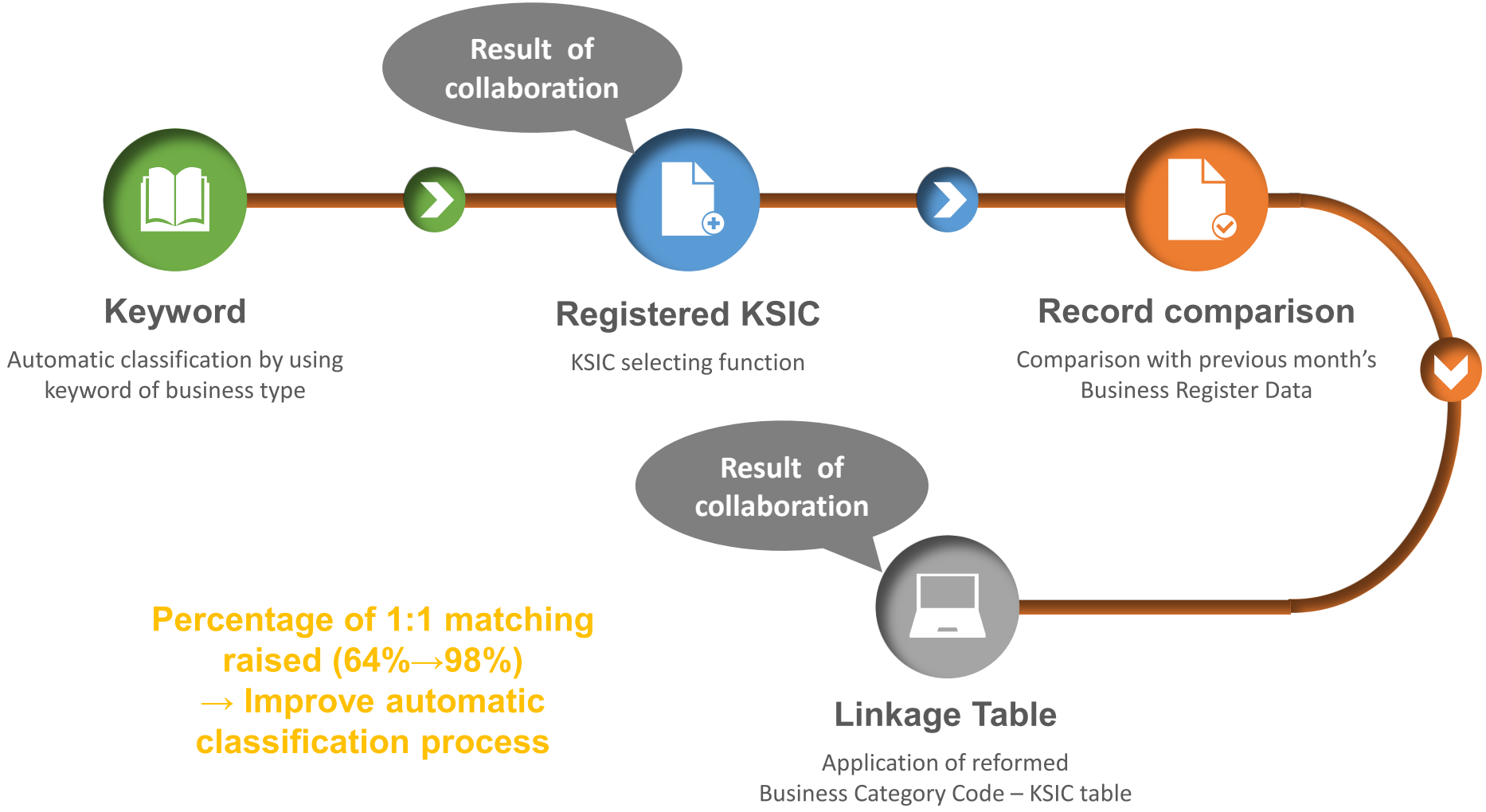
2019

Statistics Korea

National Tax Service

4. Collaboration with Data Provider

Improving Accuracy and Efficiency by applying the result of cooperation



Percentage of 1:1 matching raised (64%→98%)
→ Improve automatic classification process

5. Future Plans and Issues

Building a relevant condition for using administrative data and Data Integration

- Government led policies – Data Dam, Korean New Deal – promote favorable environment for data integration(2020.7)
- COVID-19 boosts untact methods for collecting data
- KOSTAT has carefully kept an eye on these situations for actively taking action

Effort on using and sharing more Administrative Data

- Decentralized statistical system – hard to share data
- Lack of legal basis for sharing Integrated Data with other organizations
- Build a legal framework by amending Statistics Act
- Signing MOU with administrative institute for collecting and sharing administrative data

5. Future Plans and Issues

Increasing ‘Statistical Register’ based on Administrative Data

- Expand Statistical Register by using various data sources for producing higher value-added statistics
  - Analysis on business activities and income status of low-income group
  - Analysis on business status of self-employed
  - Analysis on blind spot of maternity leave
  - Analysis on economic status caused by marriage and childbirth
- Expand linkage between Survey Data and Administrative Data for advancing Statistical Register

Conducting Economic Census Based on Administrative Data

- Register based Population and Housing Census launched since 2015
- Register based Economic Census will launch in 2021



THANK YOU