프로그램

- 일시 : 2016. 8. 31.(수) 10:00 ~ 16:00
- 장소 : 리츠칼튼호텔
- 주관 : 통계청　　　　　　　　- 후원 : 기획재정부

시 간		프로그램	
오전	10:00~10:30 (30′)	등록	
	10:30~10:40 (10′)	개회사	유경준 청장(통계청)
	10:40~10:50 (10′)	축사	최상목 제1차관(기획재정부) 김상호 원장(한국보건사회연구원)
	10:50~11:30 (40′)	빅데이터, 인공지능 그리고 통계 (The New Era of Computing)	Norishige Morimoto (CTO&VP, Technical Leadership, IBM Asia Pacific)
	11:30~12:00 (30′)	좌담	Norishige Morimoto 통계청장 이상구(서울대학교 빅데이터연구원 부원장)
12:00~13:30 (90′)		오찬	
오후	13:30~15:40 (130′)	전문세션 회의 – 총 4개 세션 운영 ※ 2개 주제 발표 후 Coffee Break (10분)	세션별 4~5개 주제 발표 / 토론

차 례 *Contents*

기조연설

빅데이터와 인공지능 그리고 통계
(The New Era of Computing)

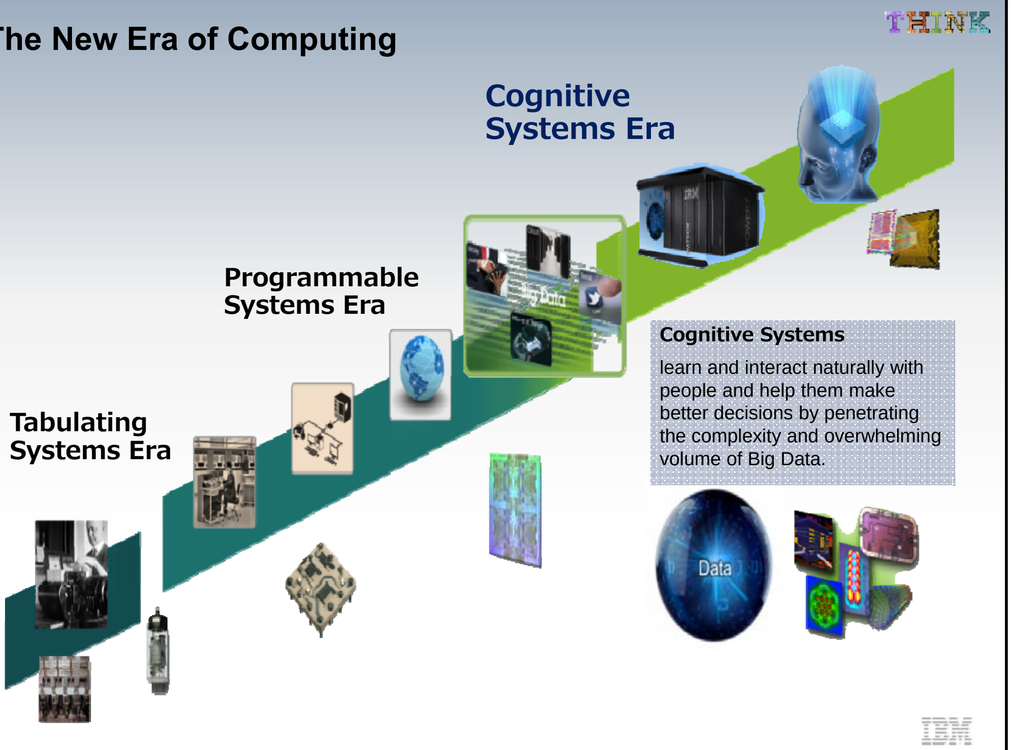
Norishige Morimoto
CTO & VP, Technical Leadership, IBM Asia Pacific

The New Era of Computing

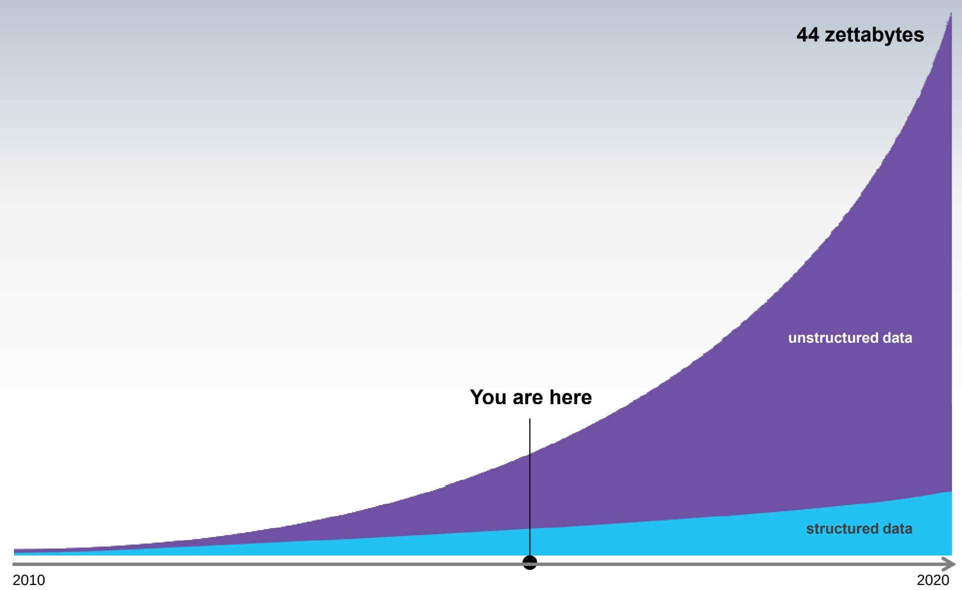
- Big data, AI and Statistics

Noly Morimoto
CTO and Vice President,
Technical Leadership, Asia Pacific
IBM

The New Era of Computing



Data is growing exponentially and demands new approaches (technology and strategy)



3

The amount of ambiguous, uncertain information is growing

(2012 – 2015)

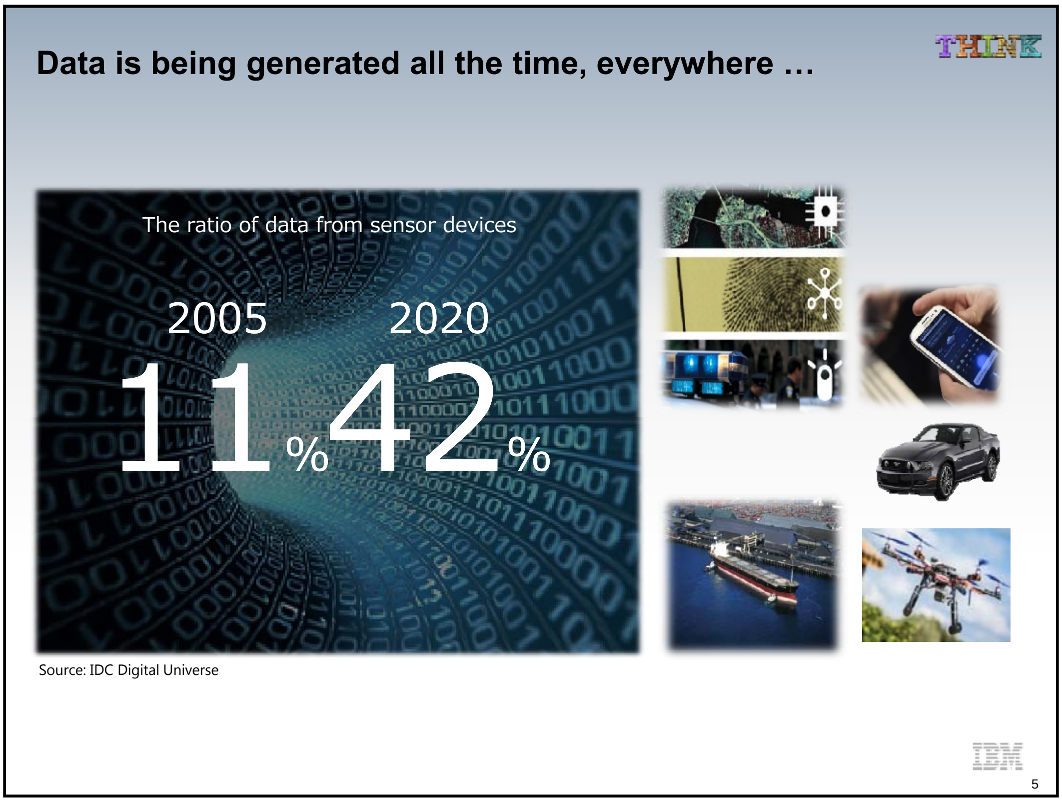
1.8B photos /day

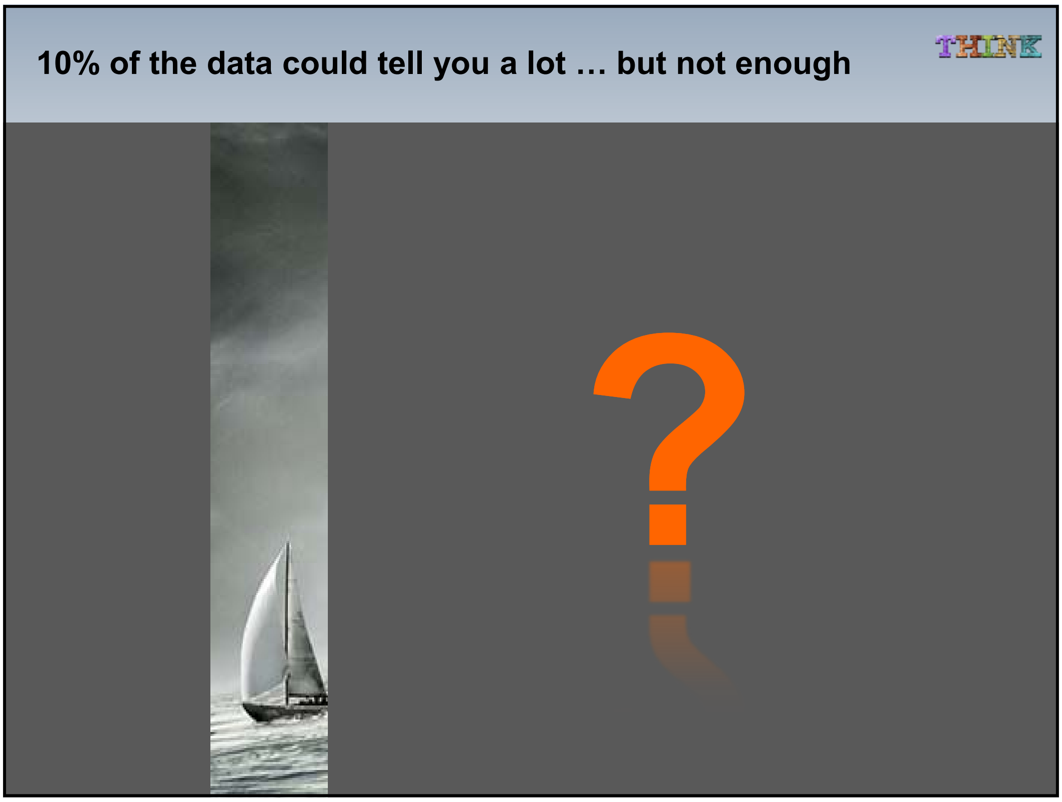
500 hours /minute

700K hours /day

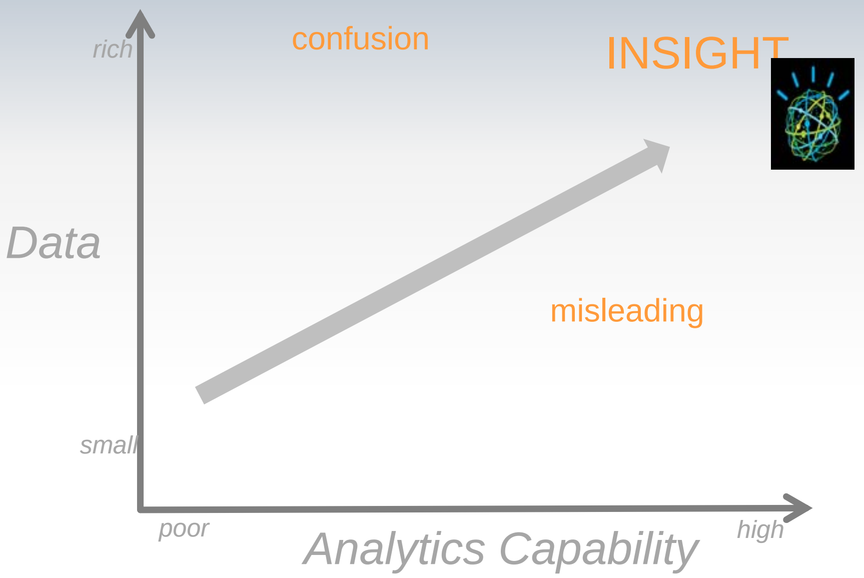
Source: Business insider / MIC / Google

4


Data is being generated all the time, everywhere …
THINK
The ratio of data from sensor devices
2005
2020
11%
42%
Source: IDC Digital Universe
IBM
5


10% of the data could tell you a lot … but not enough
THINK
?

Data is a Natural Resource … yes, but …



7

The Cognitive Era is already here …

Cognitive solutions defined as those that accelerate, enhance and scale human expertise

> **käg-nə-tiv** *(adjective)*: of, relating to, or involving conscious mental activities (such as thinking, understanding, learning, and remembering)

UNDERSTAND
Understands data – structured and unstructured, text-based or sensory – in context and meaning, at astonishing speeds and volumes.

REASON
Has the ability to form hypotheses, make considered arguments and prioritize recommendations to help humans make better decisions.

LEARN
Ingests and accumulates data and insight from every interaction continuously. Is trained, not programmed, by experts who enhance, scale and accelerate their expertise. Therefore, it gets better over time.

Source: IBM Institute of Business Value, "Your cognitive future: How next-gen computing changes the way we live and work ", 2015; Cognitive Naming Guidelines

8

What problem Cognitive Computing is addressing to …

UNDERSTAND ❯ REASON ❯ LEARN

Artificial Problems
e.g. board games, puzzles. Large but limited options with well defined set of rules. The result is countable and easy to define 'win/lose' or 'right/wrong'.

Close Domain Problems
e.g. medical diagnosis, stock investment, legal cases. Large, incremental set of 'high-quality' data with no limit of options. No obvious rules to define 'right/wrong' other than judged by the outcome.

Open Domain Problems
e.g. communication robot, open domain Q&A. Unlimited data, domain with generally available data, with no limit of options. No obvious rules to define objective or outcome, other than the user satisfaction.

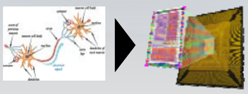
IBM

9

Technologies to drive Cognitive Computing

THINK

Neuromorphic computing



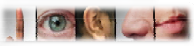
Processing heterogeneous big data

1.8B photos/day

500 hours/minute

MOBILE/IoT Connected devices

Ambiguous data and machine-generated hypotheses

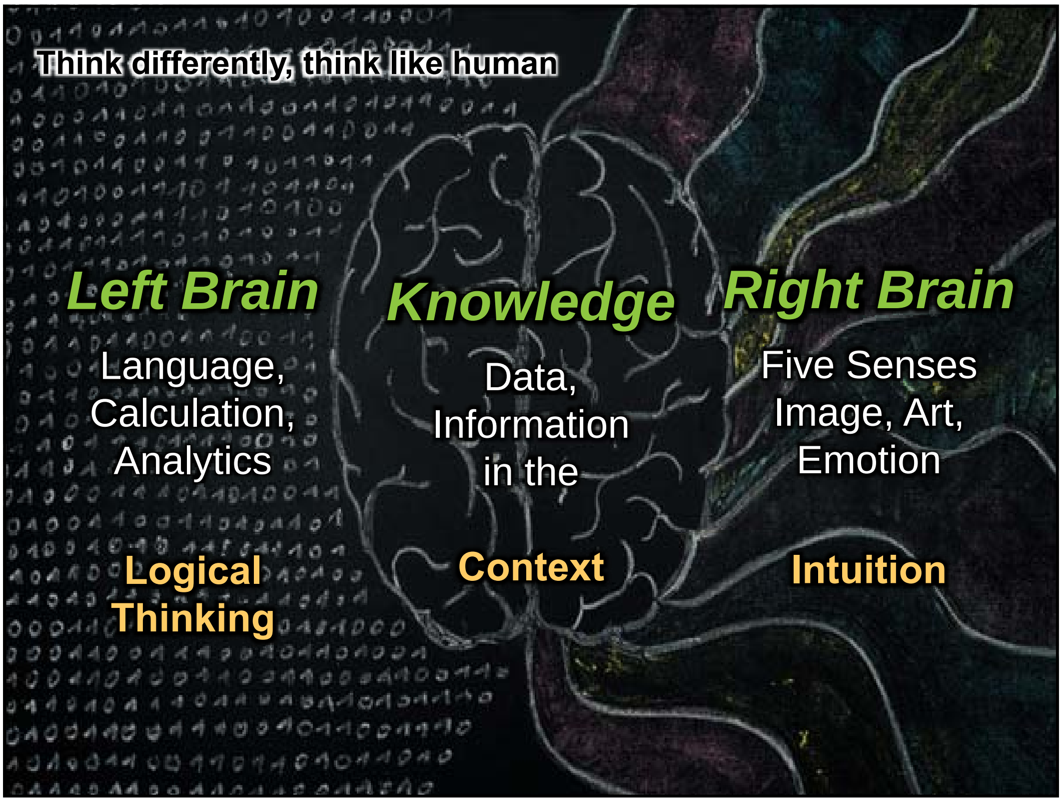


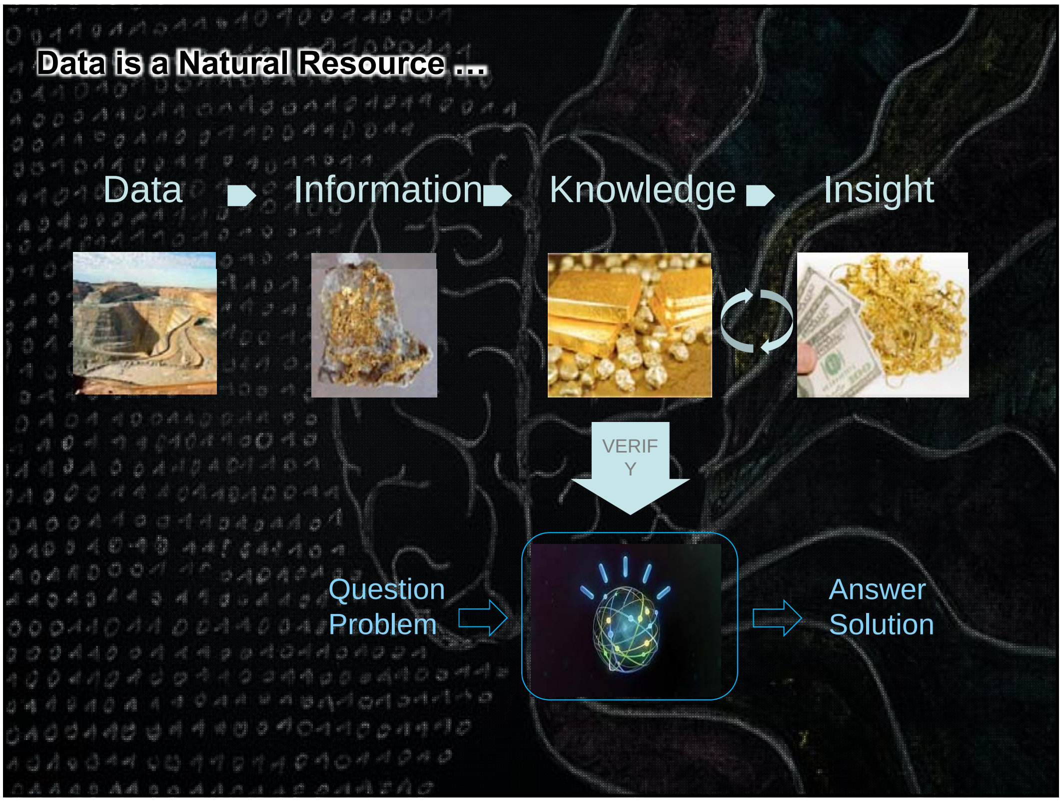
Natural Language Understanding
Natural Language Processing
Question Analysis
Machine Learning
Ontology Analytics
…

Advanced human-machine interactions, dialogue

IBM

10


Think differently, think like human
Left Brain
Language,
Calculation,
Analytics
Logical
Thinking
Knowledge
Data,
Information
in the
Context
Right Brain
Five Senses
Image, Art,
Emotion
Intuition


Data is a Natural Resource ...
Data
Information
Knowledge
Insight
VERIF
Y
Question
Problem
Answer
Solution

Cognitive Capabilities are here, but are you ready?

UNDERSTAND → REASON → LEARN

Defined Problems
e.g. assist operators, medical diagnosis, case judgment, predict failure modes. etc.

Judgment Criteria
e.g. good response, bad response, client satisfaction, disease found, cured, problem identified, shorter down time etc.

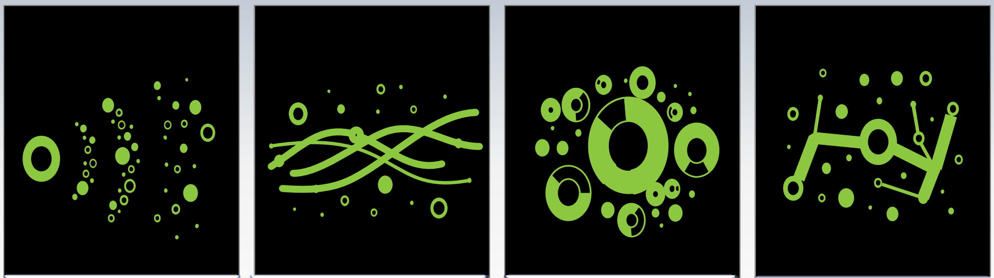
Subject Matter Experts
e.g. doctor, supervisor, who can evaluate and provide feedback and scoring to 'teach' the system.

Data Availability
e.g. call logs, EMR, HER, product specification, problem report, client claim report etc.

Domain Knowledge to be Leveraged
e.g. academic paper, medical literature, scientific papers, dictionary, etc.

IBM 13

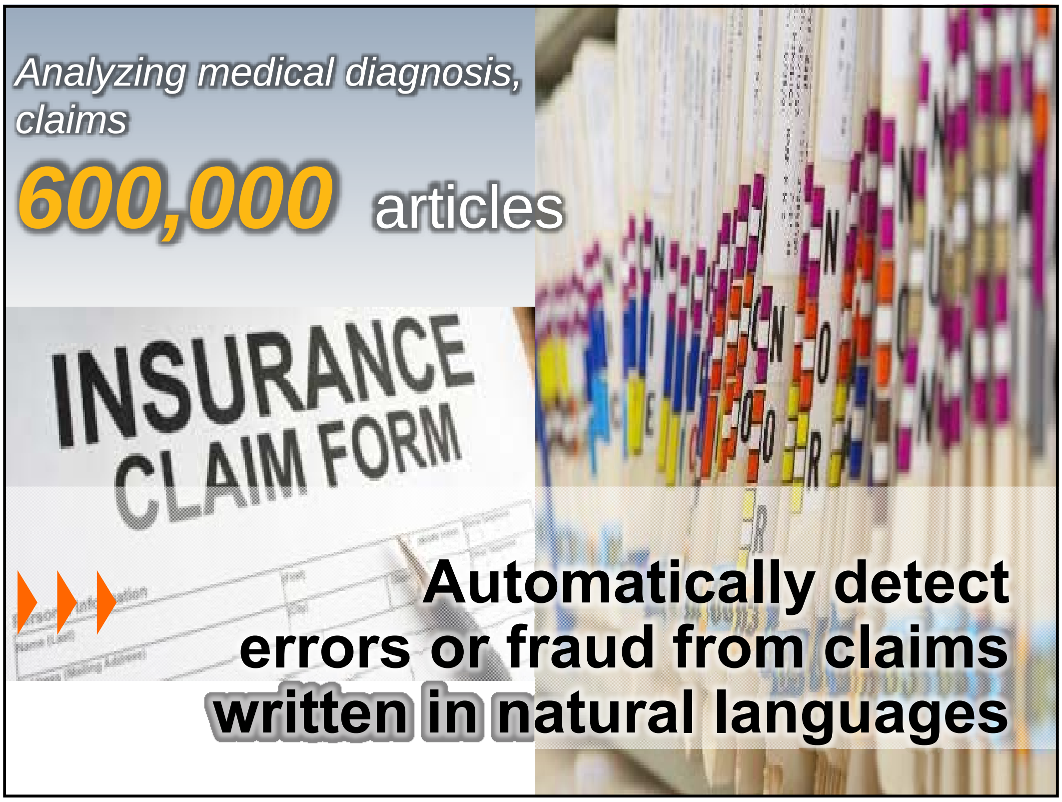
Watson in Different Use Cases



Engagement	Discovery	Decision	Exploration
Helps organizations build stronger relationships with constituents	Help people create new insights by synthesizing information	Help users make more informed evidence-based decisions	Visually depict and analyze data for clear advice

IBM

15

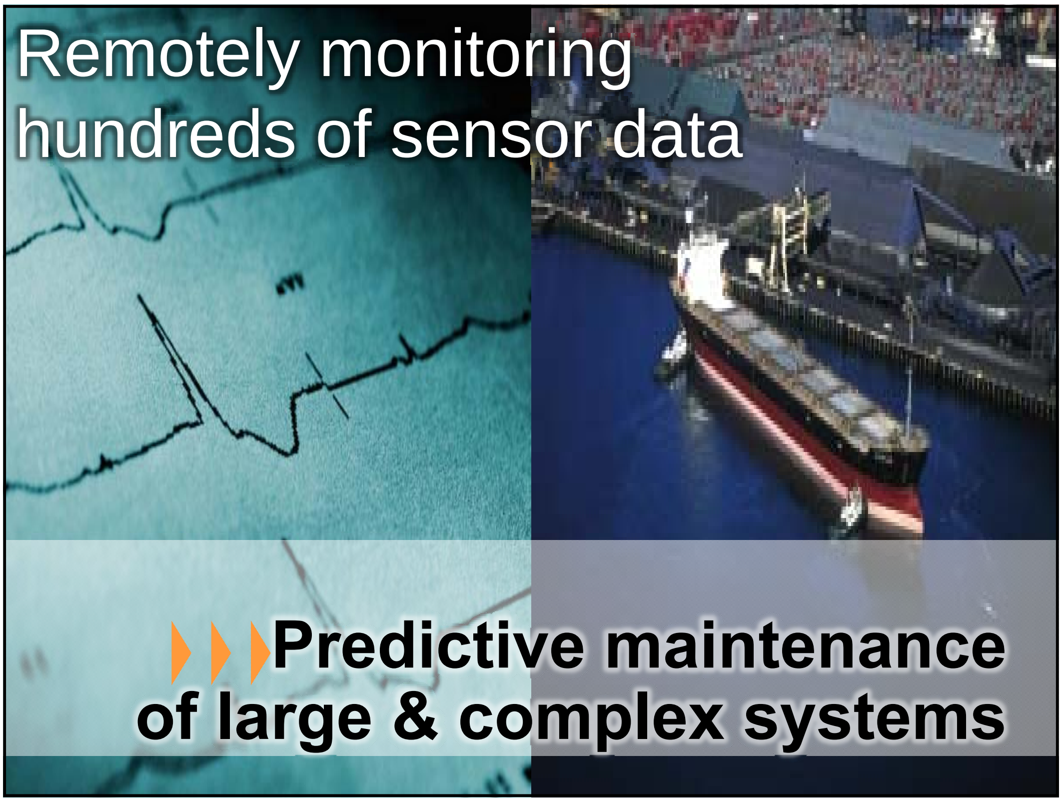

Analyzing medical diagnosis, claims
600,000 articles
INSURANCE CLAIM FORM
Automatically detect errors or fraud from claims written in natural languages


Contact Center
Hundreds of thousands of hours of conversations, chats with customers are recorded
Analyze entity, context, tone and intent

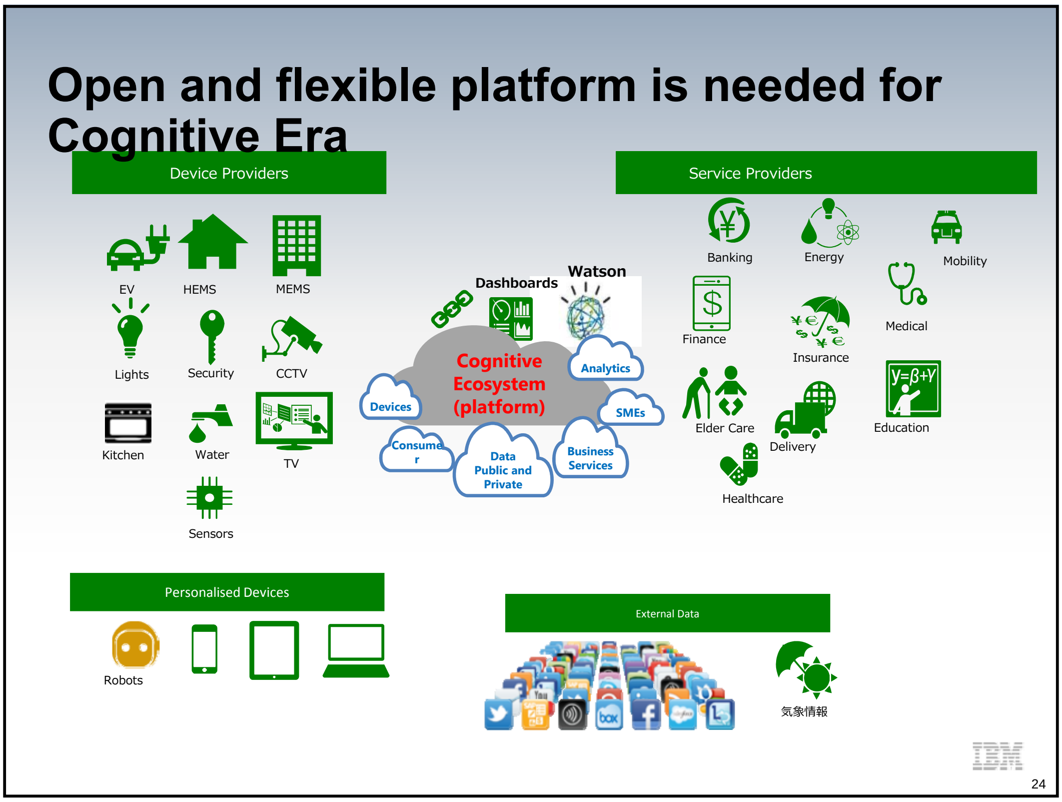

Medical Literature
300K articles / year
Assist scientists on discovery of new drug and material
3:00


Cognitive Computing for Internet of Things (IoT)
Connected Car
Smart Agriculture
Energy and Utility
Security/Surveillance
Mobile Payment
Smarter Home, City & waste
Smart Factory
Trade, Custom Port Management
Environment, pollution
Food Traceability
Transportation & logistics
Disaster response/recovery
IBM
20


IBM
UNDER ARMOUR
HEALTHBOX
YOUR CONNECTED FITNESS SYSTEM.
Personal trainer to
offer advice 24/7


Remotely monitoring
hundreds of sensor data
Predictive maintenance
of large & complex systems


Watson meets Human
Watson with PEPPER
Watson with NAO
IBM Watson
cognitoys
SoftBank
https://cognitoys.com/
IBM
23


Open and flexible platform is needed for Cognitive Era
Device Providers
EV
HEMS
MEMS
Lights
Security
CCTV
Kitchen
Water
TV
Sensors
Dashboards
Watson
Cognitive Ecosystem (platform)
Analytics
Devices
SMEs
Consumer
Data Public and Private
Business Services
Service Providers
Banking
Energy
Mobility
Finance
Medical
Insurance
Elder Care
Delivery
Education
Healthcare
Personalised Devices
Robots
External Data
気象情報
IBM
24

How to better drive Cognitive Capabilities to the industry?

UNDERSTAND → REASON → LEARN

Defined Problems
e.g. assist operators, medical diagnosis, case judgment, predict failure modes. etc.

Judgment Criteria
e.g. good response, bad response, client satisfaction, disease found, cured, problem identified, shorter down time etc.

Subject Matter Experts
e.g. doctor, supervisor, who can evaluate and provide feedback and scoring to 'teach' the system.

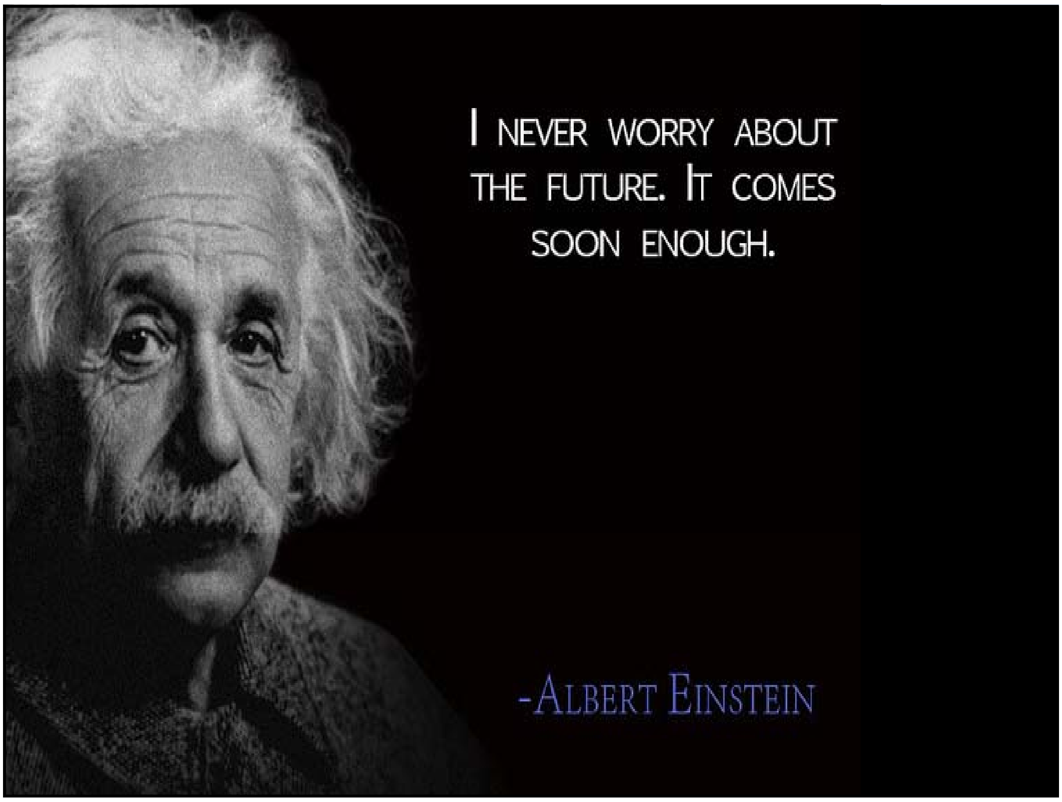
Data Availability
e.g. call logs, EMR, HER, product specification, problem report, client claim report etc.

Domain Knowledge to be Leveraged
e.g. academic paper, medical literature, scientific papers, dictionary, etc.

Government's Roles

- Standardization of Data
- Standardization of Process and Rules
- Structure to Collect and Share Knowledge
- Educate Talent of Domain Experts

25



Uber Moment … sounds familiar?

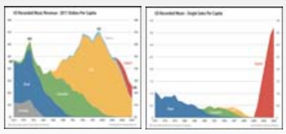
- **Evolution of technology** (experimental, often problematic)
- **Early adapter in the market** (disruptive or even illegal)
- **Disrupt traditional business mode** (and re-define market structure with strong resistance)
- **Winning party to take over the operation** (create rule and trust to gain support from consumer)
- **Evolution of new business model and eco system** (with essential owners and stake holders)

Music

P2P File Sharing

napster 1999 -

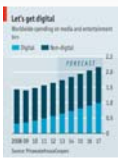
Music Unlimited



Book/Publish

eCommerce

amazon



amazon eBay PayPal NETFLIX Spotify twitter GROUPON 淘宝网 支付宝 阿里云 虾米 聚划算

Taxi/Limo

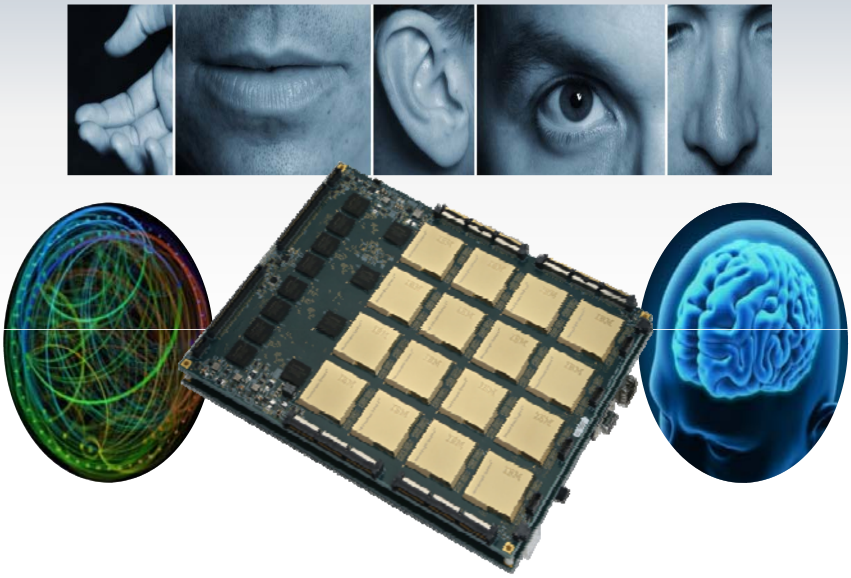
Mobile Application

UBER GRABTAXI

IBM

28

Brain-inspired systems will allow better analysis of sensory data



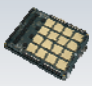
IBM

29

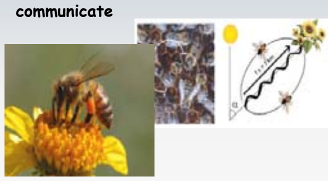
Imagine what bee(ish) brain could do …

Other than collect honey and pollinate.

Collaborate, construct and form community



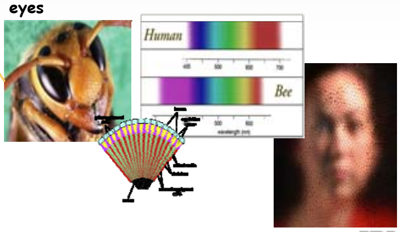
Manueuver, navigate, memorize and communicate



Brain : 1 milllion neurons

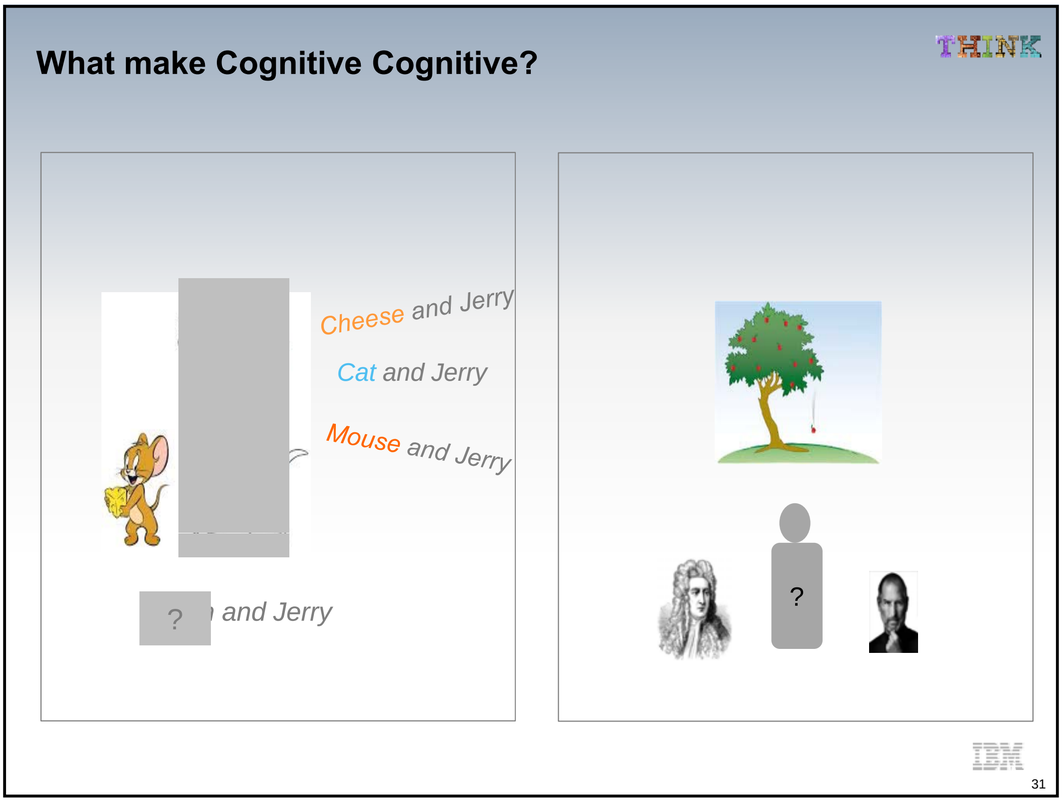
Travel 24km/h up to 10km, with 200 to 600 wing beats per second

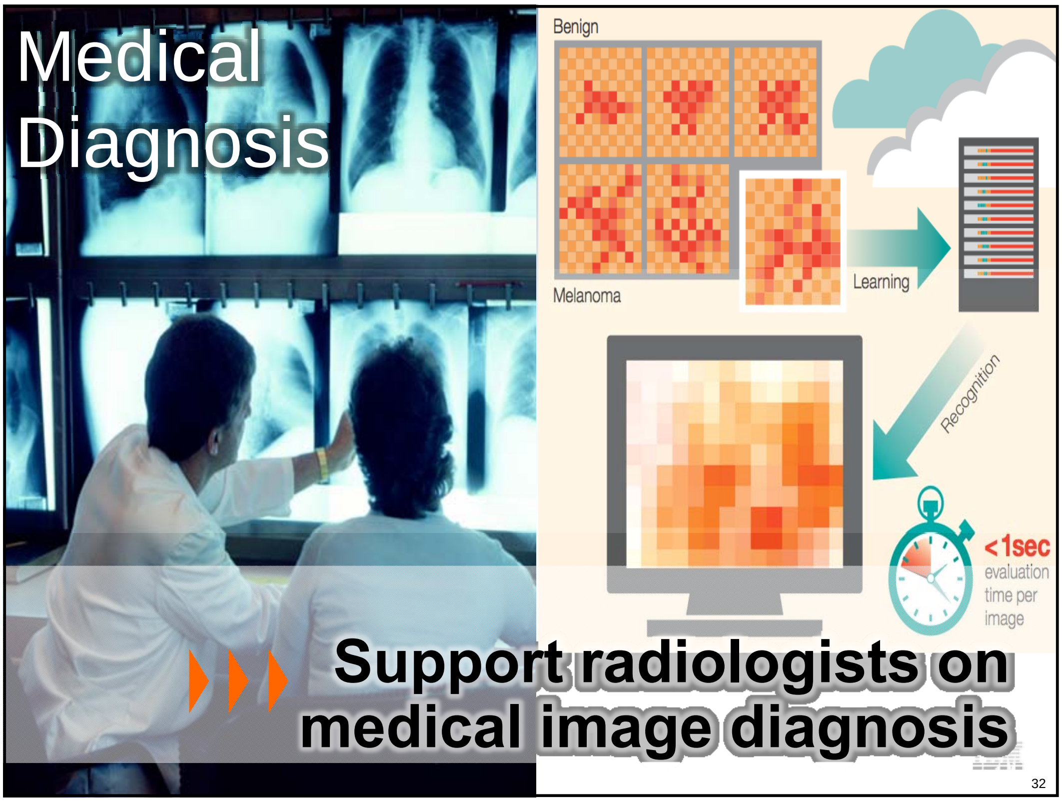
Sense light and images with 2 compound eyes + 3 eyes



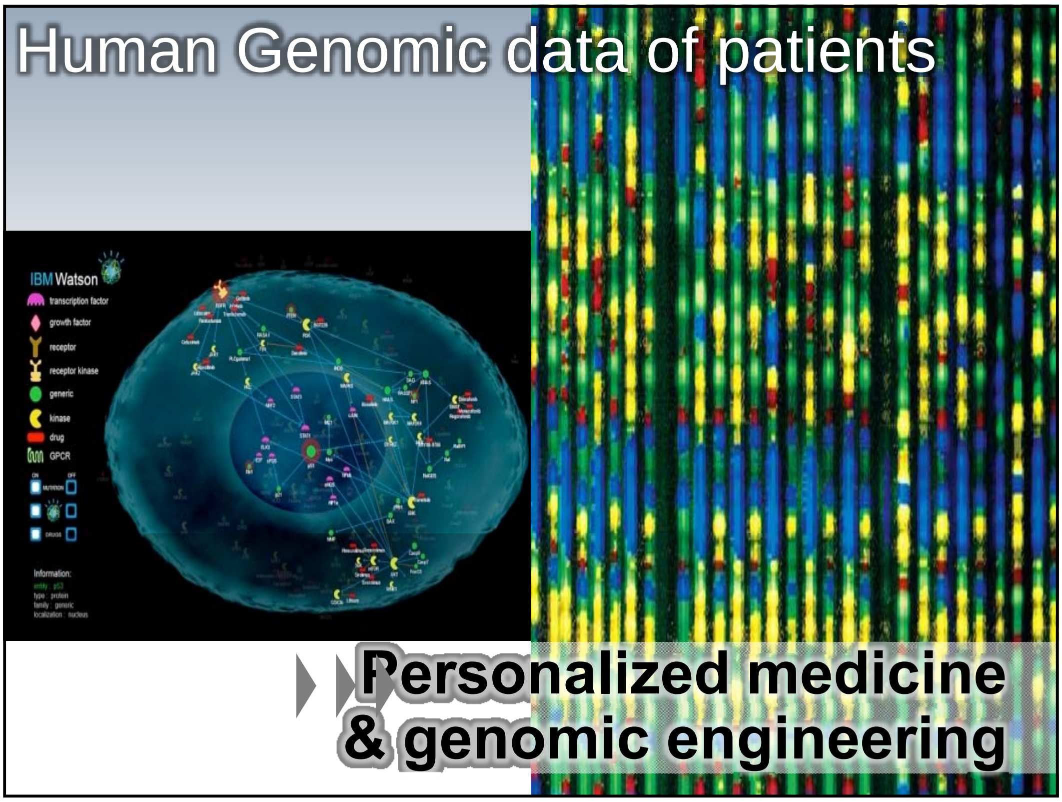
IBM

30


What make Cognitive Cognitive?
THINK
Cheese and Jerry
Cat and Jerry
Mouse and Jerry
?
and Jerry
?
IBM
31


Medical Diagnosis
Benign
Melanoma
Learning
Recognition
<1sec
evaluation time per image
Support radiologists on medical image diagnosis
32


World edible plants in the world
300,000
Cognitive Cooking with Chef Watson
Recipes for Innovation from IBM & the Institute of Culinary Education
NOPE
RED CHILE
LEMON
PEAR
BEEF
Discover new tastes without prejudice


Human Genomic data of patients
IBM Watson
Personalized medicine & genomic engineering

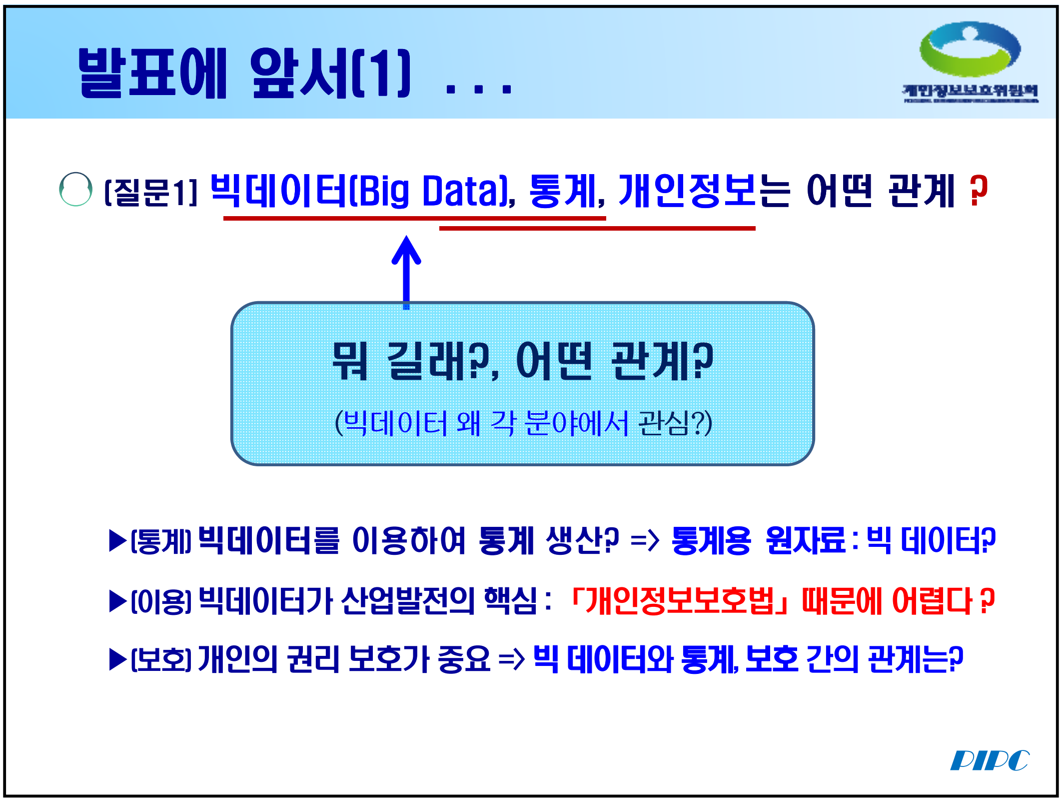

Stronger and healthier Cocoa Bean
400M/base DNA pairs
Computational biology
to help improve
productivity

제1세션

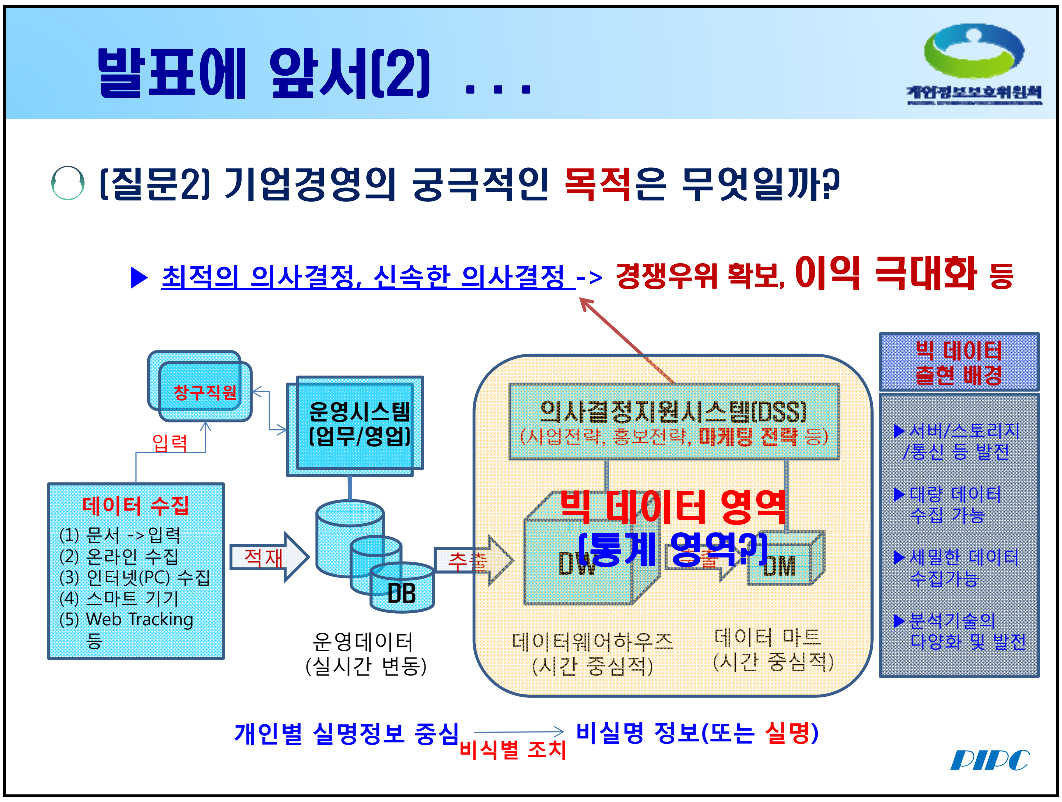
빅데이터와 국가통계

- 빅데이터와 통계, 그리고 개인정보보호 / 박원환 과장
- 빅데이터의 국가통계 활용 전략 / 정용찬 그룹장
- 빅데이터 활용 통계생산 방법 및 품질검증 / 박민규 교수
- 빅데이터 거버넌스 구축 및 활용사례 / 오세안 사무관

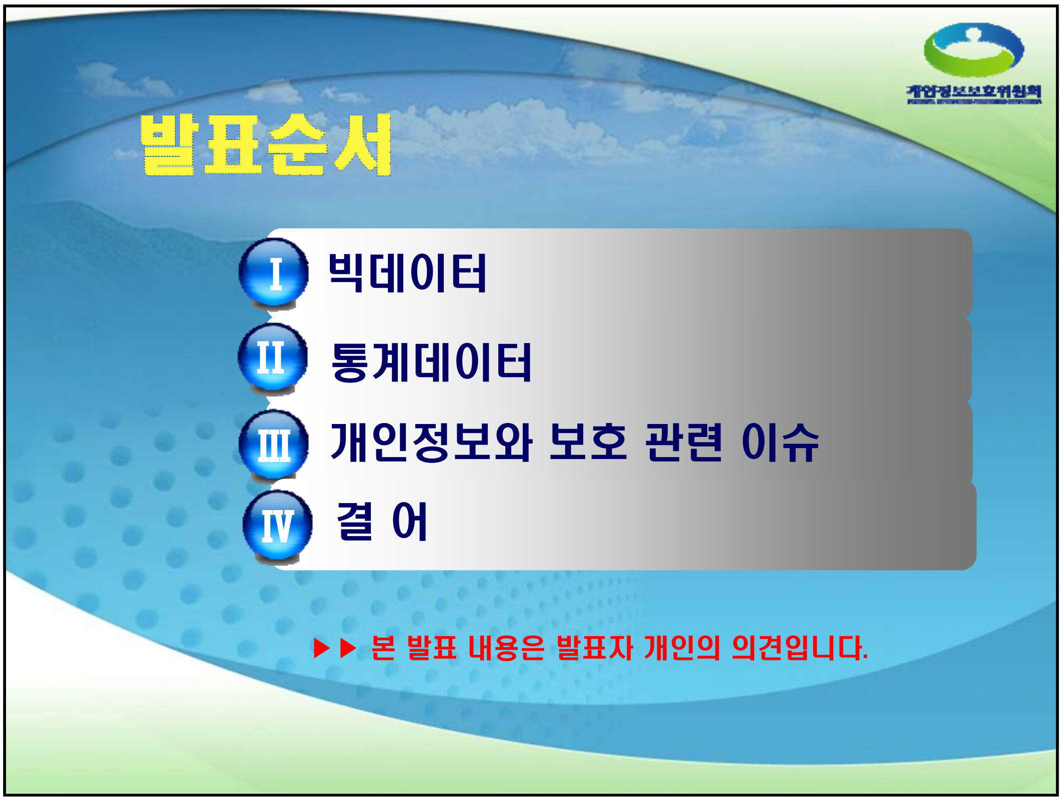
발표자료
빅 데이터와 통계,
그리고 개인정보보호
2016. 8. 31.
박 원 환



발표에 앞서(1) . . .
개인정보보호위원회
[질문1] 빅데이터(Big Data), 통계, 개인정보는 어떤 관계 ?
뭐 길래?, 어떤 관계?
(빅데이터 왜 각 분야에서 관심?)
▶(통계) 빅데이터를 이용하여 통계 생산? => 통계용 원자료 : 빅 데이터?
▶(이용) 빅데이터가 산업발전의 핵심 : 「개인정보보호법」 때문에 어렵다 ?
▶(보호) 개인의 권리 보호가 중요 => 빅 데이터와 통계, 보호 간의 관계는?
PIPC



발표에 앞서(2) . . .
개인정보보호위원회
(질문2) 기업경영의 궁극적인 목적은 무엇일까?
▶ 최적의 의사결정, 신속한 의사결정 -> 경쟁우위 확보, 이익 극대화 등
창구직원
입력
운영시스템 (업무/영업)
데이터 수집
(1) 문서 ->입력
(2) 온라인 수집
(3) 인터넷(PC) 수집
(4) 스마트 기기
(5) Web Tracking 등
적재
DB
운영데이터 (실시간 변동)
추출
의사결정지원시스템(DSS)
(사업전략, 홍보전략, 마케팅 전략 등)
빅 데이터 영역
(통계 영역?)
DW
DM
데이터웨어하우즈 (시간 중심적)
데이터 마트 (시간 중심적)
빅 데이터 출현 배경
▶서버/스토리지 /통신 등 발전
▶대량 데이터 수집 가능
▶세밀한 데이터 수집가능
▶분석기술의 다양화 및 발전
개인별 실명정보 중심
비식별 조치
비실명 정보(또는 실명)
PIPC



개인정보보호위원회
발표순서
I 빅데이터
II 통계데이터
III 개인정보와 보호 관련 이슈
IV 결 어
▶▶ 본 발표 내용은 발표자 개인의 의견입니다.

Ⅰ. 빅데이터

PIPC

Ⅰ – 1. 빅데이터(Big Data) 란? –1/2

[정의1] 빅데이터는 데이터를 수집 · 관리 · 처리 등 통상적인 작업에 사용하는 *소프트웨어의 수용한계를 넘어서는 크기의 데이터*를 말한다.[1]

▷ 출처[1] : Snijders C 외 1인, "'Big Data' : Big gaps of knowledge in the field of internet", International Journal of Internet Science, 7, 1-5, 2012.

[정의2] 빅데이터란 기존 데이터베이스 관리도구로 데이터를 수집 · 저장 · 관리 · 분석할 수 있는 역량을 넘어서는 *대량의 정형 또는 비정형 데이터의 집합*[1] 및 이러한 데이터로 부터 *가치를 추출하고 결과를 분석하는 기술*을 의미[2] 한다.

▷ 출처 [1] : James Manyika & Michael Dhui, "Big data : The next frontier for innovation, competition, and productivity.", McKinsey Global Institute, p1, 2011.5.
[2] : John Gantz & David Reinsel, "Extracting Value from Dhaos", IDC IVEW, p6, 2011.6.

PIPC

Ⅰ - 1. 빅 데이터 란? - 2/2

[정의3] 메터 그룹(현재, 가트너)의 에널리스트 더글라스 레이니의 보고서에서[1] 빅데이터를 " 3V" 로 모델화(2001)하여 3차원으로 정의

▷데이터 양(Volume), 생성/증가 속도(Velocity), 데이터 종류의 다양성(Variety)

▷ 출처[1] : Laney, Douglas, **3D Data Management** : Controlling Data Volume, Velocity, Variety, 2001.2.6.

[정의3-1] 빅데이터는 큰 용량, 빠른 속도, 그리고 높은 다양성을 갖는 정보자산으로서,
이를 통해 의사결정, 통찰지식 발견, 프로세스 최적화 등을 위한 새로운 형태의 처리방식 (2012년에 가트너가 기존의 정의를 개정)[2]

▷ 출처[2] : Laney, Dougas, The Importance od 'Big Data' : A Definition., Gartner, 21 June. 2012.

▣ **4V 정의** : 3V + 정확성(Veracity)(IBM) 또는 가변성(Variability)(Brian Hopkins)
▣ 그래도 아직까지 **3V가 대세**

PIPC

Ⅰ - 2. 출현배경

[배경1] 컴퓨팅 기술의 발전으로 빅 데이터 작업이 가능한 환경으로 변화

▷서버 통합(클라우드 컴퓨팅, 고성능화), 저장장치(대용량화, 저가 장비 등)

[배경2] 세밀한 데이터, 대량의 데이터 수집이 가능한 환경으로 변환

▷세밀한 데이터 : 휴대폰 등 스마트 기기 발전, 대형 할인마켓, 신용카드 등
▷대량 데이터 : 멀티미디어 데이터, 수집 용이성 등으로 정형/비정형 수집가능

[배경3] 분석 기술의 발전 : DB -> DW-> Big Data

▷[배경1]+[배경2] -> 대량 데이터에 직접 적용(표본->전체)
▷처리능력의 향상 -> 새로운 분석기술 개발이 가능(모의 실험시간 단축 등)

PIPC

Ⅰ – 3. 빅데이터 관련 용어

개인정보보호위원회

①빅데이터, ②빅데이터 분석, ③빅데이터 (분석 및) 활용

=> 위 세가지 용어를 총칭/혼용하여 사용 ? 시각과 관점에 따라 의미가 다름

▷빅데이터 : 수집한 데이터 그 자체(또는 분석 결과까지 포함한 데이터)
▷빅데이터 분석 : 빅 데이터에 특정기법을 적용하여 분석, 결과를 추출하는 행위
▷빅데이터 활용 : 빅 데이터 분석 결과를 특정 목적을 위해 활용하는 것

※ 데이터 분석의 목적

▷(1차) 미래 예측
↓
▷(2차) 정확한 의사결정
↓
▷(3차) 이익 극대화 등

▷보다 정확한 예측 욕망
=> 빅데이터에 대한 수요 증대

▷방대하고, 세분된 데이터 수집
▷다양한 분석 기법의 발전
▷맞춤형 서비스 등 목적 세분화

수집 → 분석 → 활용

=> 분석기법은 과거와 거의 유사하지만, 데이터 양과 정밀도가 높아졌다

PIPC

Ⅰ – 4. 빅데이터의 수집 · 분석 · 활용

개인정보보호위원회

기본 인프라 : 하둡 + NoSQL

빅데이터 → 분석기술 → 결과표현기술 → 활용사례

수집 → 분류 → 빅데이터

수집방법
(1) 문서 ->입력
(2) 온라인 수집
(3) 인터넷(PC) 수집
(4) 스마트 기기
(5) Web Tracking
(6) IoT 등

(기존 분석기술)
▷데이터 마이닝
▷기계학습(인공지능)
▷자연어 처리
▷패턴인식 등

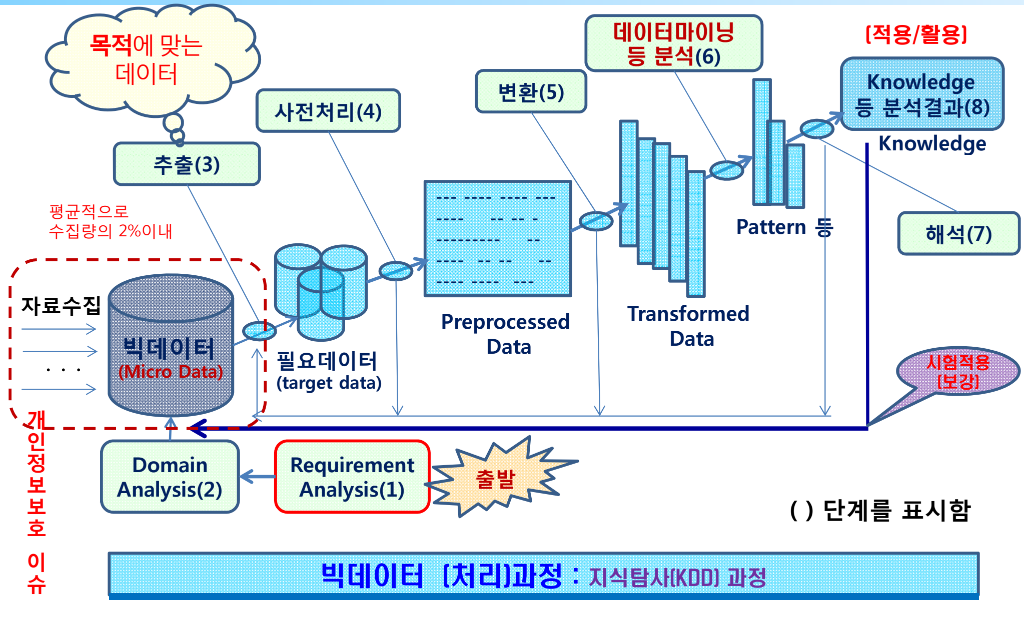
(최근 관심 기술)
▷텍스트 마이닝
▷오피니언 마이닝
▷소셜 네트워크 분석
▷군집분석 등

(주요 도구)
▷R : 무료(선호)
-구글에서 채택
▷SAS : 유료(고가)
- 연간 사용료 지불
- 여론조사 기업 등
▷SPSS : 통계분석
- 구매
▷기타

▷정치 및 사회
-미국 대통령 선거(2008)
-한국 19대 총선(2012)
▷경제 및 경영
-아마존닷컴의 추천상품
-구글의 맞춤형 광고 등
▷문화
- MLB의 머니볼 이론 및 데이터 야구
▷과학기술 응용
- 생물정보학, 의료 등
▷ 기타 분야
- 마케팅
- 구글 번역 등

PIPC

Ⅱ-5. 빅데이터 분석의 전 과정

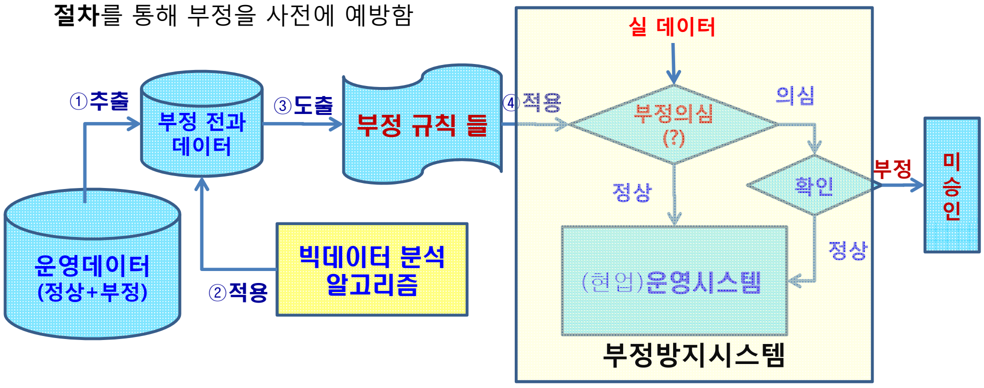


PIPC

[참고] 빅데이터 활용사례(FDS)

❑ 부정방지시스템(FDS) 개발을 위한 <u>부정규칙 발굴</u>

- FDS(Fraud Detection System)은 과거 부정한 경우의 **전과자 자료**를 추출하여, 이를 입력자료로 하는 빅데이터 분석을 통하여 **부정 규칙(Rules)**을 도출
- 도출된 부정 규칙들을 (현업)**운영시스템**에 적용하여 **유사한 경우에 경고·확인하는 절차**를 통해 부정을 사전에 예방함



※ 2014년 카드사 개인정보 유출사고 이후, FDS개발 관련 언론보도자료를 종합

PIPC

Ⅱ. 통계데이터

PIPC

Ⅱ-1. 통계작성 현황

❑ 『통계법』에 따라 390개 기관에서 935종의 통계를 작성 중이며, 이중에 개인(정보주체)과 관련되는 통계는 471종이 있음 (2015년 말 기준)

통계작성기관		작성방법(개인 관련 통계)			
구분	기관수	합계	조사통계	보고통계	가공통계
합계	**390**	**935(471)**	**414(143)**	**444(298)**	**77(3)**
통계청	1	58(22)	41(19)	2(2)	15(1)
중앙행정기관(통계청 외)	42	282(65)	129(34)	137(29)	16(2)
지방자치단체	260	428(356)	140(96)	261(260)	27(-)
지정기관(금융기관, 협회 등)	87	167(28)	104(21)	44(7)	19(-)

❑ 이 중에서 '조사통계'는 응답자로부터 직접 정보를 수집하는 경우이고,

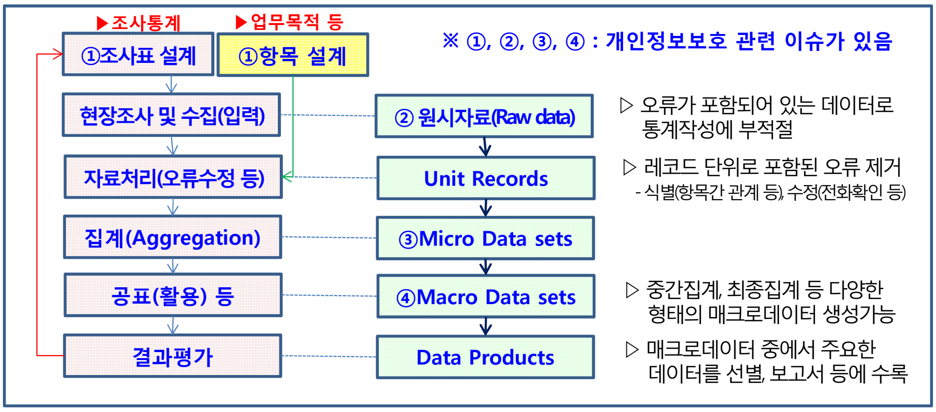
- '보고통계'는 업무목적으로 수집한 정보 중 일부를 이용하여 직접 작성한 통계 또는 통계작성기관이 정보를 제공받아 작성한 통계를 말함

PIPC

Ⅱ-2. 통계데이터 유형

❑ 현장조사 또는 업무목적 등으로 수집한 데이터를 통계작성에 적합하도록 적절히 처리(processing)한 결과 데이터를 통계데이터라 함

<통계작성 생명주기> <처리 단계별 통계데이터 유형>

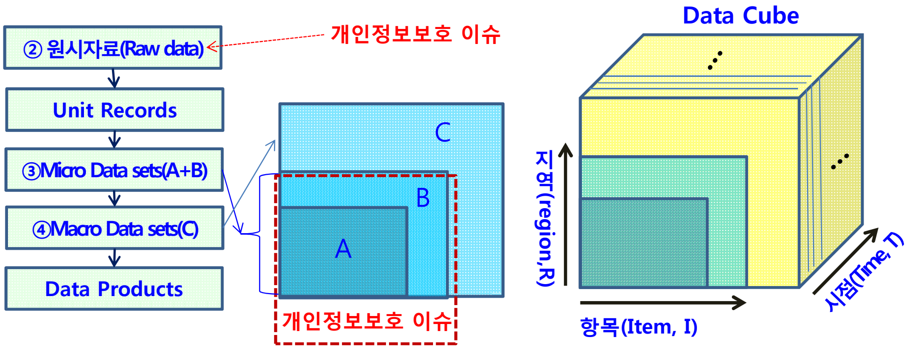


출처 : 박원환, 황조연, "통계자료의 비밀보호를 위한 방법들", 통계연구, 제9권 제2호, 2004, pp 146-172,

PIPC

Ⅱ-3. 통계데이터의 논리적 구조

❑ 통계데이터는 논리적으로 다차원(multi-dimension) 구조를 가지며, 성격과 내용이 다른 n개의 축(X, Y, Z, . . .)이 있음



A : 실제 조사한 자료,
B : 실제 조사과정에서 오류, 누락, 무응답 등을 보정한 자료(imputed data)
C : 집계 데이터(산업분류별, 지역별, 시점별)

▶ 산업분류(세세 → 세 → 소 → 중 → 대), 지역(읍면동 → 시군구 → 시도 → 전국), 시점(일→월→년 ..)

PIPC

Ⅱ-4. 통계 마이크로 데이터 서비스 유형

개인정보보호위원회

서비스 형태		자료특징 및 내용	보호위협
통계표(tables) 형태의 서비스		▷심층이용자 요구에 따라 생성하거나 생성이 가능하도록 한 통계표 ※지나치게 상세하면 **개인식별 가능성** 있음	있음
비식별 조치한 마이크로 데이터	공공이용파일 (public used file)	▷외부 이용자를 위하여 저장매체에 마이크로 또는 매크로 데이터를 수록/제공 또는 판매	없음 (무단복제 제외)
	인가파일 (licensed file)	▷ 특정한 외부 심층이용자를 위하여 제공되며, 제공 전에 법적 보증을 위한 서명이 필요한 자료	있음
원격접속제도		▷통신망을 통하여 원격에서 이용자가 마이크로 데이터를 이용하게 하는 제도	있음
데이터 실험실		▷ 통계기관이 엄격한 심사 및 감독하에 지정 장소에서 마이크로 데이터를 이용하게 하는 제도	있음

※ 현재, 통계청의 "국가 마이크로 데이터 서비스(MDIS)"에서 서비스 중

※항시 조심

PIPC

Ⅲ. 개인정보와 보호 관련 이슈

개인정보보호위원회

Ⅲ- 1. 개인정보란?

Ⅲ- 2. 개인정보 수집유형

Ⅲ- 3. 개인정보 보호 관련 법령에서의 이슈

Ⅲ- 4. 개인정보의 비식별 조치 관련 이슈

Ⅲ- 5. 빅데이터 분석과 통계(분석)의 비교

PIPC

Ⅲ-1. 개인정보란?

제2조(정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다.(개인정보보호법)

1. '**개인정보**'란 **살아 있는 개인에 관한 정보**로서 성명, 주민등록번호 및 영상 등을 통하여 **개인을 알아 볼 수 있는 정보**(해당 정보만으로는 특정 개인을 알아 볼 수 없더라도 **다른 정보와 쉽게 결합**하여 알 아 볼 수 있는 것을 포함 한다)를 말한다.

▶ **"개인을 알아볼 수 있는 정보"**의 의미는 개인정보처리자의 입장에서 특정 개인을 알아볼 수 없다면 개인정보가 아님

- 즉 제공하는 자와 제공받는 자에 따라 개인정보 여부의 판단이 다를 수 있음

▶ **"다른 정보와 쉽게 결합"**의 의미는 결합대상이 될 다른 정보의 입수 가능성이 있어야 하고, 다른 정보와 결합도 쉬워야 함

- 즉, 합법적으로 정보를 수집할 수 없거나 결합을 위해 불합리할 정도의 시간, 비용 등이 필요한 경우라면, '쉽게 결합'을 할 수 있는 상태라 볼 수 없음

※ 출처 : '개인정보 비식별 조치 가이드라인', 행자부 등 6개 부처 합동, 2016. 6.30

PIPC

Ⅲ-2. 개인정보 수집유형

(1) 제공정보ⓐ : 정보주체가 서비스 이용을 위해 자발적으로 제공한 정보,
(정보주체로 부터) **개인정보보호법 제15조**에 따라 수집한 개인정보
(예 : 이름, 주소, 주민번호 등 **식별정보(개인)**를 포함한 정보)

(2) 자동수집정보ⓑ : 서비스의 이용 과정 또는 그 결과로 생성되어지는 정보 등
(정보주체 이외로 부터) (예 : 쿠키정보, 구매정보, 서비스 이용내역, IMEI 등 **식별정보**를 포함)

※ 스마트폰의 원시DB(Native DB) 자료, 인터넷 등의 이용내역 정보, 웹 트레킹 정보, 장바구니 정보 등

※ 통계작성을 위해 행정기관, 기업 등 제3자로 부터 제공받은 데이터

(3)분석정보ⓒ : **ⓐ, ⓑ, ⓐ+ⓑ, ⓐ+ⓑ+(ⓒ), …** 등을 기초하여 분석, 재생산 정보
(재생산 정보) (예 : 구매성향, 주요 관심사, 정치적 성향 등과 같은 **정보)**

※ 식별성은 다른 정보와 결합 여부에 따라 달라짐

주) ( )는 반복적으로 입력자료로 활용될 경우도 포함함

PIPC

[참고] 개인정보의 수집유형 사례

정보주체 제공정보(ⓐ)				자동수집(ⓑ)			가공·재생산(ⓒ)	
주민번호	이름	주소	전화	통화시간	메시지	데이터	특징	영업방향
880130-	홍길동	상도동	2100	02:00:20	5	30GB	낮음	유지
970221-	설중매	잠원동	2111	03:30:20	10	20GB	보통	유지
950127-	양귀비	서초동	2122	04:20:00	30	10GB	중상	유지
990131-	김길동	방배동	2134	02:30:30	50	05GB	데이터	요금제 변경
000130-	차길동	반포동	2133	01:00:30	40	39GB	보통	유지
990222-	박길동	일산동	2130	05:20:30	70	45GB	데이터	요금제 변경
010212-	황길동	역삼동	2154	02:34:00	33	10GB	중상	유지
990313-	정길동	양재동	2142	05:00:30	90	80GB	데이터	요금제 변경

개인정보

항목 : 4개 **항목 : 7개(+3)** **항목 : 9개(+2)**

사실자료(신뢰도 ≒ 100%) **분석자료(신뢰도 ?)**

PIPC

Ⅲ-3. 개인정보 보호 관련 법령에서의 이슈(1/2)

▷**개인정보보호법 제18조 제2항 제4호**

4. 통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 <u>특정 개인을 알아 볼 수 없는 형태로</u> <u>개인정보</u>를 제공하는 경우

▶ **(이슈1) 비식별 조치를 하였는데도 "개인정보"라 한 이유?**

=> 앞서 논의한 바와 같이 비식별 조치 가이드라인에 따르면, 개인정보처리자의 입장에 따라 개인정보의 기준이 다르기 때문으로 판단함

▷ **<u>제공하는 자</u> : 개인정보, <u>제공 받는 자</u> : 개인정보가 아닐 수도 있음**

▶ **(이슈2) 위 규정에 따르면, 개인정보보호법의 적용을 받지 않는 가?**

=> 개인정보보호법의 적용대상이 아니라는 규정이 없으므로 적용 여부도 개인정보 여부에 따라 달라질 것으로 보임

▷ **<u>제공하는 자</u> : 개인정보(적용 대상임), <u>제공 받는 자</u> : 개인정보 아닐 경우만(적용대상 아님)**

PIPC

Ⅲ-3. 개인정보 보호 관련 법령에서의 이슈(2/2)

▷ **개인정보보호법 제58조 제1항 제1호**

① 다음 각 호의 어느 하나에 해당하는 개인정보에 관하여는 제3장부터 제7장까지를 제외한다.

1. 공공기관이 처리하는 개인정보 중, 『통계법』에 따라 수집되는 개인정보

▶ **(이슈1) 공공기관이 작성하는 모든 통계에 대하여 위 규정의 적용을 받는가?**

=> 모든 통계에 적용되지 않으며, 『통계법』 에 따라 작성하는 통계의 경우에 한함

▶ **(이슈2) 위 규정은 개인정보를 제공하는 자 입장인가, 아님 제공 받는 자 입장인가?**

=> 제공 받는 자에게 적용되는 규정이며, 제공하는 자는 통계법에 근거가 있음

※『통계법』 제24조, 제24의2, 제25조는 개인정보보호법 제18조 제2항 제2호 **'다른 법률에 특별한 규정이 있는 경우'**에 해당하는 것으로 보임

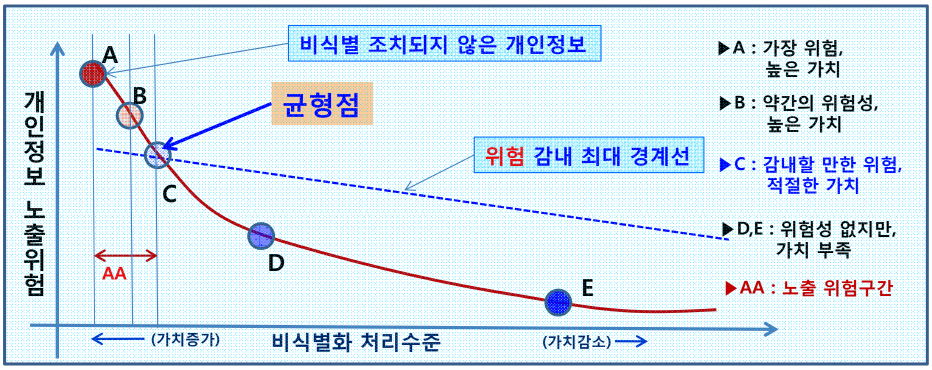
➔ 이상과 같은 규정의 적용을 받지 못할 경우에는 **정보주체의 동의** 또는 **"비식별 조치**하여 빅데이터와 통계작성을 위해 개인정보의 제공 및 활용이 가능

PIPC

Ⅲ-4. 개인정보의 비식별 조치 관련 이슈

❑ **개인정보에 대해 비식별 조치를 강하게 하면 할 수록 개인정보 노출 위험은 낮아지나, 정보의 가치도 동시에 낮아지므로**

=> 적절한 수준, 즉 개인식별이 어렵고, 데이터의 가치도 적절한 수준의 비식별 조치방법이 필요(많은 연구필요)



[출처] '통계자료의 비밀보호를 위한 익명화 방법들', 통계연구 p153, 통계청, 2004.10.

PIPC

Ⅲ-5. 빅데이터 분석과 통계(분석)의 비교

▶**분석 방법적 관점 : 빅데이터 분석 기법이 다양하고 범위가 넓음**

- 빅데이터 분석을 위한 기법에 통계학의 다양한 분석방법을 활용

※ 예 : 회귀분석, 시계열 분석/모델링, ANOVA분석 등

▶**분석결과의 활용목적 : 유사하나 빅데이터 분석결과의 응용분야가 넓음**

- (유사) 미래예측을 통한 의사결정, 정책기획 등을 위해 활용
- (차이) 지능형 장비 개발, 맞춤형 서비스 등은 빅데이터 응용 분야

▶**분석 결과에 대한 신뢰수준에 대한 기준(이슈)**

- 빅데이터 : 분석자 주관으로 결정(활용 목적에 따라 임의 설정, 예 30%, 20%도 가능)
- 통계(분석) : 이론적 기준(95%±오차 , 99% ± 오차 등)

➔ 빅데이터 분석결과는 **통계(분석)와 같은 신뢰수준 관련 기준**이 없으므로 결과의 신뢰수준 측면에서 차이가 있음

➔ 따라서, **빅데이터 분석결과의 공개·공동 활용하는 것에 대한 연구가 필요**

PIPC

Ⅳ. 결 어

Ⅳ. 결 어

PIPC

Ⅳ. 결 어

❑ 지금까지 빅데이터, 통계, 그리고 이들과 개인정보보호와의 관계 논의

❑ 빅데이터 분석과 통계작성 및 분석은 중간과정에서는 차이가 없으나, 결과의 활용 측면에서는 목적에 따라 차이가 있음

▶(빅데이터 분석) 맞춤형 서비스, 지능형 장비개발 등의 목적에서 이용자 주관적 판단 활용을 할 수 있음

▶(통계작성 및 분석) : 공표, 공개 ⇒ 공동활용(적절한 수준의 신뢰수준을 요구)

❑ 『통계법』에 따른 통계의 경우를 제외하고는 『개인정보보호법』 등 관련 법령을 준수하여야 함

▶ 개인정보를 동의 받은 목적 외의 목적 또는 법령에 근거가 있는 경우를 제외하고는 빅데이터나 통계작성 목적으로 활용이 불가함

▶그러나, "개인정보 비식별 조치 가이드라인" 에 따라 개인정보를 비식별 조치하여 이용할 경우는 그러하지 아니함

PIPC

감사합니다.

PIPC

국가통계발전포럼
'16. 8. 31. 13:30-13:50

빅데이터의 국가통계 활용 전략

2016. 8. 31.

정보통신정책연구원(KISDI)
미디어통계분석그룹장 정용찬

본 발표자료는 통계청의 통계연구 특별호(2016. 6)에 수록된 '빅데이터의 국가통계 활용 전략'의 주요 내용을 수정, 보완하여 작성함

차 례

2

데이터?

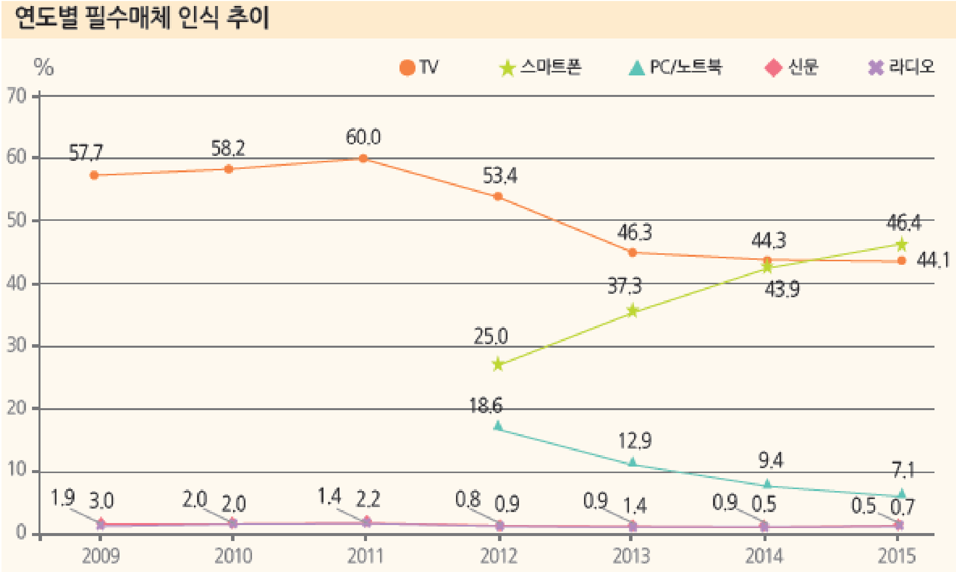
3

데이터!

신문
잡지
책
TV
라디오
MP3
게임기
…
…

4

스마트폰, TV 제치고 필수 매체로 부상



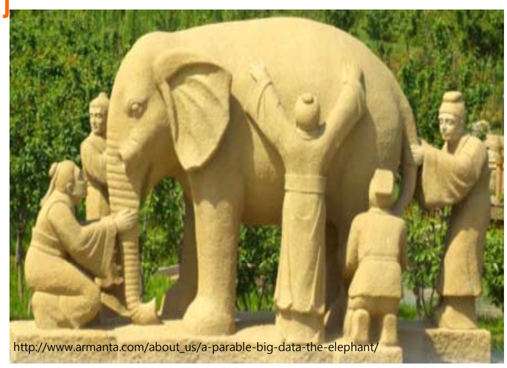
출처: 방송통신위원회, 방송매체이용행태조사, 2009~2015.

빅데이터(Big Data)

- 데이터의 양이 거대하고 형태도 다양해서 기존 방법으로 수집, 저장, 분석이 어려운 데이터

Volume(대규모)

Velocity(실시간 생성)



Value?

Variety(숫자, 문자, 영상)

"세상은 IT시대에서 DT시대로 이동 중. 데이터로 얼마나 많은 가치를 창출해 내느냐가 중요"

- 마윈 알리바바 회장, 2015 빅데이터산업 설명회

국가통계 조사 환경

● 데이터 기반 사회의 도래

- 정부 정책 의사결정에 있어서 데이터의 중요성 증대
- 데이터 기반(data-driven) 환경에서 **통계**의 **중요성**은 더욱 **강조**

● 조사 환경 악화: 조사 거부, 불응 등 응답률 저하 심각

- 1인 가구, 초핵가족 증가 등 가구원 수 감소와 재택률 감소로 응답률 저하
- 프라이버시 보호 강화, 유선전화 가입자 감소 등 통계 조사의 신뢰성 위협

※ 통계청은 광업·제조업 조사 거부 4개사에 2012년 말 과태료 부과

출처: 정용찬 외(2014) '빅데이터산업 촉진전략연구. 정용찬 외(2016) '조사환경 변화에 대응한 통계 생산체계 혁신 방안' 연구 계획서'.

조사	조사/공표 주기	조사대상	조사방법	불응률 (부재율 제외)
통계청 인구주택 총조사	5년	표본조사지역 내에 상주하는 모든 내·외국인	인터넷조사/ 면접조사	2010: 0.23% 2015: 0.88%
통계청 가계동향조사	매월 조사 분기 공표	전국에 거주하는 일반가구 약 9,000가구	면접조사	2007: 17% 2014: 22.5%
통계청 경제활동인구조사	매월 조사/공표	전국 3만 3천 가구 내 개인	면접조사	2011: 5.9% 2014: 8.2%

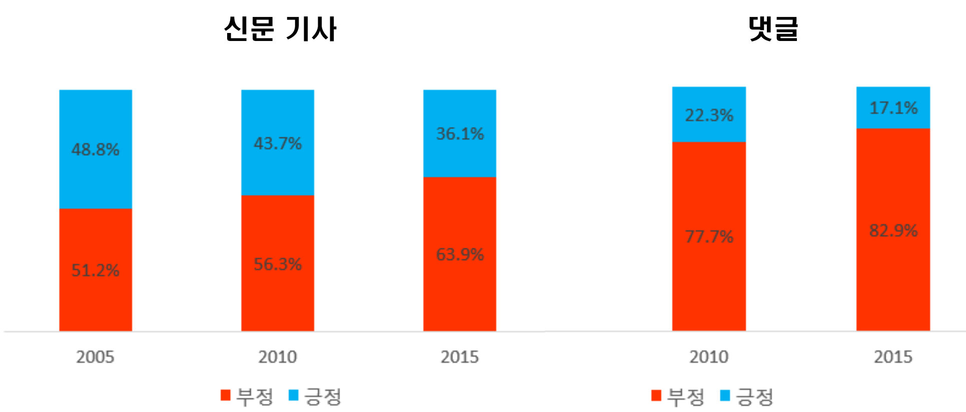
출처: 각 조사별 홈페이지(census, KOSIS), 언론사 보도 자료

7

통계 인식 변화 분석: 텍스트 마이닝

● 신문기사 내용과 댓글 분석

- '통계'와 '통계 조사'에 대한 인식의 변화 추이 조사
- 기사는 부정 감성이 7.6%p, 댓글은 5.2%p 증가(2010년 대비 2015년)



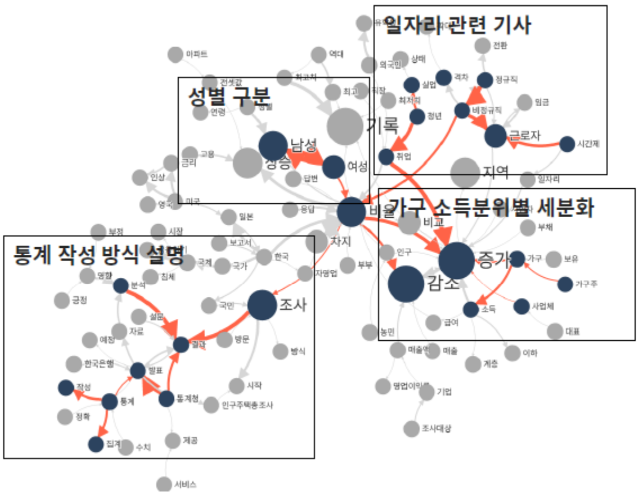
출처: 정용찬 외(2016) '조사환경 변화에 대응한 통계 생산체계 혁신 방안 연구'(진행 중).
연합신문의 2005, 2010, 2015년 각 1년간 통계, 통계조사 관련 기사와 댓글 분석

8

통계 인식 변화 분석: 텍스트 마이닝

● 신문 기사

- 기사 전문성 강화: 명확하고 세분화된 정보 전달을 위한 노력 관찰



출처: 정용찬 외(2016) '조사환경 변화에 대응한 통계 생산체계 혁신 방안 연구'(진행 중)

통계 인식 변화 분석: 텍스트 마이닝

● 댓글

- 경기 침체, 젠더 차이 고려 부족 등으로 인한 부정 감정 증가
- 경제, 사회 문화 지표인 '통계'를 통한 현실 비판

연합뉴스 '5년만의 인구총조사'…도심에서 땅끝마을까지 발품

지난 1일부터 **'2015 인구주택총조사' 방문조사**가 전국에서 일제히 시작됐다.
집집이 찾아다니며 조사하던 **과거 인구주택총조사 방식은 올해 바뀌었다.** 전 국민의 20%에 해당하는 표본 가구만 조사하는 '등록센서스' 방식을 처음 적용한 것이다.
방문·면접 대상은 지난달 24~31일 시행된 인터넷 조사에 응하지 않은 가구다. 조사 시간은 약 30분이다.
…
조사 문항은 출생, 가족 현황, 주택, 재산 등 다양하고 세밀하다. 다소 예민한 문항에는 주민들이 짜증을 내거나 조사를 거부하기도 한다.
…

질문의도가 불순하잖여. 질문이 질문같아야 응답을 하지. 왜 **개인사생활**을 터나? **인구조사**라면 이름이나 연락처 물어보지말고 가구원수나 나이 정도만 파악해야지..왜 세밀한 **개인 신상**을 터나?? 그런 **개인정보수집**해서 어따 쓸려고??

이번 통계설문은 **사생활털기** 수준의 질문위주라 응답자의 거부감이 커서 **결과도 엉터리**에 조사자체가 실패할게 자명하고.. 다음에 통계조사할 때는 **사생활침해수준**의 곤란한 질문은 하지말고 **인구조사에 꼭 필요한 핵심질문** 모아서 개괄적인 질문으로 조사해라. 응답자가 **불쾌감**이 없고 **개인신상 노출에 대한 불안감**이 없어야 정확한 통계도 나오는 것임.

연합뉴스 '결혼한 남성 10명 중 3.7명이 간통 경험'

정부 출연 연구기관인 한국여성정책연구원이 지난해 6월 성인 **남녀 2천명을 대상으로 조사**해 최근 발간한 **보고서 '간통죄에 대한 심층분석'**에 따르면 **남성** 응답자 가운데 **결혼 후 간통 경험이 있다고 답한 비율이 36.9%**를 기록했다.

또 전체 **여성** 응답자 중 결혼 후 간통 경험이 있다는 답변은 **6.5%**로 조사됐다.
…
이 설문조사는 지난해 6월 기준 **주민등록 인구 통계에 따라 지역별, 성별, 연령별 인구비례로 할당해 무작위 추출**한 패널을 대상으로 이뤄졌다.

설문이란거부터 노답. 특히 성이나 탈선 등에 있어서 **설문은 믿을게 못 된다.** 대표적인 예로 **흡연율 조사**가 있는데 조사기관에서 20 30대 남성에게 담배피냐고 했을 때 나온 흡연율과 20 30대 여성에게 담배피냐고했을 때 나온 흡연율과 거짓말 못 하는 **소변검사 결과 비교**해보니 여성은 2배가 더 높았고 남성은 10퍼정도 더 높았다. **여성**은 설문이여도 이런 류 질문은 남 시선 의식해서 **더 숨기게 될 수 밖에 없다.**

무기명 통계 조사에서 남자들은 그래도 어느 정도 사실대로 답변을 한다. 그런데 여자들은 무기명 인데도 불구하고 거짓을 답한다. 예컨대 **여성 흡연율 조사**할때 이런 사실이 단박에 드러났지. …

연합뉴스 '집안일 공평하게 분담' 부부의 비율, 韓이 북유럽의 절반수준

한국 남편들이 아내와 공평하게 집안일을 분담하는 비율이 북유럽 국가들의 절반 수준이라는 조사 결과가 나왔다.
통계청의 '한국의 사회동향 2014' 보고서 집필자인 한경혜 서울대 아동가족학과 교수와 홍승아 한국여성정책연구원 센터장은 **12개국의 만 20세 이상 기혼 남녀를 대상으로 조사한 결과**, 이렇게 나타났다고 7일 밝혔다.

12개 국가는 **한국, 일본, 영국, 독일, 프랑스, 스웨덴, 노르웨이, 핀란드, 덴마크, 멕시코, 필리핀, 대만이다.**

조사 대상 항목은 식사 준비, 세탁, 집안 청소, 장보기, 아픈 가족 돌보기, 소소한 집안 수리 등 6개다.

이 조사는 **부부의 맞벌이 유무와 근무시간을 고려해 생각할 필요**가 있다. 남편은 아침8시부터 밤 10시까지 일한다. 주부인나는 하루 평균 6~7시간 청소, 장보기, 음식준비,세탁, 정리, 아이들 학습지도 한다. 열심히 사는 편이라 나도 힘들지만 남편에 비하랴 전업주부라면 하루종일 일하는 남편에게 가사분담 요구해선 안된다. … 자발적으로 서로 도우면 좋지만 강요해선 안된다.

우리신랑 대기업사무직인데도 아침6시반에 출근해서 매일야근에 밤10시반에 집에와서 애얼굴도못보고사는데 **무슨 집안일을 반반**하란말이냐..**현실성있는 통계 좀 내라..** 내가 아줌만데도 어이없네..**남편들 사회구조가 안 그렇다고요**!!

출처: 정용찬 외(2016) '조사환경 변화에 대응한 통계 생산체계 혁신 방안 연구'(진행 중)

스포츠: STATCAST

● MLB STATCAST

- "과거에는 측정할 수 없었던" 야구 데이터를 수집하는 최신 기술
- Trackman(ball tracking) & ChyronHego(player tracking)

Source: https://swishanalytics.com/blog/2015-04-20/MLB-statcast-debuts-tomorrow

● 측정 데이터: 경기당 7.3 tera

- 투구: Pitch velocity, release point, perceived, distance from release
- 타격: Bat velocity, launch angle, ball direction, hangtime, travel distance
- 주루: First step times, acceleration, top speed, trajectory
- 수비: Distance covered, efficiency, transferring, velocity of throws

Source: https://swishanalytics.com/blog/2015-04-20/MLB-statcast-debuts-tomorrow

Source: An Introduction to TrackMan Baseball SABR40
http://m.mlb.com/news/article/128661588/statcast-yankees-pitcher-nathan-eovaldis-fascinating-low-spin-splitter

11

GE: Industrial Internet

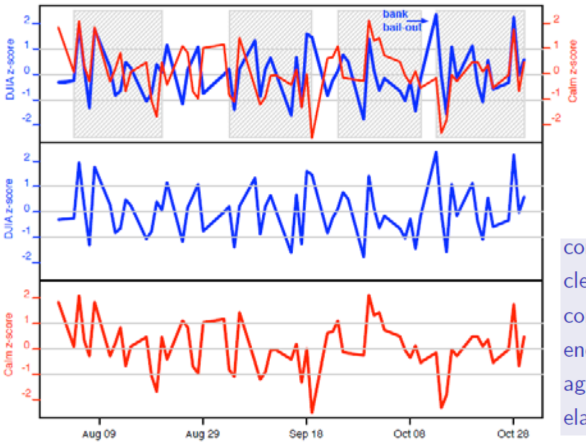
Source: General Electrics. http://www.enterrasolutions.com/2013/09/fellow-travelers-big-data-and-the-internet-of-things.html

12

주가 예측

● 트위터의 집단적 정서와 다우존스지수

- 다우존스산업평균지수 예측(상승, 하락) 정확도 87.6%
- Google-Profile of Mood States (GPOMS): Calm, Alert, Sure, Vital, Kind, Happy



Socionomics

composed/anxious : calm
clearheaded/confused : alert
confident/unsure: sure
energetic/tired: vital
agreeable/hostile: kind
elated/depressed: happy

Bollen, Johan, Mao, Huina and Zeng Xiao-Jun. Tweeter mood predicts the stock market, Journal of Computational Science, Vol 2, Issue 1, March 2011, pp 1-8.

인공지능

● IBM 수퍼컴퓨터 왓슨(Watson)

- 미국의 인기 퀴즈쇼 <제퍼디(Jeopardy!)> 출연, 인간 챔피언과 겨뤄 승리('11. 2.)
- 4테라바이트의 디스크 공간에 저장된 2억 페이지 콘텐츠 활용

http://www-07.ibm.com/innovation/kr/think/topic.html

- 의료보험 데이터 분석과 종양 진단에 활용('13, Westmed Medical Group과 폐암치료 업무협약)
- 씨티그룹(Citi group)과 금융 분야 활용 계약 체결

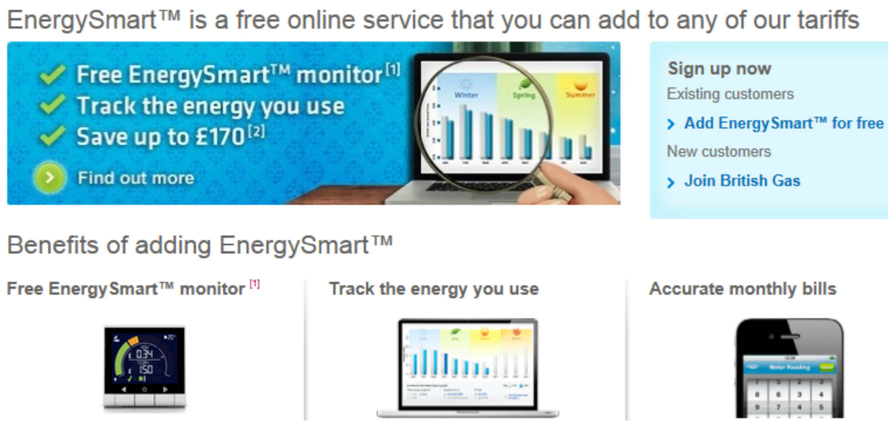
※ 퀴즈쇼 <제퍼디> : 질문 분야가 광범위하고 은유적인 표현이 포함되어 의미를 파악하기 어려움

● 구글의 알파고

에너지

● 스마트계량기 British Gas

- 과거: 한 해에 두 번 검침, 3개월 단위 요금 청구
- 토요일 전기 무료(미/영, 스마트계량기 설치 조건)), 주중 요금 인상
- 전력소비 분산, 저소득층 도움, 온실가스 감축



http://www.britishgas.co.uk/products-and-services/gas-and-electricity/energysmart.html

범죄 수사

● FBI CODIS: 등록된 범죄자 DNA로 용의자를 색출

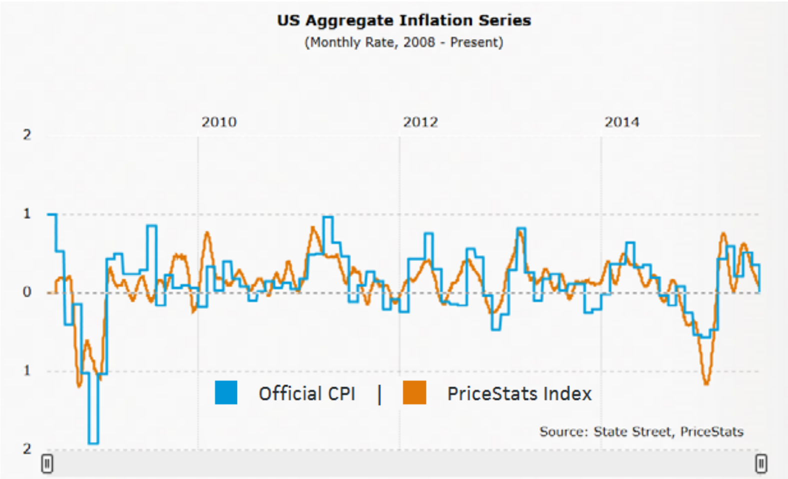
- 물잔, 껌, 봉투, 혈흔, 권총 방아쇠, 모자
- DNA Identification Act 의회 통과(1994)
- National DNA Index System(NDIS) 구축 (1998)
  : 현재 50개 주 가입, 천만 건 DNA 정보 보유

http://www.kyobobook.co.kr/product/detailViewKor.laf?ejkGb=KOR&mallGb=KOR&barcode=9788926390559&orderClick=LEB&Kc=

물가지수

● MIT의 The Billion Prices Project(BPP)

- 70개 국가의 온라인 소매업 데이터 수집, 물가 지수 매일 산출
- "정부가 독점하던 물가지수를 누구나 구할 수 있게 되었다"(NYTimes, 2013)

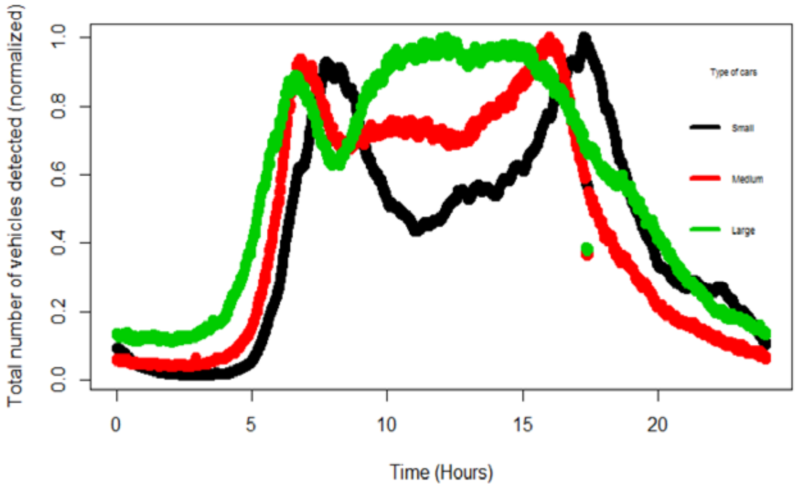


출처: http://www.pricestats.com/inflation-series?chart=1837

교통 통계

● 네덜란드 Road sensor (traffic loop) data

- 매일 2.3억건 데이터 수집, 차종별 통행속도, 통행량 분석

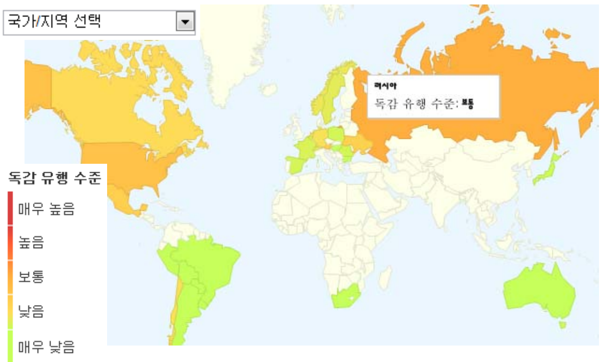


Source: Peter Struijs(2014), Traffic loop data for transport statistics, http://bpp.mit.edu/

보건 통계

● 구글 독감 동향 서비스(google.org/flutrends)

- 단어 검색어 빈도로 독감 환자 수, 유행지역을 예측
- 미 질병통제본부(CDC)보다 예측 속도 뛰어남



정부 데이터 기반 서비스

2. 활용 사례

● 온라인 부동산 장터 서비스

- 벤처기업 Zillow는 1억 천만 건이 넘는 미국 주택 데이터베이스를 기반으로 온라인 부동산 장터 서비스를 개설(시가총액 30억 달러)
- 주택 소유자, 구매자, 판매자, 세입자, 부동산 중개사, 부동산 대출 기관 등과 주택, 부동산, 담보, 주택 개조 정보 공유
- 영국 토지 등기소(UK Land Registry)의 주택 거래 데이터에 기반한 Zoopla

※ 매출(7천6백만 파운드)의 86%는 19,000여 중개사가 지불하는 정보이용료

● 농업 정보 제공 서비스

- Climate Corp는 미국 정부가 보유하고 있는 60년간 작물재배 현황 데이터와 기상청의 날씨 정보, 토양 데이터(무료 제공)를 활용하여 농부들에게 수익 개선을 위한 경작/재무 정보 제공, 날씨보험 판매
- 구글 출신 엔지니어 2명이 2006년 창업한 이 회사를 몬산토는 2013년 9억 3천만 달러에 인수

Source: The World Bank(2014), "Open Data for Economic Growth"

빅데이터의 국가통계 활용 사례(요약)

구분	데이터 유형	주요 내용
교통 통계	센서 데이터(Sensor data)	네덜란드는 도로 상에 설치한 2만여 개의 차량감지기를 통해 차종별 교통량, 속도 등의 정보 수집 경기 변동과 관련 있는 화물차량의 물동량 정보 분석, 지역별 통행량지수(traffic index) 산출
물가 통계	SNS, 블로그	네덜란드 통계청은 소셜미디어메시지에 대한 감성(sentiment)분석 감성분석 결과 경기와 상관관계가 높은 것으로 나타나 주간소비자신뢰지수 산출에 활용 가능 UN Global Pulse의 트위터와 물가 관계 분석
물가 통계	온라인 데이터	Eurostat는 인터넷 상의 물가 정보를 자동 수집, 소비자물가지수(CPI) 산출에 참고 MIT의 Billion Prices Project **중국 통계청**은 2013년 11월 알리바바, 바이두 등 11개 전자상거래업체와 빅데이터 물가 지수 개발 위한 파트너쉽 협약 체결
관광 통계	GPS 데이터	관광통계 산출에 휴대폰의 위치정보를 활용

Source: UNECE(2013), What does "Big Data" mean for official statistics?
Peter Struijs(2014), Traffic loop data for transport statistics, http://bpp.mit.edu/

21

빅데이터의 국가통계 활용 사례: Eurostat Q2016

구분	데이터 유형	주요 내용
물가 통계	온라인 데이터 (U.S. Census Bereau)	표본조사 데이터와 온라인 거래 데이터의 연계 자동차부품, 보석/시계 POS데이터 분석 소지역 추정량 산출 모형 개발
인구 통계	모바일폰 데이터 (Eurostat, 벨기에통계청)	모바일폰 데이터와 상주인구통계 비교분석 인구, 관광, 교통, 이사 통계에 활용 가능
실업 통계	온라인 데이터 독일통계청	구인구직사이트((job portal)의 데이터를 활용, 일자리 추정
관광 통계	교통량/모바일폰/신용카드 데이터 스페인통계청	빅데이터와 조사 데이터간의 비교 분석 (국적, 관광 일수, 지출 규모 등)

Source: Eurostat, Q2016 European Conference on Quality in Official Statistics, 31st May – 3rd June.
http://www.q2016.es/scientific-information/conference-programme/

22

2015년 빅데이터 시범사업(미래부/NIA)

2015년도 제조, 에너지, 유통 등 산업중심 5개 서비스 실증 지원

출처 : 미래부 보도자료('16. 5. 31), '미래부-NIA, 2016년 빅데이터 시범사업 착수'

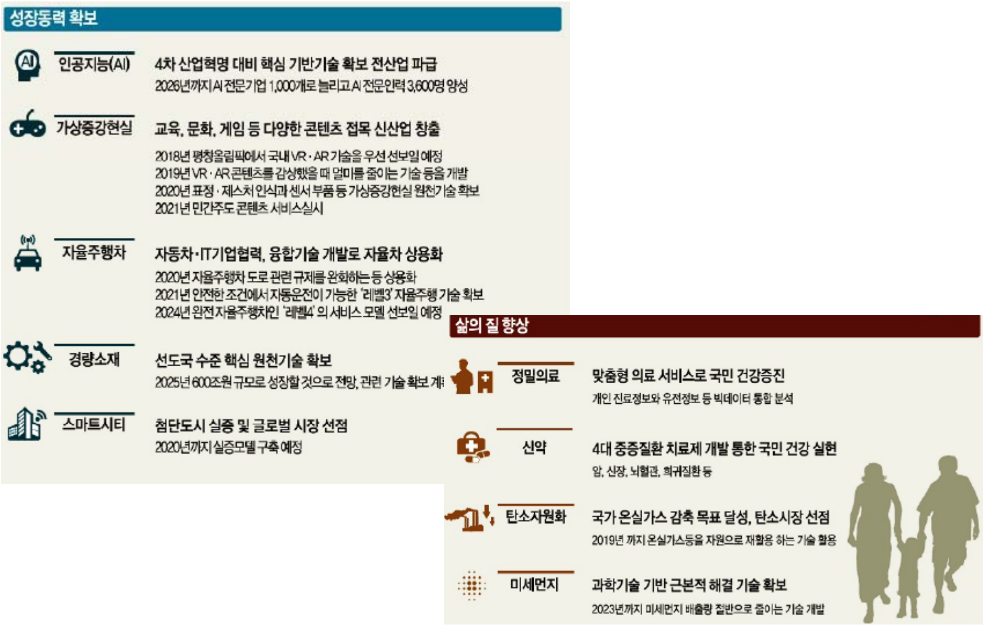
2016년 빅데이터 시범사업(미래부/NIA)

구분	주관기관/참여기관	과제명
선도 시범	KT/질병관리본부	로밍 빅데이터를 활용한 해외유입 감염병 차단 서비스
	더블유쇼핑/한동대	빅데이터 딥러닝 기술 활용 스마트 T-커머스 서비스 개발
	매일유업/ 한국그린비즈니스협회	유가공 업종 제조 생산·에너지 최적화를 위한 빅데이터 플랫폼 개발
	유라/충북대학교 등	딥러닝 기술 기반의 대용량 제조 데이터 분석 서비스 플랫폼 개발
산업 확산	ING생명/생명보험협회	생명보험 빅데이터 전략모델 개발 및 확산
	삼성중공업/현대중공업 등	제조업 빅데이터 전략모델 개발 및 실증

출처 : 미래부 보도자료('16. 5. 31), '미래부-NIA, 2016년 빅데이터 시범사업 착수'

9대 국가전략 프로젝트(2016. 8)

2. 활용 사례



출처 : 디지털타임즈(2016. 8. 11), '민관 새 협력모델 제시'

25

빅데이터와 국가통계의 특징

3. 대응 전략

구분	빅데이터	국가승인통계
관련성(relevance)	분야별 상이	높음
정확성(accuracy)	분야별 상이	높음
시의성(timeliness)	매우 높음	다소 낮음
생산 비용 (cost)	비교적 낮음	높음
응답부담(burden)	낮음	높음

출처: 정용찬(2011), '성과와 고객 관점의 통계품질진단 개선'. 국가통계포럼: 국가통계의 현안과 대응(2011. 5. 26) 발표 자료의 '품질 평가 기준' 일부 활용. 품질 평가 기준 원 출처는 Executive Office of the President of the United States(2007). "Strengthening Federal Statistics"

26

데이터 원천

- 기존 정부승인통계: 표본조사, 행정데이터(등록정보 등)

\+

- 행정데이터: 전자의료기록, 보험, 부동산, 조세…
- 상업용/거래 데이터: 신용카드, 온라인 상거래(모바일 기기)…
- 센서 데이터: 위성 사진, 도로, 교통, 기후…
- 추적(tracking) 데이터: 모바일 휴대전화, GPS…
- 행동데이터: 온라인 검색(상품, 서비스, 기타 정보)…
- 의견(opinion) 데이터: 블로그, 소셜미디어…

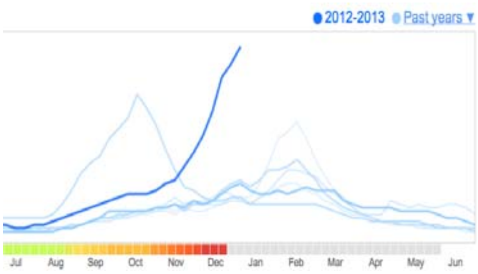
출처: UNECE(2013), What does "Big Data" mean for official statistics?.

빅데이터의 국가통계 활용

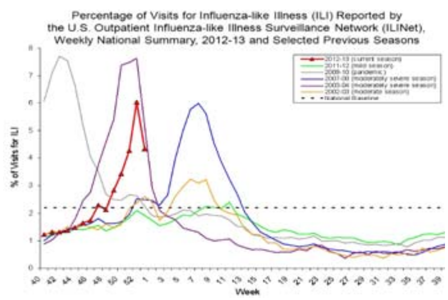
구분	내용
보완	빅데이터 통계를 국가통계의 보조 지표로 활용
대체	기존 국가승인통계를 빅데이터 통계로 대체
창조	새로운 빅데이터 기반 통계 개발

출처: UNECE(2013), What does "Big Data" mean for official statistics?.

- 구글 독감 동향



- 질병통제본부



출처: http://haicontroversies.blogspot.kr/2013/01/cdc-versus-google-trends.html

빅데이터의 국가통계 활용 시 고려 사항

- 법적 근거(Legislative): 데이터 접근과 이용 관련 법령
- 프라이버시(Privacy): 데이터 재이용 관련 사회적 공감대
- 재무적(Financial) 관점: 민간 데이터 획득 비용 대비 효용
- 관리(Management): 데이터 관리와 보호 정책
- 방법론(Methodological): 데이터 품질과 통계방법론의 적합성
- 기술(Technological): 정보기술(IT) 관련 이슈

Source: UNECE(2013), What does "Big Data" mean for official statistics?

주요국의 빅데이터 정책

구분	전략	추진조직	특징
영국	오픈데이터전략 ('12. 6, '14. 7) 데이터 역량 강화 전략 ('13. 10)	내각 사무국 공공데이터그룹(공급자) 데이터전략위(사용자)	오픈데이터를 통한 혁신, 성장 도모 데이터 역량 강화를 위한 교육 개혁
미국	빅데이터 연구개발 전략 ('12.3) 기회포착과 가치보호전략 ('14.5)	대통령실과학기술정책국 빅데이터 고위 운영그룹	연구개발을 통한 인력 양성 개인정보보호 종합대책 제시
한국	빅데이터 산업 발전전략 ('13.12) 공공데이터 개방 발전전략 ('14.9)	미래부, 안행부 공공데이터전략위원회	공공데이터 개방을 통한 경제 발전(일자리) 2017년 오픈데이터 5대 강국

출처: 정용찬 외(2014) '빅데이터산업 촉진전략연구'의 내용 수정

주요국의 정보보호 정책

구분	정책	특징
영국	데이터보호법('98) 빅데이터와 데이터 보호 보고서('14. 7)	이 법에 따라 익명화 규약 발표('12. 11) 빅데이터 활용 방안 제시. 개인정보보호 위한 익명화 방안 제시
미국	온라인 프라이버시 프레임워크('12. 2) 빅데이터: 기회 포착과 가치 보호('14. 5)	빅데이터의 활용 강조(Opt-out 허용) 프라이버시 보호 강화
일본	개인정보보호법(2003) 개정('15. 9)	익명 처리 데이터는 본인 동의 없이 제공 가능
한국	빅데이터 개인정보보호 가이드라인 ('14. 12) 개인정보비식별조치 가이드라인('16.6) 신용정보의 이용 및 보호에 관한 법률 시행령 개정안 입법예고('16. 7)	데이터의 수집 단계에서 '비식별화' 조치가 선행된 경우에는 개인정보의 수집·활용 가능

출처: 정용찬 외(2015) '통신산업의 빅데이터생태계 촉진전략'의 내용 수정

31

데이터전략

● 국가 차원의 데이터전략 수립

- 정보화 시대 컴퓨터 교육, 인터넷 교육처럼 초중등 교육 과정의 개선(컴퓨팅과 데이터 분석 능력), 전문교사 양성 등 교육제도 개편, 인재 육성 정책
- 정부 역할 감안 시 공무원의 분석 능력 강화 정책 마련에 최우선

● 공공데이터 수요자 조직 상설 운영

- 영국: 공급자인 공공데이터그룹(Public Data Group)과 수요자인 데이터전략위원회(Data Strategy Board)와 오픈데이터이용자그룹(Open Data User Group) 운영

구분	주요 기관(주요 데이터)
공급자 그룹	통계청(인구센서스, 서비스업통계조사) 행정자치부(주민등록통계) 국토교통부(지가변동, 부동산거래, 건축물, 주택통계, 도로교통량) 국세청(국세), 기상청(날씨정보) 등
수요자 그룹	정부 부처, 지방자치단체, 공공기관, 국공영기업, 민간기업, 대학, 연구소

출처: 정용찬 외(2014) '빅데이터산업 촉진전략연구'.

32

데이터 공유와 개방

● 데이터 개방 확대

- 국세, 교통, 부동산, 기상, 보건/의료 등 활용도 높은 분야의 정부 데이터를 개인 식별 정보 삭제를 전제로 과감한 개방
- 정부 개방 데이터의 품질 확보를 위한 제도 도입, 연계 활용 위한 표준화

● 정부, 민간 데이터 공동 활용 방안 모색

- 정부와 기업이 생산하는 데이터의 재사용 방안에 대한 개인, 기업, 정부 등 이해당사자 간의 소통과 합의 필요
- 단계적인 공유 전략 수립

● 소셜데이터 공동 활용을 위한 인프라 구축

- SNS, 블로그에서 생성되는 소셜데이터의 수집, 공유시스템 구축

※ 트위터 (공개)정보의 미 의회도서관 Archive 협약(2010. 4)

데이터 유통 활성화

● 개인정보 보호의 범위와 수준에 대한 사회적 합의

- 주민등록번호, 휴대전화 번호 등에 비해 민감성이 떨어지는 주소, 이메일 등의 활용 방안 모색 필요
- 빅데이터 시대에 부합하는 개인정보 개념, 개인 정보의 유통과 활용, 개인 정보 보호의 범위와 수준에 대한 사회적 논의 필요

※ 개인정보 공개 수준은 국가별로 상이. 부패인식지수 조사에서 청렴도 최상위 국가군에 속하는 핀란드의 개인 소득과 납세는 공개 정보

출처: 정용찬(2015), 빅데이터 산업과 데이터브로커, KISDI Premium Report 15-04, KISDI.

● 개인정보 비식별조치 가이드라인

- 국무조정실 등 관계부처는 현행 개인정보보호 법령 내에서 빅데이터의 활용 위해 개인정보의 비식별조치 기준과 비식별 정보의 활용 범위 제시('16. 6. 30)
- 금융위원회는 개인을 특정할 수 없는 익명화한 개인신용정보를 통계자료로 활용할 수 있도록 하는 신용정보의 이용 및 보호에 관한 법률(신용정보법) 시행령 및 신용정보업 감독규정 개정안 입법예고('16. 7. 7)

개인정보 규제 수준(아시아 국가)

자료: 나성현(2015), IoT환경에서의 개인정보보호 이슈, KISDI Premium Report. 원출처는 Hogan Lovells(2014). "Data Privacy Regulation Comes of Age in Asia." 데이터 운영 시 요구사항, 직접적인 마케팅 규제, 개인정보 국외 이전에 대한 통제, 집행 환경의 강도 등 네가지 요소를 종합 평가. 한국은 4부문 모두 1위.

개인정보보호 위반사업자 과징금 제재

구분	사업자명(앱명)	위반 유형	위반조문	과징금	과태료	시정명령
배달	㈜**우아한 형제들**(배달의 민족)	기술적·관리적 보호조치 위반	§28①	-	**1,500**	○
부동산	㈜**직방**(직방)	기술적·관리적 보호조치 위반	§28①	-	**1,000**	○
	㈜**스테이션3**(다방)	기술적·관리적 보호조치 위반	§28①	-	**1,000**	○
		개인정보 유효기간 위반	§29②	-	**1,000**	○
포인트	**CJ CGV(주)**(CJ ONE)	기술적·관리적 보호조치 위반	§28①	-	**1,000**	○
쇼핑	㈜**씨제이오쇼핑**(CJmall)	개인정보 유효기간 위반	§29②	-	**1,000**	○
	㈜**지에스홈쇼핑**(GS SHOP)	기술적·관리적 보호조치 위반	§28①	-	**1,000**	○
		개인정보 유효기간 위반	§29②	-	**1,000**	○
	㈜**현대홈쇼핑**(현대Hmall)	기술적·관리적 보호조치 위반	§28①	-	**1,000**	○
		개인정보 유효기간 위반	§29②	-	**1,000**	○
	㈜**우리홈쇼핑**(롯데홈쇼핑)	기술적·관리적 보호조치 위반	§28①	-	**1,000**	○
		개인정보 제3자 제공 위반	§24조의2①	**18,000**	**-**	○
		개인정보 유효기간 위반	§29②	-	**1,000**	○
	㈜**엔에스쇼핑**(NS홈쇼핑)	기술적·관리적 보호조치 위반	§28①	-	**1,000**	○
		개인정보 유효기간 위반	§29②	-	**500**	○
	㈜**홈앤쇼핑**(홈&쇼핑)	기술적·관리적 보호조치 위반	§28①	-	**1,000**	○
		개인정보 유효기간 위반	§29②	-	**1,000**	○
	㈜**공영홈쇼핑**(아임쇼핑)	기술적·관리적 보호조치 위반	§28①	-	**1,000**	○

출처: 방송통신위원회 보도자료(2016. 8. 11), 개인정보보호 법규 위반사업자 과징금·과태료 제재. 위반조문은 정보통신망법. 롯데홈쇼핑은 2009년 2월부터 2014년 4월 기간 고객 개인 정보를 롯데·한화·동부 등 3개 손해보험사에 몰래 판매, 37억3600만원 수익 올림 (연합뉴스, 2016. 8. 11, '롯데홈쇼핑, 개인정보 팔아 37억 벌었는데 과징금 1.8억')

데이터 역량 강화와 기술 개발

● 데이터 역량 강화 전략 수립

- DT(data Technology) 시대에 부합하는 교육과정 개편
- 정부의 역할을 감안, 공무원의 데이터 역량 강화 정책/실행 계획 수립

● 기술 개발을 통한 전문가 양성

- 데이터 보안, 암호화, 비식별화 등 개인정보보호를 위한 기술 개발이 중요

※ 미국은 프라이버시 보호 연구개발에 최대 우선순위를 두고 국립과학재단, 국립보건원, 국토안보부, 국가안보국 등이 프라이버시 연구에 매년 7천7백만 달러의 예산을 집행(White House, 2014)

- 익명화 처리 등 '개인정보보호 기술' 개발과 보급으로 데이터 제공자의 부담 경감과 유통 활성화 기반 조성 필요

※ 정부는 한국인터넷진흥원(KISA)에 '개인정보 비식별 지원센터'를 설치·운영할 예정이며, 소관부처 책임 하에 한국신용정보원, 금융보안원, 사회보장정보원, 한국정보화진흥원 등 분야별 전문기관을 지정, 운영 계획

출처: 정용찬 외(2014) '빅데이터산업 촉진전략연구'. 관계부처 합동(2016. 6.30). 개인정보 비식별 조치 가이드라인.

37

데이터 중심 문화 정착

● 데이터 기반 의사결정 체계 정착

- '데이터에 기반한 정책 수립' 정착 위한 의사결정권자의 지원 필요
- 조직 구성원의 학습과 공감대 형성 중요

● 기록관 vs 통계전문가

구분	주요 기관(주요 데이터)
공공기록물 관리에 관한 법률	제13조(기록관) ① 공공기관의 기록물을 효율적으로 관리하기 위하여 대통령령으로 정하는 공공기관은 기록관을 설치·운영하여야 한다 다음 각 호의 어느 하나에 해당하는 공공기관은 법 제13조제1항에 따라 기록관을 설치·운영하여야 한다. 1. 중앙행정기관 5. 시·도 6. 시·군·구 7. 시·도 교육청 8. 군 기관
통계법	제7조 (통계작성기관의 인력 및 예산 확보) ①통계작성기관의 장은 통계의 작성 및 보급에 필요한 인력과 예산을 확보하도록 노력하여야 한다.

38

빅데이터 시대 정부의 역할

● 정부는 데이터 공급자이자 선도자, 촉매자, 활용자

구분	내용
공급자 (supplier)	경제성장과 기업혁신에 필요한 데이터 정기적 공개 데이터 품질과 접근성 개선을 위한 지속적인 노력
선도자(leader)	정부/지방정부, 공공기관, 국공영기업, 민간기업 데이터도 공개
촉매자 (catalyst)	데이터 이용자, 애플리케이션 개발자, 데이터 기반 산업 등 데이터 생태계 활성화 위한 오픈데이터 활용 촉진
활용자 (user)	공공기관의 공공데이터 활용을 촉진과 관련 기술에 대한 투자 의사결정 개선, 신규 서비스 창출 지원 고급 분석 서비스 제공 오픈데이터를 활용한 민간의 상품과 서비스 개발 촉진

출처: 정용찬 외(2014) '빅데이터산업 촉진전략연구'. 원 출처는 Source: The World Bank(2014). Open Data for Economic Growth.

● 데이터 활용 촉진을 위한 정부의 역할(호주)

- 데이터 소유권 프레임웍 수립: 소유, 이전, 대가 산정 법/제도 마련
- 데이터 보안과 프라이버시 보호
- 정부 행정 업무 개선을 위한 데이터 활용 강화

출처: Digital Disruption: What do governments need to do? Productivity Commission Research Paper, Commonwealth of Australia, June 2016

39

빅데이터 시대의 국가통계

● UN 국가통계(Official Statistics) 10대 원칙(Fundamental Principles)

- 제5원칙. 통계 작성 목적의 데이터는 다양한 유형의 원천 활용 가능
  통계작성기관은 품질, 적시성, 비용, 응답자 부담 고려해야

- Principle 5. Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents.

Source: Fundamental Principles of Official Statistics, http://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx

"빅데이터 환경에 부합하는 데이터 수집, 분석, 전파 등 국가통계 생산 체계 재설계 필요"

40

"경청해 주셔서 감사합니다"

- 정용찬, KISDI ICT통계정보연구실 미디어통계분석그룹장. ycjung@kisdi.re.kr

- 저서 : 빅데이터(2013), 시청률조사(2013), 미디어산업통계(2013), 미디어소비통계(2013), 융합과 통섭(2012, 공저), 미디어통계의 이해(2008), 통계로 본 대한민국(2007)…
- 연구보고서: 빅데이터 산업과 데이터브로커(2015, KISDI Premium Report), 통신산업의 빅데이터생태계 촉진 전략(2015), ICT통계관리체계 개선방안 연구(2015) 빅데이터산업촉진전략연구(2014)…

41

제6회 국가통계발전포럼

빅데이터 활용 통계생산 방법 및 품질검증

2016. 08. 31.

박민규, 김강민
(고려대학교)

1

순 서

1. 빅데이터의 개념 / 특징
2. 공공데이터의 개념 / 종류
3. 빅데이터를 활용한 통계 생산
4. 빅데이터 통계의 품질 검증

Part 1 빅데이터의 개념 / 특징

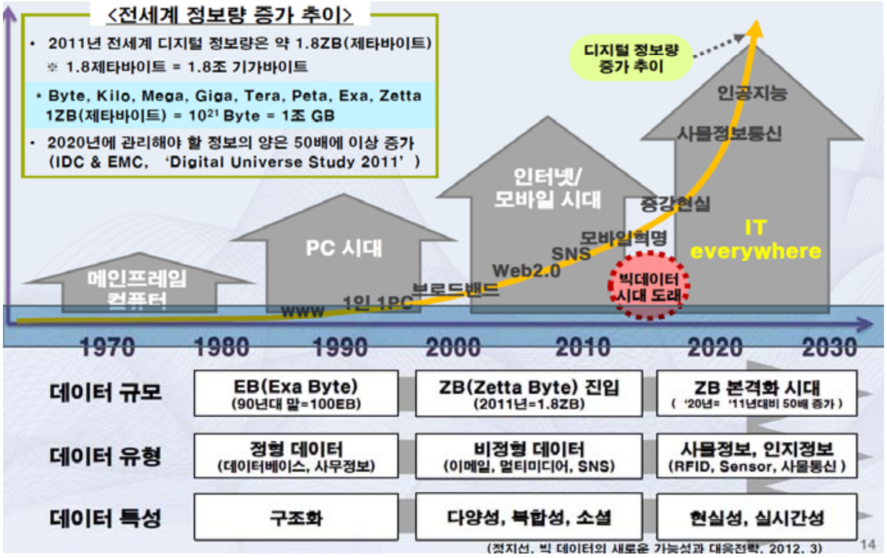
3

빅데이터 개념 / 종류

- 협의적 개념 : 기존의 관리 및 분석 체계로는 감당할 수 없을 정도의 **거대한 데이터의 집합**
- 광의적 개념
  - 기존 데이터베이스 관리 도구의 데이터 수집 · 저장 · 관리 · 분석 역량을 넘어서는 대량의 **정형 또는 비정형 데이터 세트** 및 이러한 데이터로부터 **가치를 추출하고 결과를 분석하는 기술**
- 데이터 형태에 따라 **정형, 비정형** 데이터, 출처에 따라 **외부, 내부** 데이터, 생성 주체에 따라 **기계, 소셜, 관계** 데이터로 구분
- 신용카드 거래/전자 상거래 기록 등 **거래정보**, 각종 센서에 의한 **사물정보통신 정보**, GPS와 핸드폰에 의한 **위치정보**, **SNS 데이터**

빅데이터의 등장배경

• 정보사회 패러다임의 변화에 따른 **데이터 급격한 증가**



빅데이터의 등장배경

• **소셜 네트워크 서비스**의 발달에 따른 **비정형 데이터의 폭증**

• **데이터 저장/처리 비용**의 하락(클라우드 컴퓨팅)

• 대용량, 초고속 유무선 **네트워크의 보편화**

• **그림자 정보**(위치정보, 검색 패턴, 접속기록)의 증가

• **사물 정보통신망** 확산에 따른 센서 저변 확대

빅데이터의 특징

구 분	주 요 내 용
규모 (Volume)	• 기술적인 발전과 IT의 일상화가 진행되면서 해마다 디지털 정보량이 기하 급수적으로 폭증 → 제타바이트(ZB) 시대로 진입
다양성 (Variety)	• 로그기록, 소셜, 위치, 소비, 현실데이터 등 데이터 종류의 증가 • 텍스트 이외의 멀티미디어 등 비정형화된 데이터 유형의 다양화
복잡성 (Complexity)	• 구조화되지 않은 데이터, 데이터 저장방식의 차이, 중복성 문제, 데이터 종류의 확대 등 데이터 관리 및 처리의 복잡성이 심화
속도 (Velocity)	• 사물정보(센서, 모니터링), 스트리밍 정보 등 실시간성 정보 증가 • 실시간성으로 인한 데이터 생성, 이동(유통) 속도의 증가 • 대규모 데이터 처리 및 가치 있는 현재정보(실시간) 활용을 위해 데이터 처리 및 분석 속도가 중요

Part 2 공공데이터의 개념 / 종류

8

공공데이터

- 공공데이터란?
  - 정의 : 데이터베이스, 전자화된 파일 등 공공 기관이 법령 등에서 정하는 목적을 위하여 생성 또는 취득하여 관리하고 있는 전자적 방식으로 처리된 자료
  - 정부나 공공기관이 보유하고 있는 데이터
  - 공공 기관의 업무와 밀접한 데이터
- 공공데이터의 가치
  - 활용 여부/방법에 따라 국민 생활 향상에 밀접한 연관
- 공공데이터의 분류
  - 국가통계포털(KOSIS)에 등록된 공공데이터
  - 공공데이터 포털(www.data.go.kr)에 등록된 데이터

공공데이터

- 국가통계포털(KOSIS)에 등록된 공공데이터
  - 2016. 5. 25일 기준 총 955종
  - 기관별 작성 현황

구 분	작성 기관수	작성 통계수	종류별		작성방법		
			지정통계	일반통계	조사통계	보고통계	가공통계
계	397	955	93	862	418	451	86
◎ 정부기관	303	785	75	710	316	402	67
- 중앙행정기관	43	348	58	290	175	141	32
통계청	1	57	39	18	40	2	15
이외기관	42	291	19	272	135	139	17
- 지방자치단체	260	437	17	420	141	261	35
◎ 지정기관	94	170	18	152	102	49	19
- 금융기관	8	23	10	13	9	6	8
- 공사/공단	27	49	0	49	21	26	2
- 연구기관	20	37	2	35	31	4	2
- 협회/조합	22	32	4	28	27	3	2
- 기타기관	17	29	2	27	14	10	5

공공데이터

- **국가통계포털(KOSIS)에 등록된 공공데이터**
  - **부문별 작성 현황**

구 분	작성 통계수		종류별		작성 방법별		
	통계수	구성비	지정	일반	조사통계	보고통계	가공통계
계	955	100	93	862	418	451	86
인구	35	3.7	3	32	3	22	10
고용.임금	37	3.9	8	29	29	7	1
물가.가계소비(소득)	16	1.7	9	7	15	1	0
보건.사회.복지	202	21.2	6	196	149	47	6
환경	28	2.9	3	25	12	13	3
농림.수산	50	5.2	9	41	30	17	3
광공업.에너지	31	3.2	3	28	21	8	2
건설.주택.토지	40	4.2	2	38	14	20	6
교통.정보통신	49	5.1	4	45	22	24	3
도소매.서비스	16	1.7	6	10	14	2	0
경기.기업경영	89	9.3	26	63	62	4	23
국민계정.지역계정	21	2.2	5	16	0	0	21
재정.금융	17	1.8	1	16	2	15	0
무역.외환.국제수지	11	1.2	3	8	0	4	7
교육.문화.과학	56	5.9	3	53	41	14	1
기타(시도기본통계포함)	257	26.9	2	255	4	253	0

공공데이터

- **공공데이터 포털(www.data.go.kr)에 등록된 데이터**
  - **국민 건강정보 데이터, 상권정보 데이터 등 16,887건(2016. 8. 2 기준)**

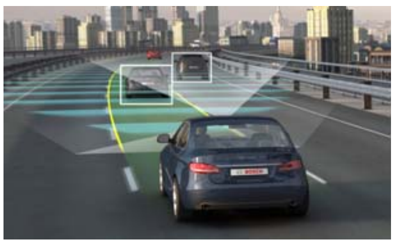
Part 3 빅데이터 활용 통계생산

13

빅데이터 활용 통계생산

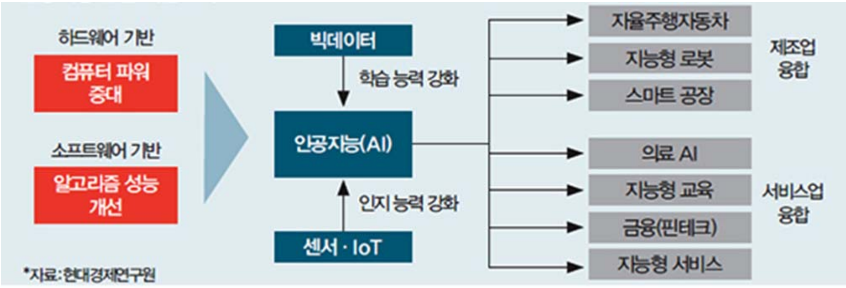
- **분석적 측면에서 빅데이터 통계생산(Big Data Analytics)**
  - **머신 러닝(Machine Learning)**
    - 컴퓨터 스스로 대량의 데이터를 수집하고 알고리즘 기반으로 학습, 분석하는 기술
    - 맞춤형 상품 마케팅, 검색어 자동 완성, 번역 서비스 등 다양한 분야에 활용 중

  - **딥 러닝(Deep Learning)**
    - 사람의 뇌가 사물을 구분하는 것 처럼 컴퓨터가 사물을 인식하고 분류하도록 훈련시키는 기술
    - 음성인식과 이미지 인식의 정밀도 향상을 바탕으로 목소리 인증, 얼굴 인식, 자동차 자율주행 기술 분야 활용



빅데이터 활용 통계생산

- 분석적 측면에서 빅데이터 통계생산(계속)
  - 인공 지능(Artificial Intelligence)
    - 인간의 지능으로 할 수 있는 사고, 학습, 자기계발 등을 컴퓨터가 할 수 있도록 하는 방법을 연구하는 분야로서, 컴퓨터가 인간의 지능적인 행동을 모방할 수 있도록 하는 것



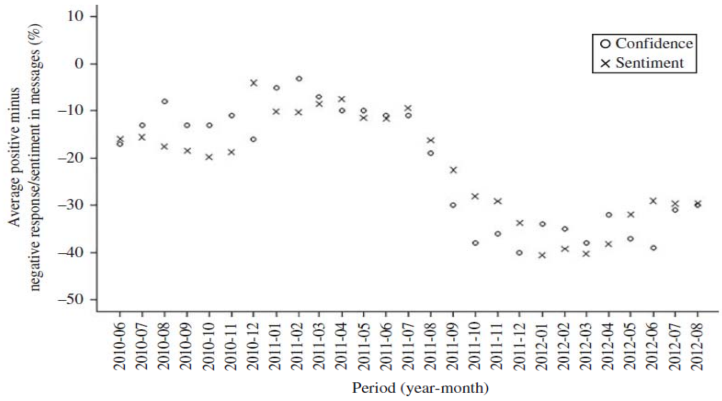
빅데이터 활용 통계생산

- 데이터 증대 측면에서 빅데이터 통계생산
  - 공공데이터와 민간데이터를 활용한 데이터 증대(Data Augmentation)
  - 데이터 증대를 위한 통계적 기법
    - 무응답 대체(Imputation)
    - 데이터 매칭(Data Matching)
    - 데이터 연계(Data Linkage)
  - 기존 조사 및 행정 자료의 활용도를 높이고 통계 이용자의 다양한 분석 수요 충족
  - 공공데이터와 민간데이터의 상호 보완을 통해 생산 통계 효율성 증대
  - 데이터의 질을 높이고 효율적 분석을 가능 ⇒ 새로운 가치 창출 및 미래 예측

해외 사례 – 소비자 신뢰지수와 SNS 감성경기 지수 비교분석(네덜란드)

- SNS 감성 경기 지수
  - 네덜란드 전체 인구의 70%가 SNS 활동중
  - 2010년 6월부터 2012년 8월까지 Twitter, Facebook등의 소셜미디어에서 총 16억 건의 메시지 수집(매월 5,200만 건의 메시지 수집)
  - 각각의 메시지를 내용에 따라 긍정, 부정, 중립으로 자동 분류
  - 긍정과 부정 메시지 개수를 비교하여 SNS 감성 경기지수 산출
- 소비자 신뢰지수
  - 전체 인구 중 매월 1000가구를 표본추출하여 현재 및 장래 재정 상태, 향후 경제 상황에 대한 인식을 5개 문항으로 설문 조사하여 결과를 지수화

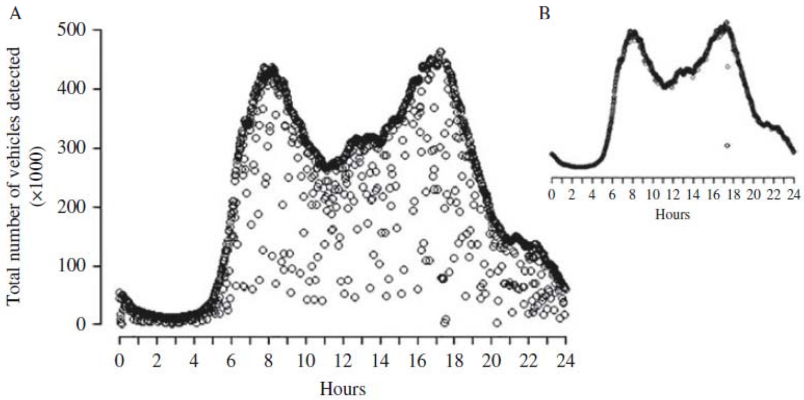
해외 사례 – 소비자 신뢰지수와 SNS 감성경기 지수 비교분석(네덜란드)



- 두 지수 간의 상관계수가 0.88로 상당히 연관성이 있음
- SNS감성 경기 지수를 활용하여 소비자 신뢰지수(공공 데이터)보다 저렴한 비용으로 자주 지수 산출 가능

해외 사례 - 도로 센서 데이터 분석을 통한 통행량 예측(네덜란드)

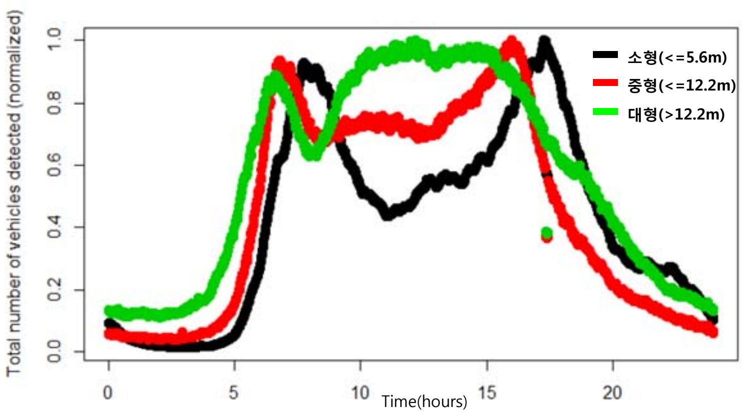
- 12,622개 센서 데이터(7,600만개)를 활용하여 분 단위(24시간) 교통 통행량 측정



  - A 그래프 : 센서 오류 등 다양한 이유로 Missing data가 발생하여 분산이 큼
  - B 그래프 : Missing data를 보정한 후 측정한 교통 통행량

해외 사례 - 도로 센서 데이터 분석을 통한 통행량 예측(네덜란드)

- 자동차 크기 별 시간에 따른 통행량 분석

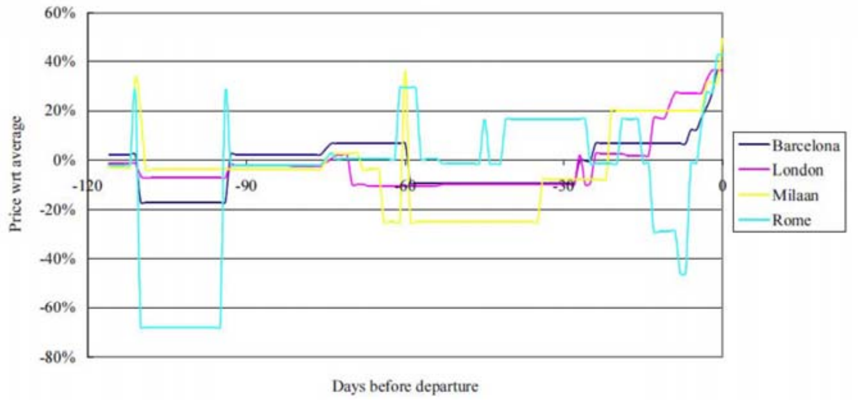


※ 자동차 크기 별 최대 대수 : 119,523(소형), 8,673(중형), 8,599(대형)

  - 소형차는 rush hour에 최대값을 나타내는 반면 중형 및 대형 화물차는 rush hour를 피하는 시간대에 최대값을 나타냄

해외 사례 – 인터넷 로봇을 활용한 온라인 항공권 가격 분석(네덜란드)

• 암스테르담에서 출발하는 4개의 노선에 대한 116일 전부터의 항공권 가격 정보 수집



• 날짜 별 티켓 가격 변동이 크기 때문에 CPI의 방식과 같이 한달에 한번 가격 정보를 수집하는 것은 무의미

⇒ 인터넷의 자동수집 로봇을 활용하여 수시로 데이터 수집 분석

Part 4 빅데이터 활용 통계의 품질검증

22

빅데이터 통계 품질검증

- **통계 품질 결정 기준[통계행정편람(2013)]**

구 분	설 명
정확성 (Accuracy)	추정값과 목표로 하는 모집단의 참값이 얼마나 근접하게 접근하고 있는가를 의미
관련성 (Relevance)	이용자에게 얼마나 의미가 있고 유용한 통계를 작성하여 제공하고 있는가를 의미
시의성 (Timeliness)	자료수집 및 생산에 소요되는 시간, 자료 수집의 빈도
접근 가능성 (Accessibility)	이용자들이 통계조사 결과를 용이하게 얻어낼 수 있는지 여부와 그 정도를 의미
비교 가능성 (Comparability)	시간, 공간이 서로 다른 자료 간에 신뢰할 만한 비교가 가능한가를 의미
일관성 (Coherence)	여러 출처에서 수집되어 작성된 자료가 서로 논리적으로 연결되고 타당성이 있는가를 의미

빅데이터 통계 품질검증

- **데이터 품질기준(한국 데이터베이스 진흥원)**

구 분	설 명
정확성	실 세계에 존재하는 객체의 값이 오류없이 저장되어 있음
일관성	정보시스템 내의 동일한 데이터 간에 불일치가 발생하지 않음
유용성	조직이 요구하는 데이터의 범위와 상세화 정도를 충족시킴
접근성	사용자가 원하는 데이터를 손쉽게 이용할 수 있음
적시성	응답시간 및 데이터의 최신성 유지와 같은 품질 요건에 얼마나 잘 대처하고 있는지를 의미
보안성	외부 및 내부 요인으로부터 데이터를 적절하게 보호함

빅데이터 통계 품질검증

- **해외사례(EUROSTAT)**
  - **유럽연합 통계청은 The role of Big Data in the Medernization of Statistical Production 프로젝트의 일환으로 빅데이터의 품질 기준을 제시**
  - **빅데이터를 공공데이터로 활용하기 위해 품질 검증을 3가지 단계로 제시**
    - **Input : acquisition, or pre-acquisition analysis of the data**
    - **Throughput : transformation, manipulation and analysis of the data**
    - **Output : the reporting of quality with statistical outputs derived from big data**
  - **빅데이터의 품질을 평가하기 위한 기본적인 원칙 제시**
    - **Fitness for use**
    - **Generic and flexible**
    - **Effort versus gain**

빅데이터 통계 품질검증

- **Input단계 품질 검증 Framework**

Hyperdimension	Quality Dimension	Factors to consider
Source	Institutional/Business Environment	Sustainability of the entity-data provider Reliability status Transparency and interpretability
	Privacy and Security	Legislation Data Keeper vs. Data provider Restrictions Perception
Metadata	Complexity	Technical constraints Whether structured or unstructured Readability Presence of hierarchies and nesting
	Completeness	Whether the metadata is available, interpretable and complete
	Usability	Resources required to import and analyse Risk analysis
	Time-related factors	Timeliness Periodicity Changes through time
	Linkability	Presence and quality of linking variables Linking level
	Coherence - consistency	Standardisation Metadata available for key variables (classification variables, construct being measured)
	Validity	Transparency of methods and processes Soundness of methods and processes
Data	Accuracy and selectivity	Total survey error approach Reference datasets Selectivity
	Linkability	Quality of linking variables
	Coherence - consistency	Coherence between metadata description and observed data values
	Validity	Coherence between processes and methods and observed data values

빅데이터 통계 품질검증

- **Throughput단계 품질검증**
  - **System Independence** : The result of processing the data should be independent of the hardware and software systems used to process it
  - **Application of quality dimensions** : In monitoring process quality, quality dimensions as articulated here and in other frameworks should be used as a guide for quality evaluation
  - **Steady states** : that the data be processed through a series of steady states, which are stable versions of the data that can be referenced by future processes and by multiple parts of the organization
  - **Quality gates** : that the NSO employ quality gates as a quality control business process.

    ※ Quality gates : Checkpoint in the business process at which the quality of the data is explicitly assessed

빅데이터 통계 품질검증

- **Output단계 품질검증 Framework**

Hyperdimension	Quality Dimension	Factors to consider
Source	Institutional/Business Environment	Type of data source Arrangements and quality assurance Type of use of the BD source
	Privacy and Security	Legislation Actual limitations in the use of data Actions undertaken
Metadata	Complexity	Data treatment; output limitations
	Accessibility and Clarity	Data and metadata accessibility Clear definitions, explanations Conformity to standards
	Relevance	Extent to which the data measures the concepts meant to be measured for its intended uses
Data	Accuracy and selectivity	Traditional measures of accuracy Selectivity
	Validity	Correlation with similar metrics Utility Conceptual soundness
	Coherence - Linkability	
	Coherence - Consistency	
	Time-related factors	Timeliness, Periodicity

토의

- 빅데이터를 활용한 통계의 생산은 여러 측면에서 판단할 때 그 필요성이 인정됨
- 활용 방안에 따라 통계적으로 적절하고 타당한 절차가 요구됨
  - 국가 통계의 작성
  - 빅데이터 분석을 통한 컨설팅
  - 보조자료로서의 활용
  - 제공정보의 확장으로서의 역할
- 각 방안 별 특성에 맞춘 품질검증 절차가 필요함
  - 데이터와 이로부터 생성된 통계의 구분
  - 국가 통계에 대한 신뢰도를 하락시킬 수 있는 요인을 배제
  - 특별히 데이터 및 통계의 대표성에 대한 품질 검증

참고문헌

- Daas, P. J. H., Puts, M. J., Buelens, B., and van den Hurk, P. A. (2015). Big Data as a source for official statistics. Journal of Official Statistics, 31(2), 249-262.
- UNECE(2014). A Suggested Framework for National Statistical Offices for assessing the Quality of Big Data. NTTS 2014 (New Techniques and Technologies for Statistics) Conference.
- Olav Bosch, Dick Windmeijer(2014). On the Use of Internet Robots for Official Statistics. Meeting on the Management of Statistical Information System(MSIS 2014).
- Daas, P. J. H., Puts, M. J.(2014). Social Media Sentiment and Consumer Confidence. Statistical. European Central Bank Statistical Paper Series
- Daas, P. J. H., M. Roos, M. van de Ven, and J. Neroni(2012). Twitter as a Potential Data Source for Statistics. Discussion paper 201221, The Hague/Heerlen: Statistics Netherlands.
- Puts, M. J. H., Tennekes, M., Dass, P. J. H., Blois, C.(2016). Using Huge Amounts of Road Sensor Data for Official Statistics. European Conference on Quality in Official Statistics(2016)

Q/A

빅데이터 거버넌스 구축 및 활용사례

- 데이터 센터를 중심으로-

2016. 8.

통계청
Statistics Korea

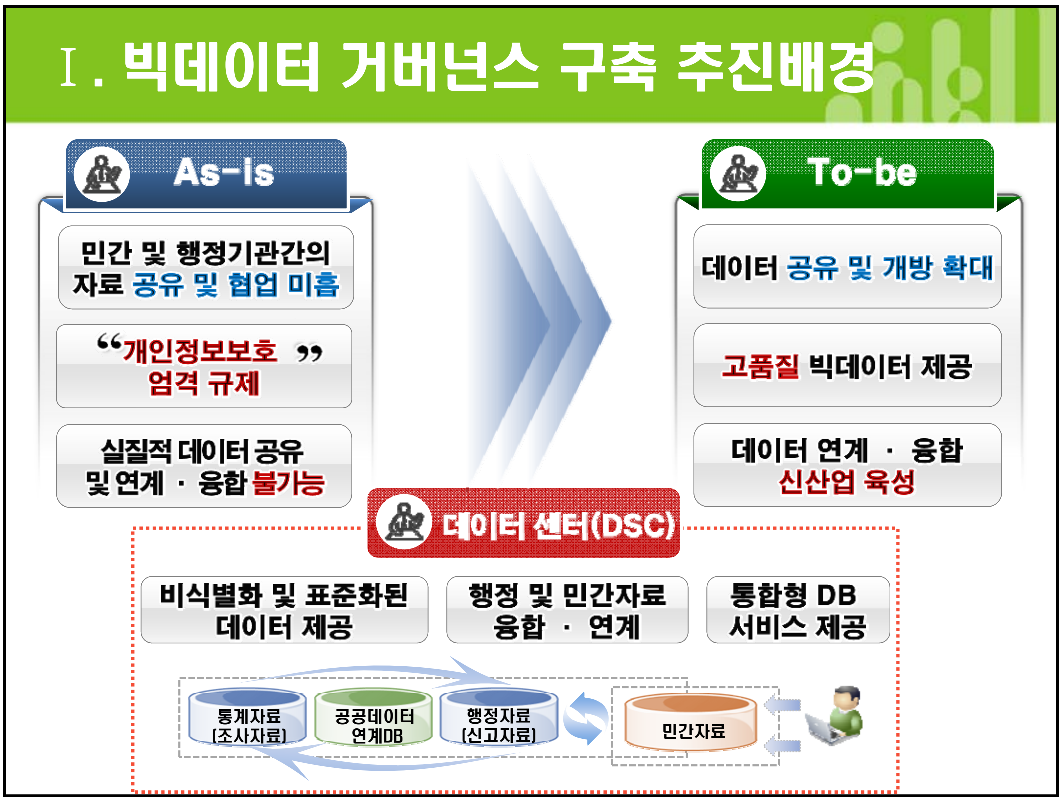
목 차

Ⅰ. 빅데이터 거버넌스 구축 추진배경

- 데이터의 내재 가치를 제대로 활용하기 위해서는 **실질적인 데이터 공유와 연계.융합에 대한 제도적 장애 개선 필요**
- 통계청이 보유한 통계데이터와 민간빅데이터간 **비식별화 방식의 연계를 통한 통계작성 활용 사례 발굴 및 확산** 필요

연계.융합으로 **데이터의 잠재 가치를 최대한 활용**

공공데이터 및 민간빅데이터 거버넌스 구축


Ⅰ. 빅데이터 거버넌스 구축 추진배경
As-Is
민간 및 행정기관간의 자료 공유 및 협업 미흡
"개인정보보호 엄격 규제"
실질적 데이터 공유 및 연계 · 융합 불가능
To-be
데이터 공유 및 개방 확대
고품질 빅데이터 제공
데이터 연계 · 융합 신산업 육성
데이터 센터(DSC)
비식별화 및 표준화된 데이터 제공
행정 및 민간자료 융합 · 연계
통합형 DB 서비스 제공
통계자료 (조사자료)
공공데이터 연계DB
행정자료 (신고자료)
민간자료


Ⅱ. 빅데이터 허브로서의 데이터센터

Ⅱ-1. 데이터센터(DSC) 개념 및 운영방향

데이터센터 개념 | 운 영 방 향

수요자가 공공데이터 및 민간자료를 편리하게 이용하고 자료 융합이 가능하도록 구축된 데이터 허브

<DSC 구축 前>

<DSC 구축 後>

Ⅱ-1. 데이터센터(DSC) 개념 및 운영방향

데이터센터 개념 | **운 영 방 향**

자료처리 전문기관의 표준화.품질점검 등을 통한 고품질의 데이터 제공

- 각 기관별로 공공데이터의 경우 품질점검 등을 실시하지 않아 자료 활용 제약

공공데이터 상호간, 공공데이터와 민간데이터의 연계 지원을 통한 데이터의 잠재적 가치 확장

- 개인정보 비식별화 처리를 지원을 통해 공공 및 민간기관간의 데이터 연계

Ⅱ-2. 데이터센터(DSC) 주요 기능 및 역할

행정자료 및 부분별 연계DB 제공

- 통계청이 입수·보유하고 있는 **각종 행정자료를 표준화 및 자료 정제** 후 통합 서비스

행정자료(155개 기관 180종)를 **표준화(주소 및 산업분류), 품질점검, 비식별화 등** 자료처리 후 DB 제공

<산업분류 표준화 사례>

국세청 업종코드	종목 (세세분류)	적용범위		산업 분류코드	한국표준산업분류 항목명
552107	· 피자 햄버거 샌드위치 및 유사 음식점업 · 치킨 전문점	피자, 햄버거, 닭튀김, 간이양식, 아이스크림 등 체인화 된 음식을 취급하는 업소	↗	56191	· 제과점업 · 아이스크림
			→	56192	· 피자 햄버거 샌드위치 및 유사 음식점업
			↘	56193	· 치킨 전문점

<주소 표준화 사례>

자료A	청운아파트	103동	1712호
자료B	청운APT	103	1712
자료C	청운아파트 103동 1712호		
자료D	청운아파트	103-1712	

⇨

자료A	청운아파트	103	1712	20*
자료B	청운아파트	103	1712	20
자료C	청운아파트	103	1712	20
자료D	청운아파트	103	1712	20

Ⅱ-2. 데이터센터(DSC) 주요 기능 및 역할

행정자료 및 부분별 연계DB 제공

- 통계자료와 행정자료, 행정자료간 데이터 연계를 통해 **연계DB 구축**하여 이용자의 편리성 제고

정제된 행장자료 DB와 통계자료 DB간 연계한 **부분별.주제별 DB*를 제공**

① 인구·가구DB : 원행정자료(주민등록자료) + 분석DB(인구, 가구) + 농림어업행정(통계)등록부 등

② 사업체·기업체DB : 원행정자료(법인세, 법인등기 등) + 분석DB (사업자등록&종사자) + 사업체모집단 + 농림어업행정(통계)등록부

③ 주택·건물DB : 원행정자료(건축물대장, 주택공시정보)

④ 경제활동DB : 원행정자료(4대보험자료, 일용·근로소득·사업소득명세서), 민간데이터(통신사, 신용카드 정보, 구인·구직정보 등)

< 부문별 DB간 연계율 ('15.12월 기준) >

연 계	연계율	비 고
사업체·기업체 → 경제활동	98.33%	사업체/기업체의 경제활동 현황
경제활동 → 인구·가구	98.10%	경제활동 인구의 인구/가구 특성
사업체·기업체 → 인구·가구	97.93%	사업체/기업체 대표자의 인구/가구 특성
인구·가구 → 주택·건물	94.33%	인구의 주거 특성
사업체·기업체 → 주택·건물	92.22%	사업체/기업체의 사업장(건물) 특성

Ⅱ-2. 데이터센터(DSC) 주요 기능 및 역할

각 종 통계자료(조사자료)를 **통계적 기법 처리 및 정제 후 통합 DB로 구축 및 서비스**

" 통계데이터 제공 "

공유 & 공개

통계 원시자료(Raw data)의 공유·공개를 통해 창조경제의 원천이 되는 **통합형 서비스 구축**

< 통합 DB 구축건수 >

	계	중앙행정기관	지방자치단체	협회 및 기타
신규	121종 (71개 기관)	43종 (13개 기관)	21종 (20개 기관)	57종 (38개 기관)

< 서비스 제공 건수 >

서비스 종수(종)/ 기관수(개)		'14년(1단계)	'15년(2단계)	'16년(3단계)	'17년(4단계)
서비스	누적	5 / (4)	21 / (16)	51 / (40)	106 / (80)

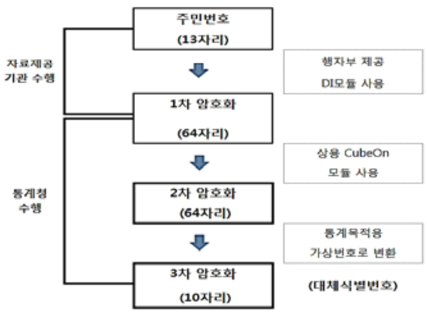
※ 통계청 통계자료 40종은 포함하지 않은 수치

Ⅱ-2. 데이터센터(DSC) 주요 기능 및 역할

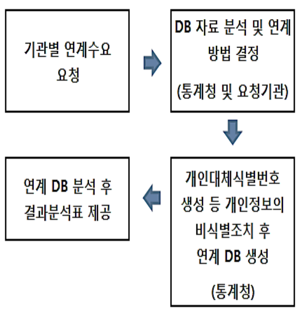
" 데이터간 연계 서비스 제공 "

통계청이 **개인정보 비식별화* 처리를 지원**함으로써 공공 및 민간기관간의 **데이터 연계 DB 사례 발굴, 확산** 및 지원

개인정보의 비식별화 과정



DB간 자료 연계 과정



Ⅲ. 공공데이터와 민간 빅데이터 연계사례

Ⅲ-1. 자료연계 및 보완방법

◇ 데이터 공유의 한 주체가 정부임을 감안하여 **완전하고 실질적인** 비식별화 조치 시행

◇ **그간의 비식별화 경험과 각국 통계청의 비식별화 방식을 반영, 단계적인 비식별화 조치 마련 (과정적 접근)**

1. 비식별화와 연계

1단계 (비식별화 조치)		2단계 (연계)		3단계 (사후관리)
• 기관별 데이터에 대한 식별자 암호화	➡	• 암호화된 식별자를 연계키로 데이터간 연계 • 연계키 삭제	➡	• 연계 DB 외부 반출 금지

Ⅲ-1. 자료연계 및 보완방법

2. 추가적 보완 조치

구분	관련 여건	보완 조치 내용
연계를 위한 데이터 제공방식	• 관련법상 제3자 제공 금지 규정 ⇒ 민간기관은 연계를 위한 1차 암호화 데이터의 제공에 한계 • 통계법상 통계자료 제공 가능 ⇒ 통계법상 식별할 수 없는 형태로 통계자료를 처리 후 제공 (제31조)	• 통계청이 민간기관에 데이터 제공 • 민간기관이 데이터간 연계 및 연계키 삭제 후 분석용 DB 작성 → 양기관 공유 * 분석용 DB는 개인정보가 아님 (법률 자문 결과)
데이터 관리 및 분석방식	• 상호간에 소관 데이터의 유출 우려	• 통계청의 독립된 제한 공간(데이터 센터)에서 자료 분석 및 통계 작성 → 민간기관의 서버를 가지고 입주 • 자료 접근권자 인가

Ⅲ-2. 연계사례 – 신혼부부 신용 DB 구축

◇ 공공데이터와 민간 빅데이터간 연계 시범사례로서 **통계청 공공데이터와 민간 신용정보회사의 빅데이터간 연계·분석**

◇ **저출산 대책 등 지원을 위한 신혼부부에 관한 통계**를 우선적으로 DB로 구축·분석

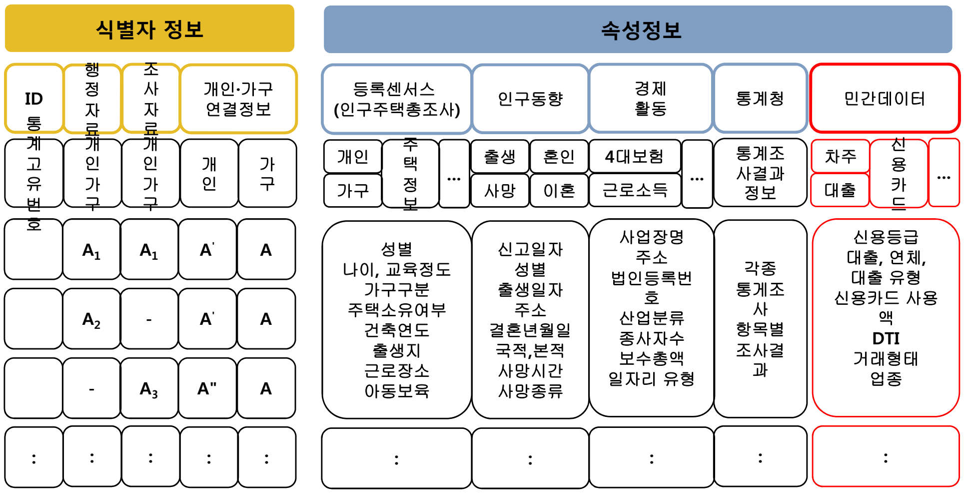
▌연계 데이터

- **통계청 통계자료 및 공공 행정자료와 신용정보회사의 신용 자료**

자료원	활용 항목	자료원	활용 항목
인구동향통계	혼인날짜, 연령, 직업, 학력, 출산자녀수	임금근로 일자리통계	전년도보수총액
인총 및 등록센서스	주택형태, 가구 유형	사업소득신고	매출액
주택소유통계	보유주택수	민간 부채자료	소득, 신용등급, 대출잔액, 연체금액, 부채상환액, 카드사용액

Ⅲ-2. 연계사례 – 신혼부부 신용 DB 구축

신혼부부 신용DB 자료연계 구성항목



Ⅲ-2. 연계사례 – 신혼부부 신용 DB 구축

▌작성대상 및 표본

- **(대상) '14.11.1. 기준 혼인 신고 후 5년이 경과하지 않은 부부**
  - '09. 11. 1.~'14. 10.31.까지 혼인신고를 하고 '14. 11. 1. 현재 혼인을 유지하고 있는 신혼부부(재혼 포함)로서 이혼·사별한 가구 제외

- **(표본) 신혼부부 모집단 154만 부부 중 횡단 및 종단 각 5만쌍**
  - 횡단표본 : '14년 기준 신혼부부 5만쌍으로 1~5년차로 구성
  - 종단표본 : '10년 결혼한 신혼부부 5만쌍

Ⅲ-2. 연계사례 - 신혼부부 신용 DB 구축

▌분석 내용

- **(횡단 분석) 신혼부부 가구의 일반현황, 사회경제적 특성 등**
  - (일반현황) 성별, 연령별, 시도별, 직업별, 교육정도별
  - (사회경제적 특성) 주택소유, 자녀수, 자녀수와 주택·부채
  - (부채 및 소득현황) 부채·소득 규모, 유형, 부채와 주택 등
- **(종단 분석) 연도별 현황, 소득분위별 현황 등 동태적 분석**
  - 인구특성 및 부채의 동태적 분석 (신혼부부의 자녀수, 직업 변동 및 대출잔액 변동 현황)
  - 소득분배 현황 분석 (신혼부부의 소득이동 및 빈곤탈출 현황)

Ⅳ-1. 신규데이터 연계 사례 발굴

▌신규 빅데이터 연계 사례 발굴

- 신용보증재단중앙회의 소상공인 개인별 자료 연계

  * 소상공인 창폐업 정보 및 가구의 특성 분석, 자영업자 보증포화지수 등 개발

- 신용카드사의 카드 승인액 자료 연계 및 물가(POS데이터) 자료 활용

  * 개인별·가구별 소비현황, 카드승인액 지수 등 민생지표 작성, 물가정보
  * 현재 여신금융협회를 통해 전체카드사의 일단위 업종별·지역별 카드승인액 자료 입수중이며, 개별 카드사의 개인별 신용카드 사용자료 입수 협의 중

- 통신사의 위치정보와 공공데이터 연계

  * 출퇴근 거리.시간.인구수 등 교통시설 공급 정책에 필요한 정보, 등록센서스 등의 실거주 여부 확인, 유동인구 등 확대된 개념의 인구 정보 제공

Ⅳ-2. 데이터 센터의 확대

▌데이터센터의 확대

- 데이터센터 구축을 위한 ISP 실시 : 2017년

  * 데이터센터 구축 및 운영을 위한 중장기 추진 전략 도출

- 데이터센터 구축 및 외부 서비스 제공 : 2018년 이후

  * 자료의 민감성 등을 고려 오프라인 형태로 구축 운영
  * 현재 통계청 내부적으로는 운영 중

- 데이터센터 온라인 서비스 제공 : 중장기

  * 통계뱅크플랫폼 및 원격분석시스템 구축을 통한 온라인 서비스 제공


Big Data
감사합니다

제 2 세 션

환경변화와 통계개발

- SDGs 글로벌지표의 국내 적용가능성 평가 / 박영실 사무관
- 국민이전계정 개발 현황과 향후 계획 / 김현경 사무관
- 시군구 GRDP 분배계정 개발과 활용 / 김시윤 사무관
- 금융위기 이후 자금순환통계의 확충 / 김성준 과장
- 기업경영분석 분위수 통계 작성배경 및 결과 / 김혜림 과장

국가통계발전포럼(2016.8.31)

SDGs 글로벌지표의 국내 적용가능성 평가

박영실(통계개발원 연구기획실)

목차

1. SDGs란 무엇인가?
2. 글로벌지표 프레임워크
3. 국내 적용가능성 평가
4. 나가며

1. SDGs란 무엇인가?

『Transforming our world: the 2030 agenda for sustainable development』

출처 : UNDG (2015)

1. SDGs란 무엇인가?

1. SDGs란 무엇인가?

1. SDGs란 무엇인가?

2015. 9.	2016. 3.	2016. 7.
UNGA	UNSC	HLPF
목표 및 세부 목표 채택	글로벌 지표 합의	국가별 평가

1. SDGs 란 무엇인가?

글로벌차원	지역차원	국가차원
HLPF 핵심역할	UN 지역위원회 동료검토,모범사례공유	이행계획수립 국가주도 정기적, 포괄적 평가
SDGs연례이행 보고서	글로벌과 국가의 연계	모든 후속조치의 기초

Statistical Research Institute 7

2. 글로벌지표 프레임워크

IAEG-SDGs 주도의 지표 개발

- 제46차 통계위원회(15.3.) 창설
  28개국 국가통계기관 대표가 회원
- 역할: 지표 개발, 이행 위한 기술적 지원, 방법론 점검, 통계역량 강화 활동, 자료수집과 활용에 관한 지원 등
- 지표 선정 기준: 명확성, 적절성, 시의성, 비교가능성 등
  제47차 통계위원회(16.3.) 지표 보고서 제출

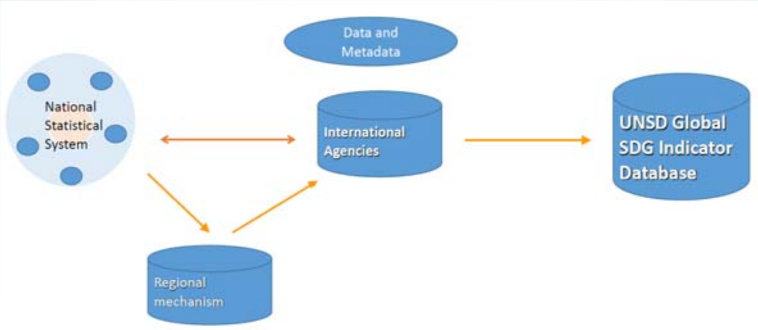
Statistical Research Institute 8

2. 글로벌지표 프레임워크

목표	세부목표 수	지표 수
1. 빈곤종식	7	12
2. 기아해소와 지속가능농업	8	14
3. 건강 및 웰빙	13	26
4. 양질의 교육	10	11
5. 양성평등	9	14
6. 물과 위생	8	11
7. 에너지	5	6
8. 양질의 일자리와 경제성장	12	17
9. 혁신과 인프라	8	12
10. 불평등완화	10	11
11. 지속가능한 도시	10	15
12. 지속가능한 소비, 생산	11	13
13. 기후변화 대응	5	7
14. 해양 생태계	10	10
15. 육상 생태계	12	14
16. 평화와 정의, 제도	12	23
17. 파트너십	19	25
합	169	241 (중복제외시 230)

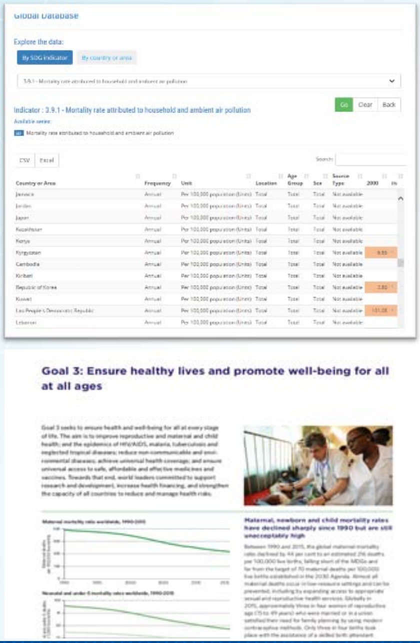
2. 글로벌지표 프레임워크

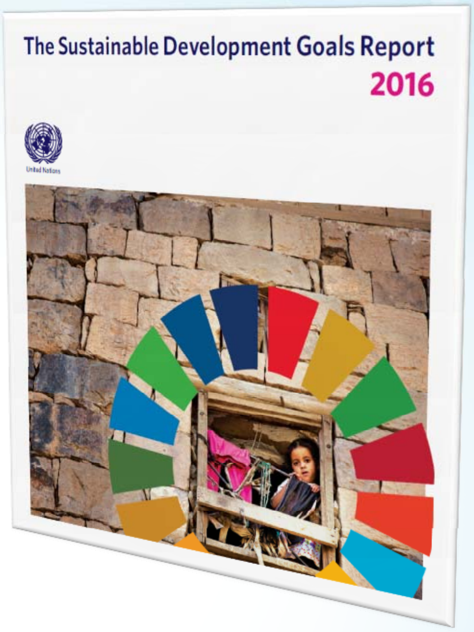
√ 비교가능하고 표준화된 공식통계 제공



출처 : IAEG SDGs (2016)

2. 글로벌지표 프레임워크





Statistical Research Institute | 11

2. 글로벌지표 프레임워크

글로벌차원	국가차원
Tier 1 개념명확 방법론 및 표준정립 정기적 자료생산o	**즉각 제공 가능** • 1-1: 현재 바로 제공 가능, 정기적 • 1-2: 산식 통해 계산해서 제공 가능, 정기적 • 1-3: 국제기구에서 제공
Tier2 개념명확 방법론 및 표준정립 정기적 자료생산x	**중단기 전략 수립** • 2-1: 시계열 부정기 (조사주기 개선 필요) • 2-2: 항목 및 조사대상 변경 등 개선 • 2-3: 새로운 조사 개발
Tier3 방법론 및 표준 미정립or개발중	**장기 전략 수립** • 3-1: 메타데이터 개념 모호

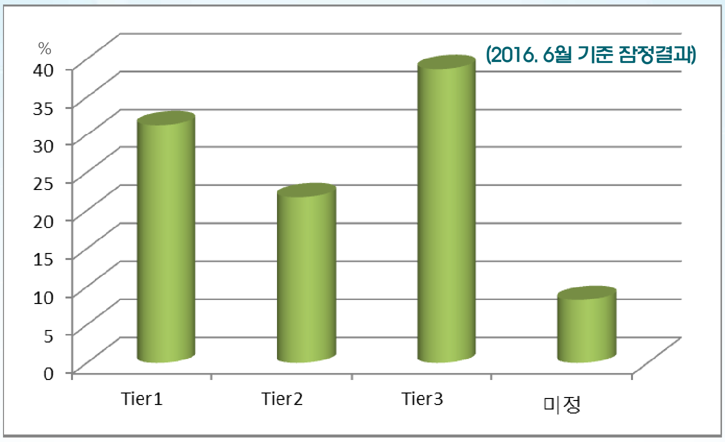
√ 비해당(1.한국맥락무관, 2.이미 달성) 및 법/제도관련지표

Statistical Research Institute | 12

2. 글로벌지표 프레임워크

Indicators	데이터베이스 (2016.7.29 기준)	
	전체	한국
1.1.1 Proportion of population below the international poverty line, by sex, age, employment status and geographical location (urban/rural)	o	
1.2.1 Proportion of population living below the national poverty line, by sex and age	o	
1.2.2 Proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions		
1.3.1 Proportion of population covered by social protection floors/systems, by sex, distinguishing children, unemployed persons, older persons, persons with disabilities,	o	o
1.4.1 Proportion of population living in households with access to basic services		
1.4.2 Proportion of total adult population with secure tenure rights to land, with legally recognized documentation and who perceive their rights to land as secure, by sex and by		
1.5.1 Number of deaths, missing persons and persons affected by disaster per 100,000 people	o	
1.5.2 Direct disaster economic loss in relation to global gross domestic product (GDP)[a]	o	
1.5.3 Number of countries with national and local disaster risk reduction strategies[a]	o	o
1.a.1 Proportion of resources allocated by the government directly to poverty reduction programmes		
1.a.2 Proportion of total government spending on essential services (education, health and social protection)		
1.b.1 Proportion of government recurrent and capital spending to sectors that disproportionately benefit women, the poor and vulnerable groups		
2.1.1 Prevalence of undernourishment	o	
2.1.2 Prevalence of moderate or severe food insecurity in the population, based on the Food Insecurity Experience Scale (FIES)	o	

3. 국내 적용가능성 평가



√ Tier 1 비중이 높은 분야: 목표 3, 7, 9

√ Tier 3 비중이 높은 분야: 목표 12, 13, 14

3. 국내 적용 가능성 평가

√ 데이터: 글로벌 요구 수준 vs 국가 제공 가능 수준 갭
ex) UN 제시기준보다 국내에서 더 정확한 자료 제공 가능
: 무상돌봄과 가사노동에 소요되는 시간 비율

√ 개념 및 측정방법: 국제 기준 vs 국내 기준 갭
ex) 영양결핍현황 국제 기준 5% 미만, 국내 기준 6.8%

Tier 구분 과정에서의 쟁점들

√ 메타데이터 개념 여전히 모호
√ 한국 상황에 적합하지 않은 지표 (맥락무관, 기달성)
√ 글로벌지표는 없으나 유사지표 및 관련지표의 존재
√ 경계성 티어의 존재 (세분화 여부 등)

Statistical Research Institute 15

3. 국내 적용 가능성 평가

글로벌차원	국가차원
Tier 1 개념명확 방법론 및 표준정립 정기적 자료생산o	**즉각 제공 가능** • 1-1: 현재 바로 제공 가능, 정기적 • 1-2: 산식 통해 계산해서 제공 가능, 정기적 • 1-3: 국제기구에서 제공
Tier2 개념명확 방법론 및 표준정립 정기적 자료생산x	**중단기 전략 수립** • 2-1: 시계열 부정기 (조사주기 개선 필요) • 2-2: 항목 및 조사대상 변경 등 개선 • 2-3: 새로운 조사 개발
Tier3 방법론 및 표준 미정립or개발중	**장기 전략 수립** • 3-1: 메타데이터 개념 모호

√ 비해당(1.한국맥락무관, 2.0이미 달성) 및 법/제도관련지표

Statistical Research Institute 16

3. 국내 적용 가능성 평가

SDGs 국가지표 개발

글로벌지표 검토	국내 유사지표 검토
• 241(230)개 지표의 국내 적용 가능성 진단	• 지속가능발전지표 • 국가주요지표 • 삶의 질 지표

Statistical Research Institute | 17

4. 나가며

자료생산패러다임의 변화

- 현장조사방법의 선진화
- 다양한 원천의 자료 활용

생산자에서 조정자로 진화

- 자료 품질관리, 표준설정 및 모니터링
- 국가통계생산기관을 넘어선 통계 생태 시스템 구축

글로벌 통계 사회에의 기여

- (세분화) 자료 제공 통한 SDGs 글로벌데이터 베이스 구축
- 개도국 통계역량 강화 지원

Statistical Research Institute | 18


Thank You
통계청
통계개발원

제6회 국가통계 발전포럼
세션Ⅱ. 환경변화와 통계개발 (2016.8.31)

통계청
통계개발원

저출산 · 고령화 사회 대비 정책 지원을 위한

국민이전계정(NTA) 개발 현황과 향후 계획

통계청 통계개발원 동향분석실
김 현 경

통계청
통계개발원

Contents

한국의 저출산 · 고령화 현상

통계청 통계개발원

※ 출처 : 정책브리핑 정책뉴스, "향후 5년 인구위기 '골든타임' … 만혼 줄이고 맞벌이 출산율 ↑"
2015.2.6. 보건복지부 (저출산고령사회위원회 운영지원단)

국민이전계정 (National Transfer Accounts, NTA)

통계청 통계개발원

▪ 국민이전계정이란?

– 세대간 자원의 흐름을 측정하는 시스템

– 연령집단간 경제적 자원 흐름을 측정할 수 있는 통계 자료

- 생애주기(유년층, 노동연령층, 노년층)별 적자/흑자 내역을 파악하여,
- 세대별로 달리 나타나는 적자/흑자 내역에 따라 자원이 이전(transfer)되거나 재배분(asset reallocation)되는 과정을 파악해 낼 수 있음
  - – 국민계정이나 소득통계는 제도단위(가계, 기업, 정부 등)나 계층(빈부, 도농 등)별 경제적 자원의 규모와 분포를 측정하는데 반해, 국민이전계정은 연령간 경제적 자원 흐름을 측정

▪ 국민이전계정의 필요성

– 경제적 생애주기(Economic Lifecycle)의 중요성 증가

- 부양기간 증가와 지속가능성의 문제, 경제적 행위와 연령의 높은 연관성, 생애복지의 필요성

– 복잡하고 다양한 세대간 경제 시스템에 대한 이해 필요

- 빈곤, 경제 성장, 세대간 형평성에 대한 함의

※ 출처 : 이상협(2012), 발표자료 "인구고령화와 국민이전계정"

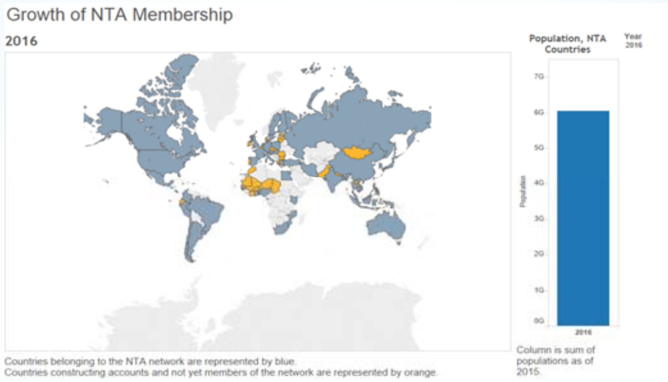
국민이전계정 (National Transfer Accounts, NTA)

통계청 통계개발원

■ **NTA Project**

– 2016년 7월 현재 49개국

- 아시아-태평양(14) : 한국, 일본, 중국, 인도, 대만, 인도네시아, 필리핀, 태국, 베트남, 호주, 방글라데시, 캄보디아, 말레이시아, 이란
- 아메리카 (12) : 미국, 브라질, 칠레, 우루과이, 코스타리카, 멕시코, 아르헨티나, 캐나다, 콜롬비아, 자메이카, 엘살바도르, 페루
- 유럽(15) : 프랑스, 스웨덴, 오스트리아, 슬로베니아, 헝가리, 핀란드, 독일, 스페인, 영국, 이탈리아, 터키, 룩셈부르크, 네덜란드, 폴란드, 러시아
- 아프리카 (8) : 케냐, 나이지리아, 남아프리카공화국, 세네갈, 모잠비크, 베닝, 가나, 이집트



※ 출처 : National Transfer Accounts Homepage (http://www.ntaccounts.org)

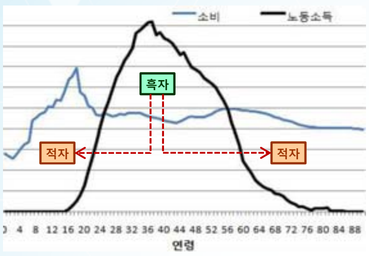
국민이전계정 (National Transfer Accounts, NTA)

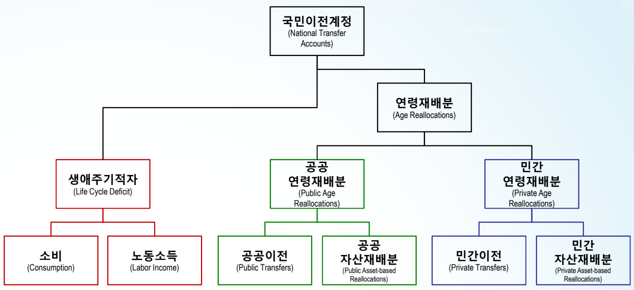
통계청 통계개발원

■ **국민이전계정 기본 개요 및 구조**

– 국민이전계정은 생애주기적자 부분과 연령재배분 부분으로 크게 구분되어 있음

- 생애주기적자 부분과 연령재배분 부분은 항상 일치되는 구조임





생애주기적자	공공연령재배분	민간연령재배분
유년/장년/노년의 생애주기별 적자/흑자 내역 파악	조세 · 연금 · 건강보험 등의 공적 이전 등에 관한 정보 파악	가계 내 상속, 증여 등 부모-자녀간 사적 이전 등에 관한 정보 파악

국민 전체의 연령별 노동소득과 소비, 공적 이전과 가구내 및 가구간 사적 이전 등의 내역을 파악

· 이전(transfer) : 재화와 서비스 또는 현금을 한 연령집단에 속하는 개인으로부터 다른 연령집단에 속하는 개인에게 기대 없이 이전하는 거래 (예. 교육 및 보건 분야의 정부 지원)
· 자산재배분(asset reallocation) : 자산처분 또는 취득을 수반하여 현재의 소득을 과거 또는 미래의 소비를 위한 소득으로 이동시키는 거래로 자산소득(예. 이자, 임대료 등)과 저축으로 구성

국민이전계정 결과 시산

통계청 통계개발원

■ 국민이전계정 결과표

– 연령집단 간 경제적 자원 흐름을 생애주기별 노동 소득과 소비, 세대간 이전 등으로 각각 구분하여 측정

- 다양한 기초 미시 자료를 활용하여 국민계정체계(SNA)와 일관성을 유지하면서 국민소득계정을 연령별로 분해

구 분	연 령					관 계 식
	총량값	0세	1세	…	85세+	
생애주기적자						= 소비 – 노동소득
소비						= 공공소비 + 민간소비
공공소비						= 공공교육소비 + 공공보건소비 + 공공기타소비
민간소비						= 민간교육소비 + 민간보건소비 + 민간기타소비
노동소득						= 임금소득 + 자영자노동소득
임금소득						-
자영자노동소득						-
연령재배분						= 생애주기적자
공공연령재배분						= 공공이전 + 공공자산재배분
공공이전						= 공공이전(유입) – 공공이전(유출)
공공자산재배분						= 공공자산소득 – 공공저축
민간연령재배분						= 민간이전 + 민간자산재배분
민간이전						= 민간이전(유입) – 민간이전(유출)
민간자산재배분						= 민간자산소득 – 민간저축

국민이전계정 결과 시산

통계청 통계개발원

■ 국민이전계정 시산을 위해 필요한 기초 자료 및 시산 작업 과정

<국민이전계정 시산에 필요한 기초 자료>

기초 자료	관련 세부 항목
국민계정	제도부문별 소득 계정, 가계 및 가계봉사 비영리단체의 목적별 최종소비지출, 일반정부의 기능별 최종소비지출
가계동향조사	민간 소비 (민간 교육 소비, 민간 보건 소비, 민간 기타 소비) 재산 소득, 이자 소득, 이자 비용 가구간 이전, 공적 연금, 기초노령연금, 사회수혜금, 사회적 현물이전
노동패널조사	임금 소득, 혼합소득 (자영자 노동 소득), 순생산 및 수입세
교육 통계 연보	공공 교육 소비
건강 보험 통계 연보	공공 보건 소비

<국민이전계정 기본 시산 작업 과정>

총량값 구성
(국민계정 자료 이용)

연령 프로파일 작성
(개별 미시 자료 이용)
→ 평활화 작업
(Super Smoothing)
→ 최종 연령프로파일 작성
(총량값 적합)

NTA 총량값 시산 결과

통계청 통계개발원

- **국민계정 자료를 활용한 NTA 총량값 시산 결과**
  - 생애주기적자 부분과 연령재배분 부분은 항상 일치
  - 전체적으로 총량값 기준으로 생애주기적자가 발생하고 있음
    - 일반적으로 소비의 원천은 노동 소득만이 아니기 때문에, 대부분의 국가들에서는 생애주기적자가 발생하고 있음

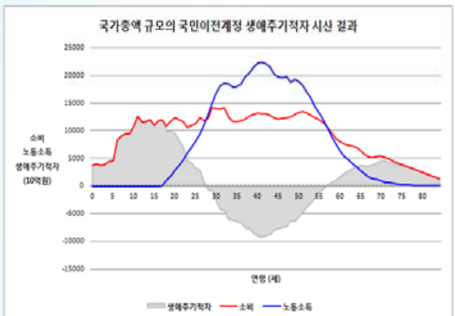
	2008 SNA 반영 기준년 개편 전		2008 SNA 반영 기준년 개편 후			
구 분	2009년 총량값	2010년 총량값	2010년 총량값	2011년 총량값	2012년 총량값	2013년 총량값
생애주기적자	**91,827.8**	**93,166.8**	**112,050.0**	**129,725.6**	**135,995.4**	**135,010.9**
소비	640,094.1	676,036.6	699,637.3	750,797.9	785,936.8	816,060.0
노동소득	548,266.3	582,869.8	587,587.3	621,072.3	649,941.4	681,049.1
연령재배분	**91,827.8**	**93,166.8**	**112,050.3**	**129,725.4**	**135,995.1**	**135,011.0**
공공연령재배분	-36,561.3	-47,735.7	-55,720.8	-60,798.2	-60,290.4	-54,950.7
공공(순)이전	0.0	0.0	0.0	0.0	0.0	0.0
공공자산재배분	-36,561.3	-47,735.7	-55,720.8	-60,798.2	-60,290.4	-54,950.7
민간연령재배분	128,389.1	140,902.5	167,771.1	190,523.6	196,285.5	189,961.7
민간(순)이전	-626.3	-3,539.8	-6,124.9	-5,208.6	-6,150.2	-4,546.8
민간자산재배분	129,015.4	144,442.3	173,896.0	195,732.2	202,435.7	194,508.5

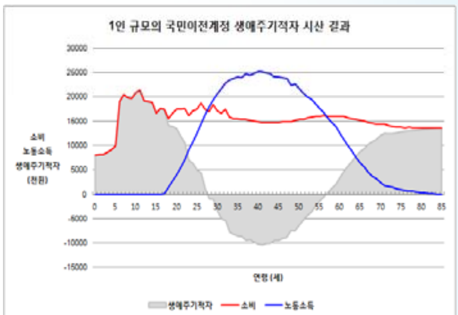
Statistical Research Institute 제6회 국가통계 발전포럼 – 국민이전계정(NTA) 개발 현황과 향후 계획 | 9

NTA 생애주기적자 부분 시산 결과

통계청 통계개발원

- **생애주기적자 부분 최종 시산 결과**
  - '생애주기적자(Lifecycle deficit or surplus, LCD)'는 '소비'에서 '노동소득'을 차감하여 계산
  - 유년층과 노년층에서는 생애주기적자가, 노동연령층에서는 생애주기흑자가 각각 발생
    - 기존 연구에서도 생애주기적자 구조는 유년층 적자-노동연령층 흑자-노년층 적자의 3단계로 나타남
    - 2012년 기준 : 유년층(0세~27세) 적자, 노동연령층(28세~56세) 흑자, 노년층(57세~) 적자 발생





Statistical Research Institute 제6회 국가통계 발전포럼 – 국민이전계정(NTA) 개발 현황과 향후 계획 | 10

NTA 공공연령재배분 부분 시산 결과

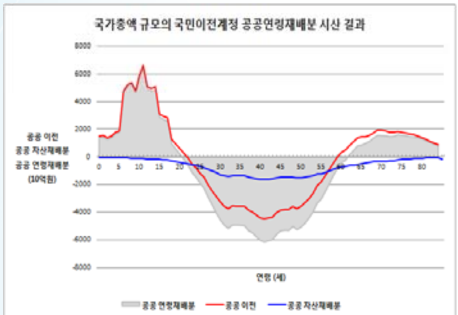
통계청
통계개발원

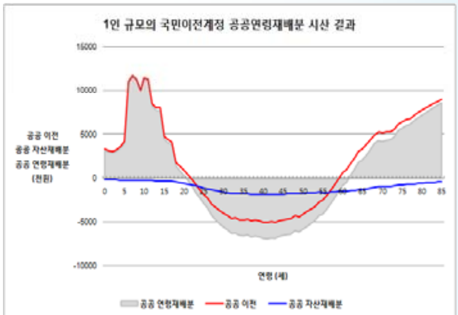
▪ 공공연령재배분 부분 최종 시산 결과

– '공공 연령재배분'은 '공공 이전'과 '공공 자산재배분'의 합계로 계산

– 우리나라의 경우, '공공 연령재배분' 구조는 '공공 이전' 구조와 유사한 구조를 가지고 있음

- 절대적인 규모 면에서 '공공 이전'이 '공공 자산재배분'보다 훨씬 큰 규모를 갖고 있음
  - '공공 (순)이전'은 유년층과 노년층에서는 순유입이, 노동연령층에서는 순유출이 각각 발생
  - '공공 자산재배분'은 모든 연령대에서 음(-)의 값을 가짐
- 2012년 기준 : 유년층(0세~20세) 순유입, 노동연령층(21세~61세) 순유출, 노년층(62세~) 순유입 발생





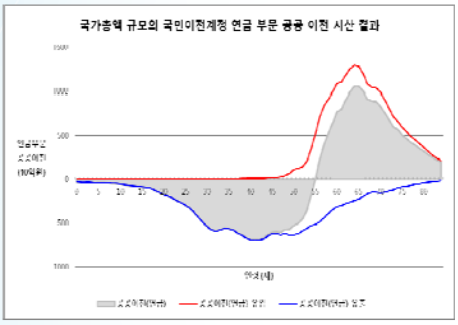
NTA 공공연령재배분 부분 시산 결과

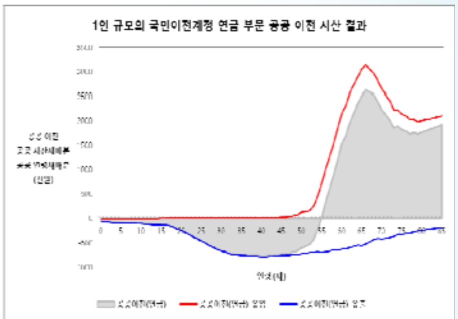
통계청
통계개발원

▪ 공공연령재배분 부분 최종 시산 결과

– 연금 부분 공공 이전 최종 시산 결과

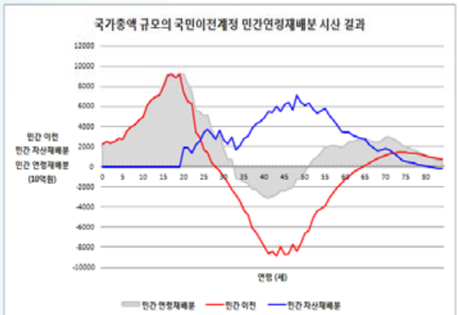
- '연금 부분 공공 이전'은 '연금 부분 공공 이전 유입'에서 '연금 부분 공공 이전 유출'을 차감하여 계산
- '연금 부분 공공 이전 (유입)'은 60세와 70세 사이의 연령층에 집중되어 발생
  - 70세 이후 연령대에서 '연금 부분 공공 이전 (유입)'이 감소되는 원인은 우리나라의 국민 연금 제도와 관련
    - » 국민연금제도 : 1988년 1월 시행, 1999년 4월 1일 가입대상자 범위 확대

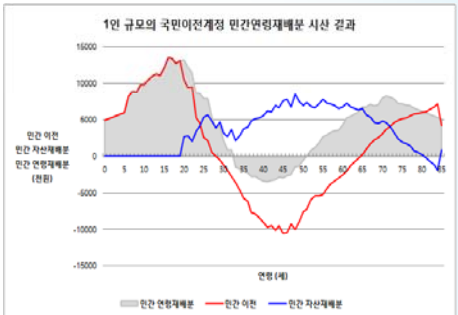




NTA 민간연령재배분 부분 시산 결과

- **민간연령재배분 부분 최종 시산 결과**
  - '민간 연령재배분'은 '민간 이전'과 '민간 자산재배분'의 합계로 계산
  - 우리나라의 경우, 민간연령재배분이 유년층에 집중하여 발생
    - '민간연령재배분'은 유년층과 노년층에서는 순유입이, 노동연령층에서는 순유출이 각각 발생
      - '민간 (순)이전' 은 유년층과 노년층에서는 순유입이, 노동연령층에서는 순유출이 각각 발생
      - '민간 자산재배분'은 '공공 자산재배분'의 경우와는 반대로 모든 연령대에서 양(+)의 값을 가짐
    - 2012년 기준 : 유년층(0세~32세) 순유입, 노동연령층(33세~50세) 순유출, 노년층(51세~) 순유입 발생





국민이전계정 최종 시산 결과

- **국민이전계정 최종 시산 결과 (2012년)**
  - 생애주기적자 = 연령재배분 = 공공연령재배분 + 민간연령재배분

연령	국가총액 규모 기준 (10억원)				1인 규모 기준 (천원)			
	생애주기 적자	연령재배분	공공 연령재배분	민간 연령재배분	생애주기 적자	연령재배분	공공 연령재배분	민간 연령재배분
0	3,655.59	3,655.59	1,408.47	2,247.1	8,021.79	8,112.23	3,090.74	4,931.05
1	3,911.96	3,911.96	1,434.75	2,477.2	8,112.23	8,112.23	2,975.25	5,136.98
…	…	…	…	…	…	…	…	…
15	10,925.49	10,925.49	2,898.39	8,027.1	16,613.94	16,613.94	4,407.46	12,206.48
16	11,822.66	11,822.66	2,701.16	9,121.5	17,528.68	17,528.68	4,004.83	13,523.85
17	11,835.43	11,835.43	2,588.85	9,246.6	17,226.74	17,226.74	3,768.14	13,458.60
…	…	…	…	…	…	…	…	…
53	-3,040.29	-3,040.29	-3,970.91	930.6	-3,741.40	-3,741.40	-4,886.63	1,145.23
54	-1,968.95	-1,968.95	-3,344.98	1376.0	-2,533.41	-2,533.41	-4,303.93	1,770.52
…	…	…	…	…	…	…	…	…
84	1,375.45	1,375.45	840.42	535.0	13,579.23	13,579.23	8,297.07	5,282.16
85+	5,917.27	5,917.27	3,729.29	2,188.0	13,585.86	13,585.86	8,562.33	5,023.53
합계	**135,995.44**	**135,995.44**	**-60,290.40**	**196,285.8**	-	-	-	-

맺음말

통계청 통계개발원

▪ 주요 결과 및 시사점

– 국민계정 기준년 개편(2008 SNA) 내용 반영 작업 완료

– 우리나라의 경우, 절대 규모 측면에서 '공공 이전(예. 교육 및 보건 분야의 정부 지원, 사회수혜금 및 사회적 현물이전)'이 '공공 자산재배분(예. 공공 이자, 공공 임대료)'보다 훨씬 큰 규모를 가짐

– 민간연령재배분은 민간 이전(가구간 이전+가구내 이전)을 중심으로 유년층에 집중하여 발생

– 국민이전계정은 경제적 생애주기 분석에 필요한 유용한 틀을 제공

• 경제 생애주기의 패턴과 생애 재분배의 형태에 따른 정책 시뮬레이션 가능

▪ 향후 연구 및 계획

– 국가통계위원회 상정(분과위원회 2016년 6월, 본회의 2016년 하반기 예정) : 국민이전계정 개발 계획(안)

– 국민이전계정 (시계열) 결과를 활용한 인구변화의 경제효과 분석 (2016년 하반기 국제공동연구)

– 국민계정 기준년 개편 결과 반영 및 국민이전계정 결과 세분화(성별, 지역별 생애주기적자계정 구축) (2016년중)

– 국민이전계정 작성 결과 및 분석 결과 제공 (2017년 이후)

향후 국민이전계정(NTA) 시산 결과 및 연구 결과를 공공 · 민간 · 학계에 정책 자료 및 연구보고서 형태로 제공
→ 관련 부처(기획재정부, 보건복지부 등) 및 민간·학계에서 NTA 결과가 활용(국가 예산, 연금, 보건, 복지 등의 분야)될 수 있도록 할 계획

Statistical Research Institute 제6회 국가통계 발전포럼 – 국민이전계정(NTA) 개발 현황과 향후 계획 ❙ 15

감사합니다

통계청 통계개발원

통계청 통계개발원 동향분석실
E-mail) khk144@korea.kr
Tel) 042.366.7304

시군구 GRDP 분배계정 개발과 활용

SNA 소득계정

소득발생	소득분배	소득재분배	현물소득 재분배	소득사용
• 피용자보수 • 생산및수입세 • 보조금(-) • 영업잉여/ 혼합소득	• 영업잉여/ 혼합소득 • 피용자보수 • 생산및수입세 • 보조금(-) • 재산소득 • 본원소득	• 경상이전 지급 및 수입 • 처분가능 소득	• 현물사회이전 지급 및 수입 • 조정처분가능소득	• 최종소비 지출 • 연금기금 조정 • 저축

분배계정

- **소득분배계정: 제도부문 및 소득유형별 본원소득 기록**
  - **(본원소득)** 생산과정 참여나 생산에 필요한 자산소유로 제도단위에게 일어나는 소득

- **지역소득(GRDP*) 분배계정**
  - 경상이전을 통한 소득 재분배를 나타내는 2차소득분배 계정을 포함하여 분배계정으로 편제

* Gross Regional Domestic Prodoct

- **시도 GRDP와 시군구 GRDP**
  - 시도 GRDP: 생산, 분배 ,지출계정을 포괄하여 생산
  - 시군구 GRDP: 생산계정만 편제 중으로 분배계정 필요

지역계정에서 분배계정

	소득발생계정	소득분배계정
특성	- 생산계정을 보다 정교화한 계정 - 정부단위와 생산에 직접 참여한 단위에게 발생한 소득 기록	- 본원소득의 제도부문별 분배 계정 - 소득발생계정에서 지급해야 할 항목이 수취되어질 곳을 보여주는데 제도단위에 의해 수취, 지급되어야 할 재산소득을 포함
대상	생산자로서 능력 있는 거주 제도단위	본원소득의 수혜자인 제도단위
집계	제도단위, 사업체, 산업별 가능	제도단위 또는 부문으로만 가능
거주	사업체(장) 소재지	제도단위 소재지
차이	집계 단위와 거주 차이로 생산과 차별 되어 분배계정 필요 → GRDP 완결성	
* 기업의 지사, 거주이전의 자유 등으로 지역소득의 유출입이 국가계정보다 大		

4

시도별 소득 유출입

	2013 (십억원)			2014 (십억원)		
	지역내 순생산	순본원 소득	소득 순유출입	지역내 순생산	순본원 소득	소득 순유출입
서 울	270,513	322,973	52,460	279,176	327,395	48,220
부 산	58,224	69,033	10,809	60,885	72,042	11,158
대 구	37,288	44,915	7,627	38,834	47,961	9,128
인 천	50,805	55,192	4,387	54,637	58,745	4,109
광 주	24,737	27,235	2,498	26,079	28,935	2,856
대 전	26,060	29,451	3,391	27,143	31,636	4,493
울 산	55,566	41,156	-14,411	54,269	39,059	-15,210
경 기	248,341	283,521	35,180	258,924	287,265	28,342
강 원	27,849	24,049	-3,800	29,301	25,411	-3,890
충 북	36,808	30,310	-6,498	38,736	30,755	-7,981
충 남	73,312	49,094	-24,218	78,777	51,935	-26,842
전 북	33,895	29,126	-4,769	35,287	31,125	-4,162
전 남	49,262	34,259	-15,003	49,086	34,971	-14,115
경 북	67,982	47,954	-20,028	70,577	53,441	-17,136
경 남	80,027	62,243	-17,784	81,300	66,690	-14,611
제 주	10,484	10,842	358	11,189	11,515	326

개발 배경

- 세부지역 단위 소득지표 부재
- 통계이용자의 요구에 부응
- 지자체 통계기반 강화 지원
- 시군구 지역계정의 완성 지향: ‘시군구 GRDP에서 시군구 지역계정으로’

기본방향

- 단계적 추진
  - (2016년) 통계청 주도 추계방안 도출, 15개 시도 시산·정제 및 매뉴얼 작성
  - (2017년) 교육 및 관련 기술지원 제공 등을 통한 시도 자체 작성 유도
- 공유 협업 지향
  - 개발과정 전반을 시도와 공유

필요 기초자료

소득유형	합계
본원소득계정	**54**
피용자보수	3
영업잉여	6
순생산및수입세	1
고정자본소모	6
재산소득	38
2차 소득분배계정	**85**
소득·부등에대한경상세	6
사회부담금	24
사회수혜금	22
기타경상이전	33
합 계	**139**

추계프로그램: 작업흐름

1. 기초자료 수집 (분담 수집)	**(통계청)** 통계청이 보유한 시도 공통 시군구 단위 자료 수집 **(시도)** 해당 시도 시군구 단위 세분가능 자료 수집
2. 시산 추계	**(통계청)** 15개 시도 기초자료를 취합하여 소득분배 계정 결과표 작성 및 검토
3. 시산결과 검토 (공동 검토)	**(통계청, 시도)** 시산된 결과를 통계청과 각 시도가 공동으로 검토하여 문제점 및 개선사항 도출
4. 보완, 정제	**(통계청)** 도출된 문제점을 보완하여 프로그램 정제
5. 추계프로그램 개발	**(통계청)** 추계프로그램 개발 보고

추계프로그램: 추계방안

- 하향식 추계원칙
  - 추계결과의 시군구 하향 배분(top-down) 및 시도별 동일 배분지표 적용 추계 원칙
- 추계방법 표준화
  - 추계결과의 정합성 확보 및 시도 간 추계방법 차이에 따른 편차 발생 최소화

추진일정

□ 개발 기본방향 시도 의견수렴 * 추계방향, 역할분담 및 추진일정 설명	: '16. 2월
□ 개발 기본계획 수립 및 보고	: '16. 3월
□ 시도 담당자 소득분배계정 이론 교육 * GRDP 시도 추계지원 1차 집합교육 시 실시	: '16. 5월
□ 기초자료 수집·발굴 및 시산	: '16. 5 ~ 8월
□ 시도 담당자 시산결과 공동 검토 * 기초자료 및 추계결과에 대한 개선의견 * GRDP 시도 추계지원 2차 집합교육 시 실시	: '16. 9월
□ 추계프로그램 수정·보완	: '16.10 ~11월
□ 추계프로그램 개발 완료 및 결과 보고	: '16.12월
□ 추계프로그램 교육 및 기술지원 등	: '17.

기대효과

- 기초자치단체의 합리적인 지역경제발전정책 수립 ·평가 기반 마련

  - 광역자치단체에 이어 기초자치단체에서도 생산활동은 물론 소득분배활동에 이르는 경제 활동에 대한 종합적인 평가가 가능

  - 지역개발정책에 따른 다각적인 지역 및 산업간 파급효과 분석 등 지역경제와 관련 연구활동을 위한 유용한 도구 제공

THANK YOU

금융위기 이후 자금순환통계의 확충

한국은행

김성준

2016.8.31

Overview

1. 금융위기 이후 자금순환통계의 중요성
2. 국제적 논의 흐름
3. 해외 사례
4. 향후 과제

금융위기 이후 자금순환통계의 중요성

In the context of financial crisis

- 금융위기 이후 국민경제 각 부문(sectors)의 **금융불균형(financial imbalances)**, **금융중개현황(financial intermediation)**을 보다 포괄적으로 파악하기 위한 통계의 필요성에 대한 인식 제고
- 금융거래내역에 관한 속보성 및 정확성을 겸비한 경제통계의 부재로 글로벌 금융위기를 유발할 수 있는 가능성이 사전에 제대로 포착되지 못했다는 견해
  - 특히 **부외거래(off-balance sheet transaction)**, **금융기관간 상호거래(financial interlinkage)** 및 **위험기반상품(risk-based instruments) 정보** 등에 관한 미시 데이터가 부족했다는 지적
- 국제기준(SNA)에 따라 모든 경제부분에 대해 가장 포괄적이고 일관성있는 금융거래현황을 보여줄 수 있는 자금순환통계의 확충 및 개선의 필요성이 높아짐

국제기구의 논의 현황

G20 DGI(Data Gaps Initiative)-개요

- G20 DGI 2nd stage에서 추진중인 자금순환통계관련 Action plans

1. Sectoral Accounts : From Whom to Whom
2. Securities Statistics : Issuance, holidings, from-whom-to-whom
3. Shadow Banking Statistics
4. Public Sector Debt Statistics

- DGI 2단계사업은 2016~2021년까지 5년간 진행되며 각 과제별로 Leading IAG의 주도하에 G20국가들이 과제를 이행

국제기구의 논의 현황

우리나라의 진행상황-개요

- From Whom to Whom : 08SNA 개편(2012년)시 금융상품별로 상세자금순환표(from whom to whom table) 편제가 가능하도록 편제시스템을 구축중
- Securities statistics : 예탁결제원의 Security by security 미시정보를 이용하여 해당 통계의 작성가능성을 점검, 작성방안 마련중
- Shadow banking statistics : 국제기구(FSB)의 기준에 맞추어 시험편제 중
- Public sector debt statistics : 현재 정부가 연간확정통계 작성, 공표중이며, 정부·한은 공동으로 분기잠정통계 작성방안 마련, 시험편제중

해외사례-미국

미국 : Enhanced Financial Accounts

1. information about which economic sectors or subsectors are borrowing from which other sectors, using what type of financial instruments, over what time frame, using what type of collateral(if any), and at what pricing.
2. information on "off-balance sheet" activities such as the use of financial derivatives to transfer market risk, credit risk, interest-rate risk, and exchange-rate risk.

해외사례-미국

미국 : Enhanced Financial Accounts

- FRB는 금융중개(financial intermediation)현황을 보다 상세하고 시의성있게 파악하기 위해 Financial Accounts of the United States를 업그레이드 하기 위한 **장기프로젝트(Enhanced Financial Accounts)**를 진행
- 기존의 일상적인 통계개선업무와의 차이는 **higher-frequency information, more-disaggregated data**를 제공하는 것
- 기존 Financial Accounts와 통합(integration)이 어려운 경우 **memo items, supplementary tables**를 사용하여 반드시 기존의 틀내에 통합되지 않더라고 유용성이 높다고 판단되는 경우 신규 작성

해외사례-미국

미국 : Enhanced Financial Accounts

- **Consolidated balance sheets**
  - US-chartered depository institutions이 해외에 보유하고 있는 자산/부채를 포함한 모든 금융자산/부채 현황(1990:1Q)
- **Syndicated loan portfolios of financial institutions**(total exposure to syndicated loans)
  - syndicated loan : a loan extended by a group of financial institutions(a loan syndicate) to a single borrower
  - syndicated loan은 은행, 비은행, 보험사, 연기금, 뮤추얼펀드 등이 참여하여 대출을 일으키고, 유동화된 이후 유통시장(secondary markets)에서 거래될 수 있음
  - syndicated loan은 개별대출금융기관의 대차대조표에 포함되지만 syndicated loan이라는 항목으로 기록되지는 않음.

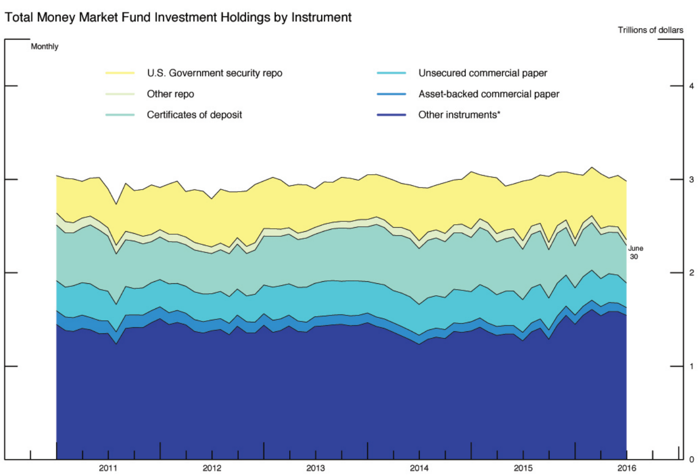
해외사례-미국

미국 : Enhanced Financial Accounts

- **Money Market Mutual Funds**: Investment Holdings Detail by Month
  - additional detail on the investment holdings of U.S. money market funds, based on a monthly dataset of security-level holdings for all U.S. money market funds
- **Off-balance-sheet items**
  - off-balance-sheet items은 contingent assets or liablilities(unused loan commitments, letters of credit, derivatives)을 의미하며, 대차대조표에 기록되지 않으나 credit risk, liquidity risk, counterparty risk가 있음 (1990:1Q)

해외사례-미국

미국 : Enhanced Financial Accounts



Source: SEC form N-MFP filings.
*Other instruments of money market funds include U.S. Treasuries, U.S. Agencies, and tax-exempt securities.

해외사례-미국

미국 : Enhanced Financial Accounts

Syndicated Loan Portfolio of Domestic Entities (1)
Billions of dollars; levels, not seasonally adjusted

		2010	2011	2012	2013	2014 Q1	2014 Q2	2014 Q3	2014 Q4	2015 Q1	2015 Q2	2015 Q3	2015 Q4	2016 Q1	
	Depository institutions (2)														
1	**Total credit amount outstanding (3)**	319.1	376.0	459.0	492.1	511.0	531.5	553.7	593.2	609.9	631.1	649.8	659.5	694.6	1
2	With missing default risk	96.3	112.9	13.2	29.3	29.7	25.4	29.0	30.2	31.4	36.3	34.6	27.6	33.4	2
3	With no default risk	13.5	7.7	33.4	28.5	28.1	30.9	29.8	33.7	33.1	33.6	31.7	38.2	54.0	3
4	With default risk between 0 percent and less than 0.5 percent	71.1	120.9	229.6	237.8	247.4	255.9	269.0	290.4	290.7	307.1	316.5	333.9	340.7	4
5	With default risk between 0.5 percent and less than 1 percent	44.4	44.1	60.7	94.3	96.2	105.9	104.3	113.1	121.7	114.8	119.0	105.8	106.0	5
6	With default risk between 1 percent and less than 2.5 percent	36.7	48.3	80.3	60.7	63.3	66.4	72.5	76.7	76.2	78.3	80.8	83.7	76.6	6
7	With default risk between 2.5 percent and less than 5 percent	21.9	17.7	11.8	19.1	23.7	22.0	21.4	24.5	27.9	28.0	28.6	25.1	23.6	7
8	With default risk between 5 percent and less than 10 percent	5.7	8.9	13.7	8.4	9.6	11.1	11.9	10.1	12.0	11.9	11.5	12.1	13.2	8
9	With default risk between 10 percent and less than 25 percent	8.0	4.4	8.3	8.2	9.8	10.6	9.6	8.5	10.9	13.0	16.1	21.7	30.4	9
10	With default risk between 25 percent and less than 100 percent	5.7	3.8	2.9	0.5	0.4	0.6	0.8	0.5	0.6	0.6	0.5	1.2	1.6	10
11	In default	15.7	7.2	5.1	5.3	2.9	2.6	5.3	5.5	5.5	7.4	10.6	10.1	15.0	11
	Other financial institutions														
12	**Total credit amount outstanding (3)**	556.5	566.4	565.9	644.4	661.5	665.2	684.3	708.2	716.1	718.1	723.1	728.3	691.6	12
13	With missing default risk	134.6	203.4	160.7	185.8	195.1	195.8	206.0	205.0	208.9	205.8	209.0	163.5	153.1	13
14	With no default risk	155.4	79.3	144.6	170.7	160.2	161.0	167.4	168.9	165.5	155.4	156.8	157.8	159.8	14
15	With default risk between 0 percent and less than 0.5 percent	52.5	69.7	90.0	93.0	112.6	117.9	125.0	129.2	129.2	131.9	140.6	154.4	157.3	15
16	With default risk between 0.5 percent and less than 1 percent	34.4	26.7	23.6	34.3	31.4	32.5	33.1	38.6	41.5	37.8	41.8	48.4	45.1	16
17	With default risk between 1 percent and less than 2.5 percent	41.8	51.9	47.8	44.8	52.7	63.3	59.5	59.1	53.5	57.8	65.2	74.8	53.8	17
18	With default risk between 2.5 percent and less than 5 percent	35.0	34.4	32.8	27.3	33.2	36.2	36.5	44.4	51.9	53.7	46.8	63.0	47.0	18
19	With default risk between 5 percent and less than 10 percent	14.9	27.0	22.7	26.9	26.7	27.9	25.4	24.7	21.7	25.0	19.9	21.6	20.9	19
20	With default risk between 10 percent and less than 25 percent	22.3	11.0	12.7	9.8	12.1	14.4	13.2	14.1	13.7	17.6	17.7	19.6	20.7	20
21	With default risk between 25 percent and less than 100 percent	5.5	3.3	5.2	3.5	3.2	4.5	4.5	3.1	1.9	1.2	2.4	1.5	2.1	21
22	In default	60.2	59.6	25.8	48.2	34.3	11.8	13.7	21.2	28.3	31.9	22.9	23.7	31.7	22

(1) Default risk is the two-year through-the-cycle probability of default provided by the reporting financial institution.
(2) Includes bank holding companies, financial holding companies, national banks, nonmember banks, state member banks, federal savings banks, state savings banks, credit unions, and savings and loan associations.
(3) Outstanding amounts on term loans, drawn revolving lines of credit, and other loans.

해외사례-영국

영국 : The UK Flow of Funds Project(Comprehensive reivew of the UK Financial Accounts)

- 2014년 하반기부터 **Bank of England와 Office for National Statistics가 공동**으로 진행
- 주요 목표는 금융자산, 부채의 거래 및 잔액을 **from-whom-to-whom** 기반으로 작성하는 것
- 2015.7월 1997-2014까지 주요 금융상품에 대해 from-whom-to-whom 시험통계(experimental estimates)를 작성하여 발표

해외사례-영국

영국 : The UK Flow of Funds Project(Comprehensive reivew of the UK Financial Accounts)

- Instruments
  - currency, deposits
  - short-term debt securities, long-term debt securities
  - equity and investment funds shares/units
  - insurance, pension and standardized guarantee schemes
  - financial derivatives

해외사례-영국

Table 3: A from-whom-to-whom balance sheet matrix for total deposits, 2014 unconsolidated current prices £ billion

Liabilities:

Assets:	PC	PNFC	MFI	OFI	ICPF	CG	LG	HH+ NPISH	RoW	Total
PC			2.8			1.8			0.0	4.6
PNFC			357.7			3.3			144.8	505.8
MFI			916.8			0.0			1,620.5	2,537.4
OFI			697.4			17.4			680.7	1,395.6
ICPF			94.7			0.0			52.6	147.3
CG			18.1			2.7			7.0	27.8
LG			25.9			1.0			1.0	27.9
HH+NPISH			1,169.3			113.0			72.0	1,354.3
RoW			2,678.2			1.3				2,679.5
Total			5,961.0			140.5			2,578.6	

해외사례-영국

Table 5: A from-whom-to-whom balance sheet matrix for total long-term debt securities, 2014 unconsolidated current prices £billion

Liabilities:

Assets:	PC[a]	PNFC	MFI	OFI	ICPF	CG	LG	HH+ NPISH	RoW	Total	*Unknown c/p*
PC	-	-	-	-	-	0.5	0.0	-	0.1	0.6	*0.0*
PNFC	-	-	1.2	-	-	2.8	0.0	-	36.2	71.1	*31.0*
MFI	0.1	-	18.9	-	-	562.9	0.0	-	416.5	1,285.6	*287.2*
OFI	20.7	0.0	16.1	-	-	124.1	0.0	-	583.6	1,025.5	*281.0*
ICPF	5.0	-	44.8	-	-	476.0	2.7	-	317.2	1,080.1	*234.4*
CG	0.0	-	0.0	-	-		0.0	-	45.3	46.1	*0.8*
LG	-	-	-	-	-	1.1		-		1.3	*0.2*
HH+NPISH	0.0	-	0.0	-	-	76.0	1.5	-	7.9	92.2	*6.8*
RoW	-	30.2	-	13.6	3.0	441.3	0	-		1,471.8	*983.6*
Total	25.7	309.9	688.9	923.8	29.2	1,684.8	4.2	0.9	1,406.9		
Unknown c/p	*0.0*	*279.7*	*608.0*	*910.2*	*26.2*	*0.0*	*0.0*	*0.9*	*0.0*		

향후 과제

G20 DGI(Data Gaps Initiative)

1. Sectoral Accounts : From Whom to Whom
2. Securities Statistics : Issuance, holidings, from-whom-to-whom
3. Shadow Banking Statistics
4. Public Sector Debt Statistics

향후 과제

Sectoral Accounts-G20이행과제

- 2021년까지 제도부분별 비금융(경상 및 자본)계정, 금융계정(거래 및 잔액) 통계를 국제적으로 합의된 양식에 맞추어 연간데이터는 9개월이내에, 분기데이터는 4개월이내에 편제 및 공표할 것
- Intermediate steps and timelines : G20 국가별로 부문통계 편제의 발전단계가 상이하므로 2016-2021년에 중간 스텝과 시한에 대한 논의가 토론될 것임
- 채권자-채무자 연계통계의 구체적 템플릿을 발전시키기 위한 준비단계로 각국은 IMF 표준양식(Standardized Reports Forms)에 따라 국가별로 가능한 데이터를 거래 및 잔액에 대해 공표할 것

향후 과제

Securities Statistics-G20이행과제

- 2021년까지 각국은 발행섹터, 통화, 만기, 이자유형, 발행시장별로 구분하여, 분기별로 4개월이내에 HSS에 맞추어 명목 및 시장가격으로 거래액 및 잔액에 대해 채권통계를 보고할 것
- 우선순위(first priority)는 2018년까지 BIS에 다음과 같은 통계를 제공할 것
  - 4개 부문별 : 기업, 금융, 정부, 가계및비영리단체
  - 통화별 : 자국통화, 외화
  - 발행만기별 : 단기, 장기
- 2021년까지 완성된 데이터를 보고할 것

향후 과제

Shadow Banking Statistics-G20이행과제

- 2016년에 개선될 예정인 FSB template에 맞추어 「Global Shadow Banking Monitoring Report」 작성을 위한 그림자금융 데이터를 FSB에 제출할 것
- 2018년말까지 SFT 통계를 FSB 기준에 맞추어 FSB에 제출할 것
- OECD에서 금융부문에 대한 추가적인 세분화가 결정되면 각국은 이에 맞추어 자료를 제공

향후 과제

Public Sector Debt Statistics-G20이행과제

- World Bank/IMF/OECD Public Sector Debt Database의 상품포괄범위를 기준으로 분기별로 일반정부 및 중앙정부의 부채통계를 보고할 것
- Intermediate steps : 중요상품(대출, 예금, 채권)에 대한 중간 목표 설정에 관한 논의가 있을 것임

향후 과제

Sectoral Accounts

- Instruments
- Sectors
- Data sources
- Valuation method
- Consolidation

향후 과제

Sectoral Accounts

현행 자금순환표

		제도부문									
		금융		정부		비금융		가계		국외	
		자산	부채	자산	부채	자산	부채	자산	부채	자산	부채
금융상품유형	금 및 SDR										
	현금 및 예금										
	보험 및 연금										
	채권										
	대출금										
	주식 및 출자지분										
	파생금융상품										
	기타금융자산부채										

상세 자금순환표

			채권자				
			금융	정부	비금융	가계	국외
채무자	현금 및 예금	금융					
		정부					
		⋮					
		국외					
	채권	금융					
		정부					
		⋮					
		국외					
	⋮						

향후 과제

Sectoral Accounts

상세자금순환표 편제 범위

금융상품

	구분
현금 및 예금	현금통화
	결제성예금
	단기저축성예금
	장기저축성예금
	표지어음
	양도성예금증서(CD)
	환매조건부채권(RP)
	금전신탁
	기타예금
대출금	단기대출금
	장기대출금
	정부융자
	콜론·콜머니
채권	단기채권
	장기채권
	기업어음(CP)
	파생결합증권
기타	금과 SDRs
	비거주자발행주식
	투자펀드지분
	직접투자
	기타대외채권채무

부문분류

대	중	소
금융 법인	중앙은행	
	예금취급기관	은행
		비은행
	MMF	
	Non-MMF	
	보험회사	생명보험
		손해보험
		재보험
		공제기관
		표준보증기구
	연금기금	
	기타 금융중개기관	증권기관
		여신전문기관
		공적금융기관
	금융보조기관	
	지주회사, 유동화회사, 대부업자	지주회사
		유동화회사
		대부업자
일반정부		
비금융법인		공기업
		민간기업
가계 및 비영리단체		
국외		

향후 과제

Securities Statistics : Issuance, holidings, from-whom-to-whom

- 채권통계 이행과제는 크게 ① **발행통계(Debt Issuance Statistics)**, ② **보유통계(Debt Holding Statistics)**, ③ **채권-채무통계 (From-Whom-to-Whom Debt Statistics)**로 구성
- 채권통계에 포함되는 내용(main features)은 ①가격(명목, 시장가격), ②만기(발행 및 잔존만기), ③금리유형(변동 또는 고정금리), ④발행통화(자국통화 또는 외국통화) 등임
- 채권통계는 잔액(position) 및 거래액(flow)을 모두 작성하며, 거래액은 순거래(transaction), 평가에 따른 변동(revaluations), 기타변동으로 구분
- 섹터(sectors)별 구분은 기업, 금융기관(중앙은행, 예취기관, MMF, 기타, 보험및연금), 일반정부, 가계및비영리단체 등 총 13개 섹터임

향후 과제
Shadow Banking Statistics

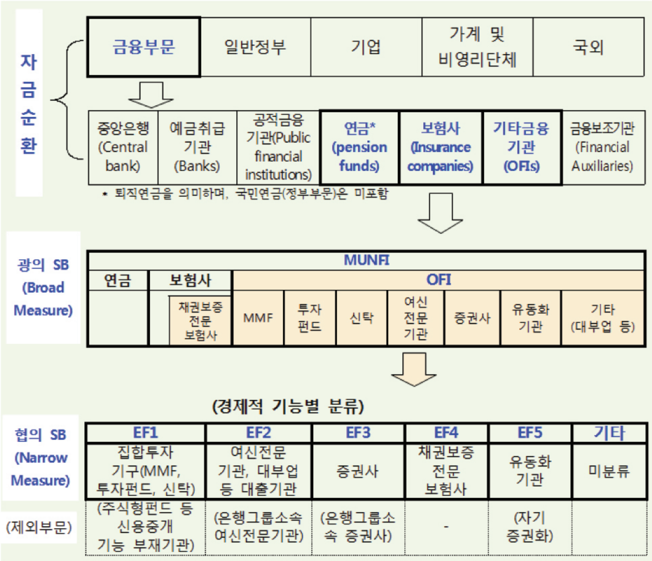
- Definition(FSB) : credit intermediation involving entities and activities outside of the regular banking system
- 일반적으로 감독당국의 규제내에 있는 예금취급기관(은행+ 비은행예금취급기관(신협, 상호금융 등))을 은행시스템내 기관으로 간주
- OFI : Other Financial Institutions
  - 일반적인 광의 SB 개념으로 이용되는 OFI에는 증권사, 투자펀드, 신탁, 여전사, 유동화기관 등이 포함되어 있음
- MUNFI (Monitoring Universe of Nonbank Financial Intermediation)는 OFI에다 보험 및 연금이 포함된 그림자금융 모니터링집단
  - 보험 및 연금이 광의SB에 포함된 이유는 보험기관 중에서 모노라인(Monoline, 채권보증전문 보험사) 등은 보증기능을 통해 비은행의 신용중개활동을 원활히 하는 SB활동을 담당하기 때문

향후 과제
Shadow Banking Statistics

- **광의SB** :「은행시스템 밖에서 은행과 유사한 신용중개기능을 수행하는 기관 또는 활동」으로 정의되며, 기타금융기관(이하 OFI) 또는 SB모니터링집단(이하 MUNFI)을 광의 SB로 선정
- **협의 SB** : 「만기 및 유동성 변환, 불완전한 신용리스크의 이전, 레버리지 확대 등을 통해 시스템적 리스크를 유발하거나 규제차익을 추구하는 비은행 신용중개활동」으로 정의되며 광의SB 중에서 해당되는 기관(또는 활동)을 엄밀한 기준으로 간추려(Risk-focused narrowing down) 협의 SB를 선정
  - 광의SB를 경제적 기능(Economic Functions)에 따라 5개로 분류하고 (EF1~EF5) 해당기관(또는 활동)중 시스템 위험 발생과 무관한 기관(또는 활동)을 제거함으로써 협의 SB를 선정

향후 과제

Shadow Banking Statistics



향후 과제

Public Sector Debt Statistics

- **부채(total debt liabilities)**는 미래 특정시점에 상환해야하는 원금 및 이자로 정의되며, 총부채(total liabilities)중 지분증권 및 투자펀드, 파생상품을 제외
  - 지분증권, 파생상품은 부채항목에서 제외되는 데, 전자는 상환의무가 없기 때문이며, 후자는 원금 및 이자에 해당되지 않기 때문임
- 공공부문(public sector)이란 정부나 공기업이 직간접적으로 **지배(control)**하는 모든 제도단위(institutional unit)이며, 이에 따라 일반정부(중앙정부, 지방정부, 사회보장기구), 공기업(금융공기업, 비금융공기업)이 해당
  - 지배성(control)의 판단기준은 ① 과반수 이상의 의결권보유, ② 이사회 혹은 주요 협의체 기구에 대한 통제, ③ 집행간부 등 주요 인사에 대한 임명권 혹은 파면권 보유 등
- 정부보증채무(Publicly guaranteed debt), 우발채무, 국민연금 충당부채(Nondebt liabilities, explicit contingent liabilities, and net obligations for future social security benefits) 등은 제외

Thank you

Thank you

2016년 국가통계발전포럼
2016. 8. 31(수)

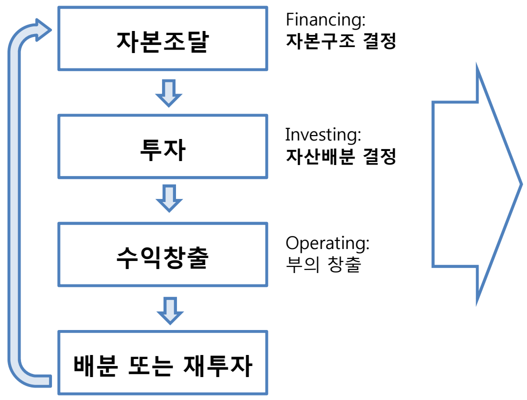
기업경영분석 분위수 통계 작성배경 및 결과

한국은행 기업통계팀
과장 김혜림

차 례

1. 한국은행 기업경영분석 통계

(1) 기업의 의사결정과 집행결과를 보여주는 재무제표



- 재무상태표(Balance sheet)
- 손익계산서(Income Statement)
- 현금흐름표(Statement of Cash Flow)
- 주석(Footnote) 등

(2) 기업경영분석 통계란?

우리나라
개별기업의 재무제표를
종합하여 만든
업종별 / 기업규모별
재무제표와
이를 바탕으로 산출한
각종 재무비율

1962년 첫 발간

2014년 책자

5

(3) 기업경영분석 통계 편제 현황 - 1

❖ 조사대상

국세청 법인세 신고기업중 금융보험업 이외의 영리법인, 다만 통계의 성격상 작성대상으로 부적합한 일부 업종과 결산월 1~5월 법인, 공정거래위원회 지정 지주사 제외

동일한 경영환경 시점의 문제

❖ 이용자료

국세청 법인세 신고자료에 첨부된 조사대상 법인의 재무제표

덕분에 2011년부터 전수집계조사가 가능해짐
기업의 응답부담 ↓ 조사비용↓ 정확성↑

6

(3) 기업경영분석 통계 편제 현황 -2

❖ 작성방법

조사대상 각 법인의 재무제표를 편제대상 업종별, 규모별로 분류하여
합산 재무제표를 집계하고 이를 기초로 업종별, 규모별 재무비율을 산출

✓ 더하기 먼저, 나누기는 나중에

- 업종분류 : 제9차 한국표준산업분류 (전산업 등 139개 분류)
- 규모분류 : 법인세 신고시 중소기업으로 신고한 법인을 중소기업으로, 그 이외는 모두 대기업으로 분류

- 매출액, 자산총액 등 기업규모가 일정수준보다 작을 것
- 상호출자제한기업집단(대기업집단)에 속하지 않을 것
- 세제혜택 지원 등 조세목적상 도박유흥업 등 일부 업종 제외

7

2. 문제의 제기

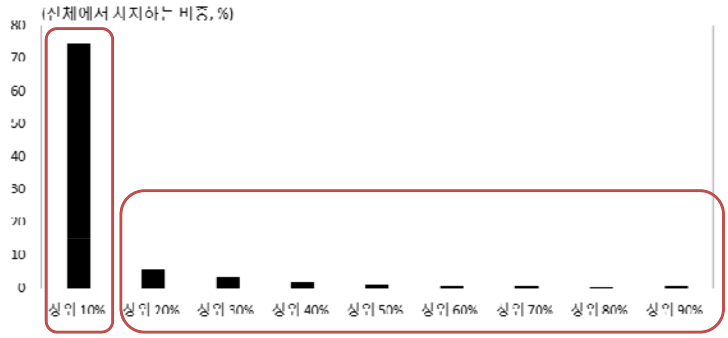
(1) 문제의 제기 : 미시정보, 총량지표에 감춰지는 사실들 1

매출액 상위 10%가 전체의 약 75%를 차지

(자료: 국세청 국세통계)

개별 기업 미시정보가 총량지표에 반영되기 어려움

법인세 신고기업의 매출분위별 매출점유율

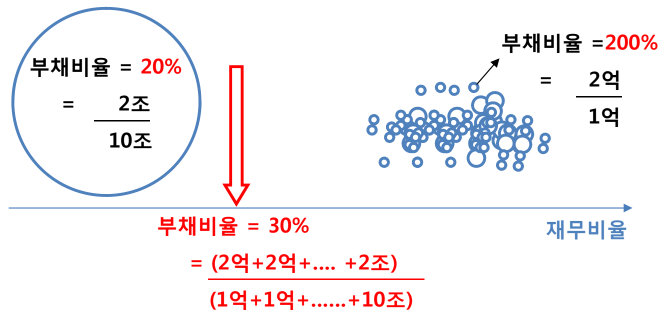


9

(1) 문제의 제기 : 미시정보, 총량지표에 감춰지는 사실들 2

대규모 기업의 특이 실적은 합산 재무비율을 기초로 작성한 총량지표를 크게 좌우

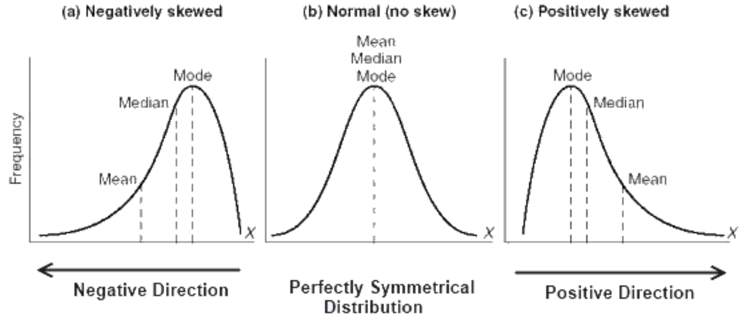
< 예 시 >



10

(2) 좋은 대표값에 대한 고민 1 : 평균, 중앙값, 최빈값

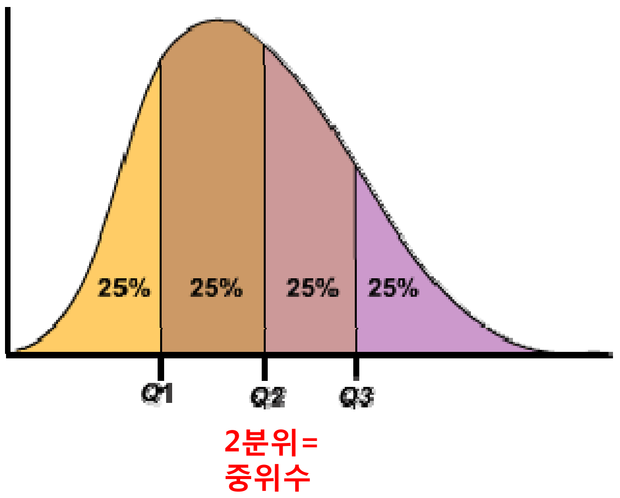
완벽한 대칭 종모양 분포
평균 = 중앙값 = 최빈값



꼬리가 긴 분포에서는
평균 ≠ 중앙값 ≠ 최빈값

(2) 좋은 대표값에 대한 고민 2: 사분위수의 개념

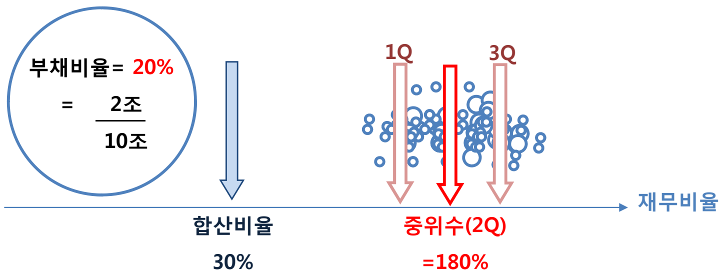
사분위수는 전체 정렬된 데이터를 네 개 그룹으로 나누는 세 포인트



1사분위수(1Q)는 25번째 백분위, 2사분위수(2Q, 중앙값)은 50번째 백분위, 3사분위수(3Q)는 75번째 백분위

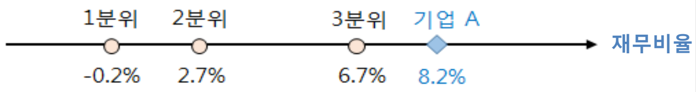
(2) 좋은 대표값에 대한 고민 3: 사분위수의 장점

1.
분위수에서는
상대적 가중치가
모두 동일하여
대기업의 극단치가
미치는 영향 축소



2.
분포에 대한 정보도 제공

예시 : 매출액영업이익률 사분위수



☞ "동 업종에서 적어도 25%의 업체들이 적자 영업을 하고 있으며, 25%의 업체들이 6.7% 보다 높은 영업이익률을 기록하고 있음. 기업 A는 동종업계 75% 업체들보다 높은 영업이익률을 기록"

13

3. 기업경영분석 분위수 통계 개발 결과

(1) 기업경영분석 분위수 통계의 개발

2015. 10. 27일 '2014년 기업경영분석' 을 발표하면서 분위수 통계를 참고지표로 함께 공표

- 개발배경 : 합산 재무제표에 나타나기 어려운 다수 기업들의 경영성과와 분포에 관한 정보를 제공
- 작성개요 : 편제대상 기업들의 재무비율을 각 재무비율 별로 오름차순 정렬 후 25%(1분위), 50%(2분위, 중위수), 75%(3분위)에 해당하는 값을 산출
- 대상업종 : 제조업 및 제조업 세부업종
- 대상기업 : 당해년 기업경영분석 통계 편제대상 기업 중 당기 및 전기의 총자산 또는 매출액이 0보다 큰 기업

(2) 제조업 사분위수 통계

제조업 주요 재무비율

(%, %p)

		사분위수			합산 재무비율		
		FY2013	FY2014	변화	FY2013	FY2014	변화
매출액증가율	1분위	-15.6	-16.0	-0.4			
	2분위(중위수)	**3.2**	**2.5**	**-0.8**	**0.5**	**-1.6**	**-2.1**
	3분위	36.3	32.8	-3.5			
매출액영업이익률	1분위	-0.7	-1.4	-0.7			
	2분위(중위수)	**3.9**	**3.8**	**-0.1**	**5.3**	**4.2**	**-1.1**
	3분위	7.7	7.7	0.1			
부채비율	1분위	71.0	68.6	-2.4			
	2분위(중위수)	**176.4**	**174.8**	**-1.7**	**92.9**	**89.2**	**-3.7**
	3분위	447.0	456.2	9.2			

(3-1) 제조업 사분위수 통계 : 매출액증가율

(%, %p)

	사분위수			합산 재무비율		
	FY2013	FY2014	변화	FY2013	FY2014	변화
1분위	-15.6	-16.0	-0.4			
2분위(중위수)	**3.2**	**2.5**	**-0.8**	**0.5**	**-1.6**	**-2.1**
3분위	36.3	32.8	-3.5			

- **매출액증가율 오름차순 정렬 후 25%, 50%, 75%에 해당하는 값**
- **2014년 제조업 매출액증가율 중위수 = 2.5% vs. 평균비율 -1.6%**

  ⇒ 큰 기업이 평균비율을 아래로 끌어당김
- **중위수와 평균 : 전년대비 변화 방향은 같지만 중위수가 하락폭 작음**

17

(3-2) 제조업 사분위수 통계 : 매출액영업이익률

(%, %p)

	사분위수			합산 재무비율		
	FY2013	FY2014	변화	FY2013	FY2014	변화
1분위	-0.7	-1.4	-0.7			
2분위(중위수)	**3.9**	**3.8**	**-0.1**	**5.3**	**4.2**	**-1.1**
3분위	7.7	7.7	0.1			

- **매출액영업이익률 오름차순 정렬 후 25%, 50%, 75%에 해당하는 값**
- **2014년 제조업 매출액영업이익률 중위수 = 3.8% vs. 평균비율 4.2%**

  ⇒ 큰 기업이 평균비율을 위로 끌어당김
- **중위수와 평균 : 전년대비 변화 방향은 같지만 중위수가 하락폭 작음**
- **1분위 -0.7 → -1.4% : 25% 커트라인이 더 낮아짐 → 적자기업 상황 악화**

18

(3-3) 제조업 사분위수 통계 : 부채비율

(%, %p)

	사분위수			합산 재무비율		
	FY2013	FY2014	변화	FY2013	FY2014	변화
1분위	71.0	68.6	-2.4			
2분위(중위수)	**176.4**	**174.8**	**-1.7**	**92.9**	**89.2**	**-3.7**
3분위	447.0	456.2	9.2			

- **부채비율 오름차순 정렬 후 25%, 50%, 75%에 해당하는 값**
- **2014년 제조업 부채비율 중위수 = 174.8% vs. 평균비율 89.2%**

  ⇒ 큰 기업이 평균비율을 아래로 끌어당김
- **중위수와 평균 : 전년대비 변화 방향은 같지만 중위수가 하락폭 작음**
- **1분위와 3분위의 거리(분위수격차) 확대**

19

감사합니다.

제 3 세 션

통계생산 패러다임 변화

- 통계의 활용도 제고를 위한 통계사회의 노력 / 홍은표 교수
- 공유경제와 통계 / 조윤정 교수
- 인구주택총조사의 패러다임 변화 / 전광희 교수
- 기업등록부(BR) 구축 및 활용 / 안은진 사무관

통계의 활용도 제고를 위한 통계사회의 노력

국가통계발전포럼
"통계생산 패러다임 변화"

2016년 8월31일

홍은표
상명대학교 금융통계학과
국가통계위원회 1분과 위원

발표 순서

1. 목적 및 배경
2. 글로벌 공식통계의 탄생 및 진화
3. 통계의 정책친화성 제고

1. 배경 및 목적

•배경: 2차세계대전이 종전되어 가며 통계작성을 위한 글로벌 기준이 마련되기 시작했고, 그 이후 통계 활용도를 제고하기 위한 노력을 지속적으로 해 오고 있음

•글로벌 공식통계의 진화와 관련한 글로벌 통계사회의 노력에 대해 살펴보고, 시사점을 발굴

2. 글로벌 공식통계의 탄생 및 진화(1/2)

•IMF

- •Balance of Payments Manual (1948), 5차례에 걸쳐 개정 1950, 1961, 1977, 1993 및 2008

•UN

- • Index Numbers of Industrial Production (1950), 2010년에 개정 (International Recommendations for the Index of Industrial Production 2010)
- •System of National Accounts (1953), 3차례에 걸쳐 개정 1968, 1993 and 2008 (1993 SNA and 2008 SNA 는 UNSD, WB, OECD, Eurostat 및 IMF 등 5개 기관이 함께 참여)

2. 글로벌 공식통계의 탄생 및 진화(2/2)

- ECB 및 Eurostat
  - European System of Accounts (1970), 2차례에 걸쳐 개정 1995 and 2010
  - Harmonized Indices of Consumer Prices: Their Conceptual Foundations (2002) EU국가의 물가수준을 포괄적으로 파악
- OECD
  - Directives for Reporting Aid and Resource Flows to Developing Countries (1962) and ODA (1969, 1972)
  - Programme for International Student Assessment (PISA, 2000)
  - PISA for Development (2019?)
  - Index of Services Production (2007)

3. 통계의 정책친화성 제고

- 목적 : 기존통계 한계 보완 및 새로운 경제사회환경에 적합한 통계 개발 및 작성
- 예 : Beyond GDP (GDP), Trade in Value Added (Total Trade), CPA (ODA), 생산성측정/실제근로시간 등

- 방향
  - ✓ 기존통계의 한계 보완
  - ✓ 완전히 색다른 통계 작성 : Beyond GDP (예 Well being index)
  - ✓ 기존통계를 활용하여 작성 :
    - ➢ Trade in Value Added (Total Trade): 글로벌 가치사슬 반영
    - ➢ CPA (ODA): 성과개선
    - ➢ 실제근로시간 : 생산성 측정의 정도 제고

3-1. Beyond GDP(1/2): 배경 및 현황

- 목적 : 경제적인 성과와 사회발전의 측정
- Measurement of Economic Performance and Social Progress 위원회: 프랑스 Sarkozy 대통령의 제안으로 결성되었고, INSEE가 첫 번째 미팅을 2008년에 개최 http://www.insee.fr/fr/publications-et-services/dossiers_web/stiglitz/doc-commission/Minutes_first_plenary_session_Commission_Measurement_Economic_Performance_Social_Progress.pdf

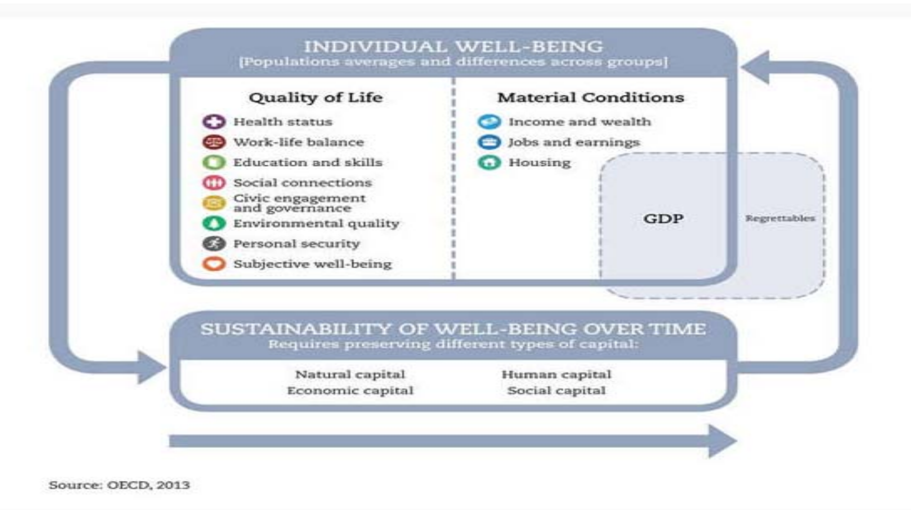
- Stiglitz-Sen-Fitoussi Commission (2009)가 작성한 리포트가 well-being을 측정하는 중요한 기폭제로서의 역할. 이후 개별국가과 국제사회가 GDP의 대안을 마련하기 위해 노력함.

- 고위급전문가 그룹(High Level Expert Group) : 통계정책가(Martine Durand, OECD 통계국장 ; Jil Matheson, UK 통계청장; Walter Radermacher, Eurostat국장), 유럽 및 미국의 대학교수(Joseph E. Stiglitz 콜롬비아대 교수, Jean-Paul Fitoussi 프랑스 씨앙스포대 교수, Angus S. Deaton 프린스톤대 교수, Thomas Piketty PSE 교수 등) 등이 포함되어 2013년에 결성
- 현재 HLE Group은 OECD가 사무국이 되어 다음과 같은 작업을 하고 있음
- 웰빙의 측정에 관한 국제프로젝트(OECD 작업 포함)에 대한 검토
- 지표와 경제이론과의 연결
- 불평등/지속가능성 등에 대한 분석작업
- 성과 및 부족한 부분에 대한 검토

3-1. Beyond GDP(2/2): OECD의 How's life Index

- 목적 : 웰빙과 발전의 측정
- OECD는 지난 10여년의 노력을 통해 2011년에 사회발전의 측정을 위한 Better Life Initiative 를 작성하여 제안함(OECD 설립 50주년을 기념).

Well-being과 Progress 측정을 위한 OECD의 프레임워크



3-2. Trade in Value Added (1/4): 배경 및 현황

- 목적 : 가속되고 복잡해 가는 국제무역의 실체(예 : 누가 무엇을 만들고 누가 무엇을 소비하는지?)를 파악하기 위함

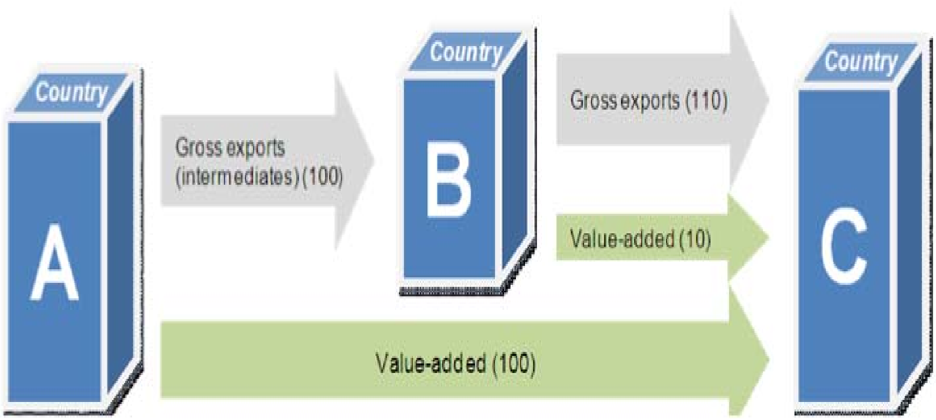
- 배경 : 1995년이후 보다 가속화되고 복잡해진 Global Value Chain으로 인해 누가 무엇을 생산하고, 누가 무엇을 소비하는지에 대한 정확한 정보가 없어 국제무역의 위험성이 보다 커지고 있음

- OECD (STD/DSTI)와 WTO가 공동작업을 하여 시안을 작성하여 2013년 1월16일 에 공표

- 지속적으로 개선하기 위한 노력을 하고 있음

3-2. Trade in Value Added (TiVA, 2/4)

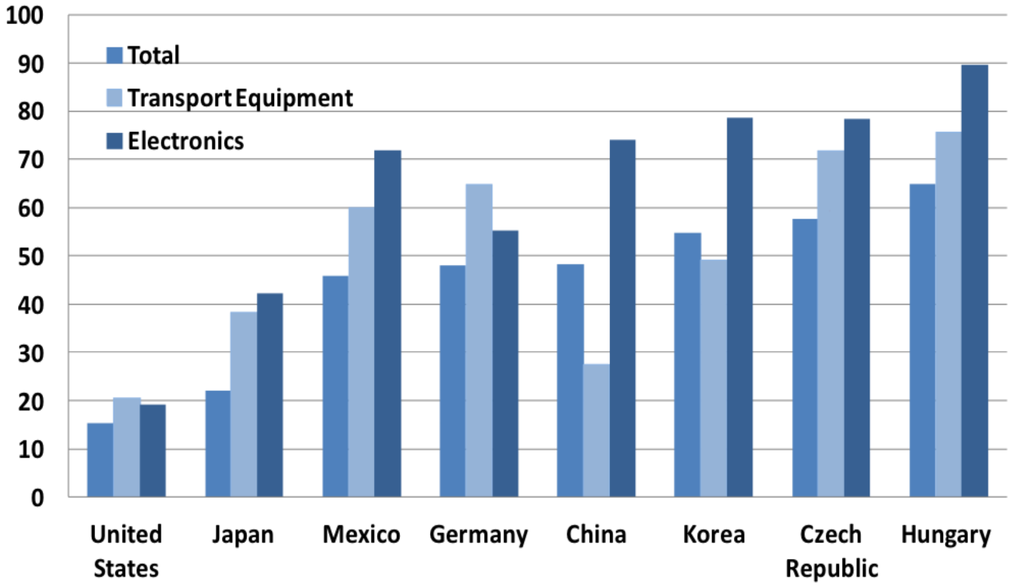
What is TiVA?



Trade in Value Added (3/4)

수출에 포함된 수입중간재, 2009

% of total intermediate imports



3-2. Trade in Value Added (4/4)

현존하는 다양한 정보를 활용해서 새로운 통계작성: **TEC => TIVA**

TEC (Trade by Enterprise Characteristics)

Year	.....		Statistical Units: Thousands															
Ownership			Total										Foreign owned			Domestic owned		
			Exports as per cent of Output strata										Output strata			Output strata		
Flow:	Industry	Size Class	0%	more than 0%	Between 0% and less than and equal to 25%	Between more than 25% and less than and equal to 50%	Between more than 50% and less than and equal to 75%	More than 75%	Less than and equal to 50%	More than 50%	*Unknown*	Total	0%	more than 0%	Between 0% and less than and equal to 25%	0%	more than 0%	Between 0% and less than and equal to 25%
	A (01-03)	Total																
		0-9																
		10-49																
		50-249																
		250+																
		Unknown																
	Unknown	*Total*																
		0-9																
		10-49																
		50-249																
		250+																
		Unknown																
Total																		

3-3. 통계청의 BR 작성 : 행정/조사자료 통합

- 국세청 : 사업자. 부가가치세, 법인세, 사업소득,
  근로소득지급명세
- 한국고용정보원 : 고용보험
- 산업통상자원부 : 공장등록
- 관세청 : 수출입무역통계
- 통계청 : 조사자료

4. 시사점 및 결론

- 글로벌사회에서 통계는 단순한 숫자가 아니라 정책

  "Statistics is not a number , but a policy!"

  (OECD Observer 2010?)

- (글로벌) 통계사회는 통계의 정책활용도를 제고하기 위해 지속적으로 노력해 오고 있음
  - ✓ 기존 통계의 보완
  - ✓ 새로운 통계 개발
  - ✓ 현존하는 다양한 마이크로 데이터의 결합
  - ✓ 데이터의 제공 기법 개발
  - ✓ 등 등

고맙습니다!!!

공유경제 (Sharing Economy)와 통계: 숙박 및 교통

국가통계발전포럼

조윤정 교수
KDI 국제 정책 대학원

공유경제의 프래임

- 개인소유를 목적으로 하는 과거의 경제 활동보다는 여럿의 사회 구성원들이 지원 공유를 통해 만들어가는 공유경제가 더 효율적이고 합리적으로 다수 소비자들의 필요를 동시 충족.
- 보다 광의적인 개념으로 공유가치창출 (Creating Shared Value, CSV) 과 같은 맥락에서 이해하기도 함 (김기연 2014).

공유경제의 정의

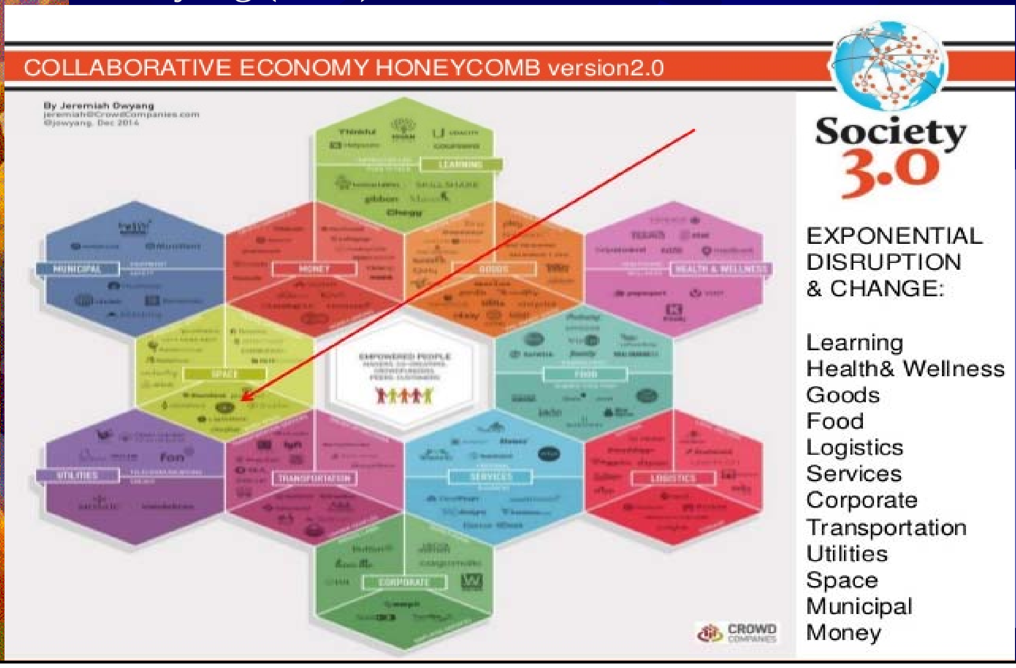
- ‘특정 서비스의 수요자’와 ‘해당 서비스를 창출하는 유휴자산’을 보유한 ‘공급자’간 해당 유휴자산을 이용한 ‘시장거래’를 ‘ICT 플랫폼이 중개’하는 경제.

공유경제의 정의

- 다양한 용어로 불림
- 소비자 대 소비가 공유 (“consumer-to-consumer sharing”)
- 개인 대 개인의 시장 (“peer-to-peer marketplace”)
- 공동 소비 (“collaborative consumption”)
- 공동 네트워크의 하이브리드 경제 (“Hybrid Economies of Collaborative Networks”)
- 접속 기반 소비 (“Access-Based Consumption”)
- 인터넷 활성화된 소비 (“Internet-facilitated Sharing”)

공유경제의 범위

- Owyang (2015)은 벌집예로 들어 버전 3를 소개.



공유경제의 분류

- 숙박공유:
  - 플랫폼 예시: 에어비엔비 (US), 코자자, 마이리얼트립 (한국)
  - 거래형태: 플랫폼이 이용하지 않는 방이나 집(유휴자산)의 보유자와 숙박수요자를 중개하면, 보유자는 숙박수요자에게 방이나 집을 렌탈하는 형태
- 교통공유 (개인차량공유)
  - 플랫폼 예시: 릴레이라이즈(RelayRides), 겟어라운드(GetAround) (US)
  - 거래형태: 플랫폼이 일시적으로 이용하지 않는 차량(유휴자산)의 보유자와 차량수요자를 중개하면, 차량보유자는 유휴차량을 차량수요자에게 대여하는 형태

2016-08-24 Copyright*Dr. Cho 6

공유경제의 분류

- 교통공유 (승차공유)
  - 플랫폼 예시: 우버(Uber) (US), 카카오택시, 티클 (한국)
  - 우버는 여분의 좌석이 있는 ‘자가용차량’을 수요자에게 중개.
  - 카카오택시 등 콜택시는 여분의 좌석이 있는 ‘영업용택시’를 중개
  - 자가용차량의 유상운송은 현행법(여객자동차운수사업법, 81조)상 불법
- 예외적 사례: Zipcar, 쏘카, 그린카 등 B2C
- 금융공유(크라우드펀딩): 조파(Zopa), Kickstarter(US),
- 지식·재능공유: 태스크래빗(TaskRabbit) (US)
- 공간공유: 쉐어데스크(ShareDesk) (US)

2016-08-24 Copyright*Dr. Cho 7

공유경제의 특성

- 단면시장은 기업과 소비자가 직접거래하는 경제를 지칭하는 반면, 양면시장은 플랫폼기업이 수요자와 공급자를 매칭(matching)하여 거래를 성사시키는 경제를 지칭.
- 인터넷, 스마트폰 등 정보통신기술의 발전으로 인해 태동한 ICT 플랫폼 산업*이 양면시장의 대표적인 예.
- 공유경제는 공유 플랫폼이 유휴자산보유자와 수요자를 매칭하여 거래를 성사시킨다는 측면에서 양면시장의 한 예.
- 양면시장의 핵심적인 특성은 간접적 망외부성(indirect network externality)임.
  - 사용자 1명이 플랫폼의 중개서비스로부터 얻게 되는 편익은 해당 플랫폼을 사용하는 반대 쪽 사용자 그룹의 규모(size)에 비례하여 증가한다는 것

2016-08-24 Copyright*Dr. Cho 8

공유경제의 경제적 효과

	긍정적 효과 (기대요인)	부정적 효과 (우려요인)
단 기 자산이 주어진 (fixed) 경우	신규거래 창출효과 주어진 자산의 활용도를 높여 상호호혜적(mutually beneficial) 거래를 성사시키고 이에 따라 수요자, 자산보유자, 플랫폼기업의 후생(surplus)이 모두 증가	기존거래 구축효과 기존공급자로부터 서비스를 구입했던 일부 수요자가 자산보유자로 구입대상을 전환함에 따라 기존공급자의 후생 감소
장 기 자산은 가변 (flexible), 자산을 생산하는 자원이 주어진 경우	고부가가치생산 창출효과 주어진 자원을 경제적 가치가 보다 높은 자산의 생산에 대체 투입함으로써 경제성장 제고	저부가가치생산 구축효과 기존자산 생산업체의 생산감소에 따른 경제성장 위축효과

2016-08-24 Copyright*Dr. Cho 9

공유경제의 기대 효과

- 정보통신기술(ICT)을 활용하여 상호호혜적 거래를 성사시키고 소비자의 편익을 제고하며 공급자에게 새로운 홍보수단을 제공하고 신성 장동력으로 기능하는 등 다양한 측면에서 국가경제에 이바지할 것으로 기대.
- 주요 기대효과:
- 1) 사회후생 증가효과: 거래에 참여하는 수요자, 공급자(유휴자산 보유자) 및 거래를 중개하는 플랫폼 등 모든 거래 당사자의 후생(혹은 효용)을 증가
  - 소비자 1천명을 대상으로 진행한 2014년 서베이 결과에 따르면 미국 공유경제 소비자의 83%는 공유경제가 편리하고 효율적이라고 답했고 72%는 적어도 향후 2년간 지속적으로 공유경제에 참여할 것이라고 대답 (PWC)

2016-08-24 Copyright*Dr. Cho 10

공유경제의 기대 효과

- 2) 경제성장 효과:
  - 공유거래의 공급자인 유휴자산 보유자는 사회적 약자인 저소득층, 노인, 여성, 청년층인 경우가 많으므로 공유거래를 통한 공급자 후생 증가는 경제적 효율성 뿐 아니라 사회적 배려계층에 대한 분배적 정의도 제고
  - '공유경제의 도입'으로 인해 중위소득 이상 계층의 후생은 평균 1.42% 증가한 데 그쳤으나 중위소득 이하 계층의 후생은 평균 5.26% 증가 (Fraiberger and Sundararajan, 2015)
- 3) 환경비용 저감효과
  - 교통공유의 주된 기대효과로 형태별 환경비용 저감효과가 다소 상이 (개인차량공유, 업체차량공유, 승차공유, 카풀 등)

공유경제의 우려요인

- 1) 기존의 사업 모형을 일부 대체함에 따라 기존사업과의 마찰 초래 할 수 있음.
  - 기존사업과 마찰을 일으키지 않는 형태의 공유경제도 가능 (예: 지식/재능/공간 등 기존사업자가 존재하지 않음)
  - 주택, 차량 등 유휴자산의 활용도가 증가하여 해당 자산에 대한 매입수요가 감소할 것이므로 자산생산도 감소할 수 있음.
  - 공유경제 활성화를 통해 신규 수요가 창출될 경우 기존 생산기업들이 공유경제에 참여하여 손실을 줄이거나 추가적 수익원을 확보할 수 있음.
  - 예) 지엠(GM)은 계열사인 온스타가 판매하는 차량ICT 서비스를 정기구매(subscription)하는 차량 소유주는 차량공유를 쉽게 할 수 있도록 기술적 편의*를 제공함으로써 차량 ICT 서비스 판매를 촉진

공유경제의 우려요인

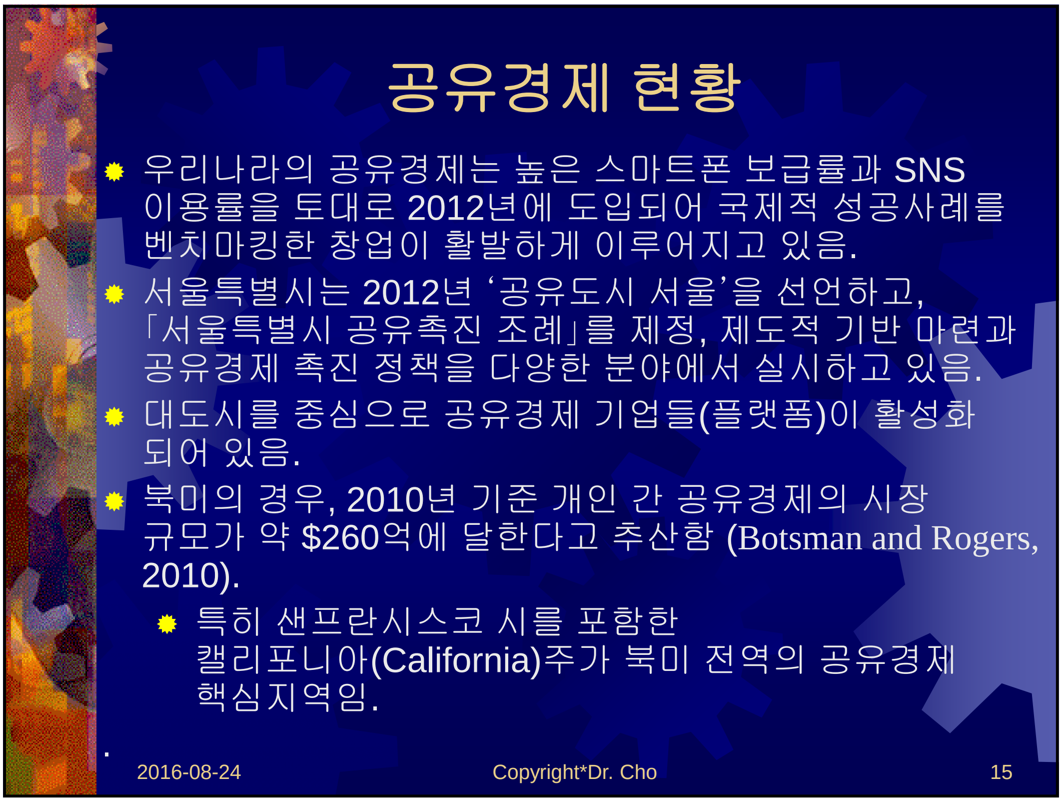
- 2) 제도적 기반의 미비: 관련 법제도가 미비하여 공유거래가 합법인지 불법인지에 대한 명확한 기준이 없고 이에 따라 수요자, 공급자, 플랫폼 등 거래당사자들은 법적리스크 (세금, 안전 등)를 부담
- 3) 공유거래는 서로 알지 못하는 수요자와 공급자간 거래이기 때문에 정보의 비대칭성 (information asymmetry) 문제를 수반하고 이에 따라 거래상 위험에 노출
- 4) 사회적 위험: i) 거래상 위험의 정도가 심해져 사회적 위험으로 비화되는 형태, ii) 공유거래가 부정적 외부효과 (교통사고, 화재위험, 집값상승 등)를 일으키는 형태, iii) 양질의 일자리를 저 품질의 일자리로 대체하는 노동문제 등으로 나눌 수 있음.

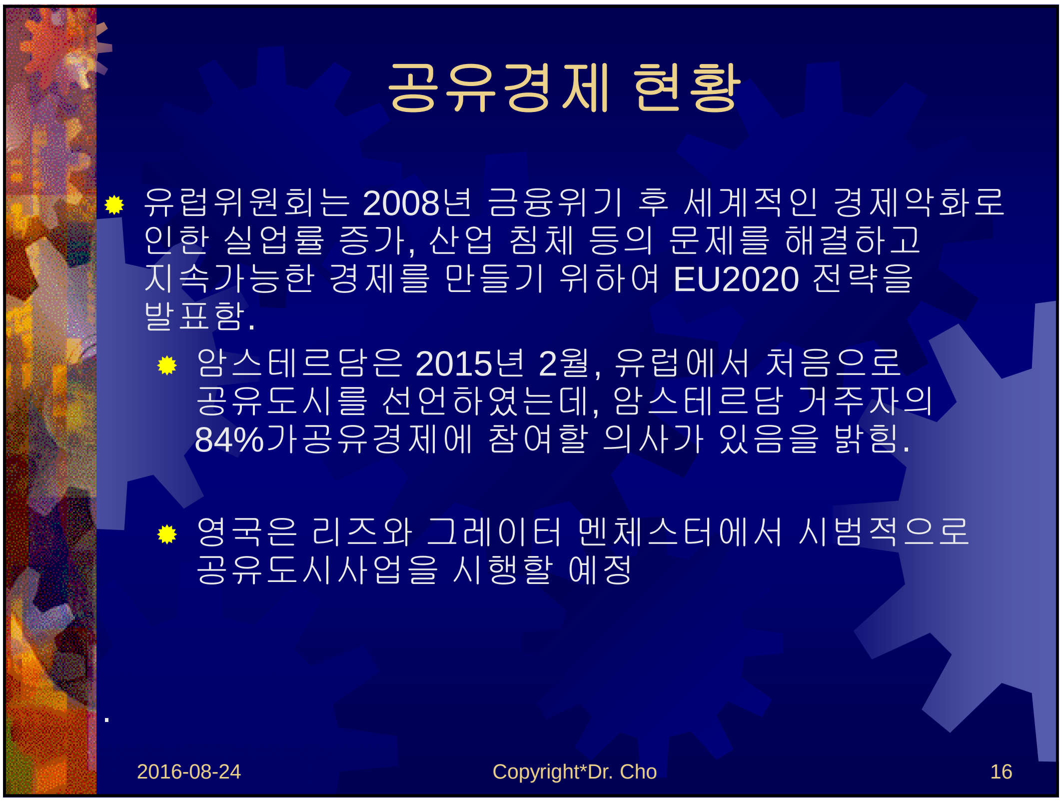
2016-08-24 Copyright*Dr. Cho 13

공유경제의 활성화 필요성 및 제도개선방향

- 1) 다양한 기대효과를 수반: 거래당사자들의 편익과 소득을 증가시키고, 질적 경제성장에 기여하며, 환경비용을 저감.
- 2) 세계적 흐름: 한국 뿐 아니라 미국, 유럽, 중국 등 주요국가에서 공통적으로 확산되고 있음.
- 3) 시대적 흐름: 밀레니엄 세대, 스마트폰 사용자계층, 1인 가구 등 공유경제의 잠재적 주사용자 계층이 인구의 다수를 차지함에 따라 공유경제는 거부할 수 없는 시대적 흐름으로 간주 됨.
- 4) 제도개선을 통해 관련 우려요인을 상당부분 경감시킬 수 있다는 점에서 활성화가 필요.

2016-08-24 Copyright*Dr. Cho 14


공유경제 현황
우리나라의 공유경제는 높은 스마트폰 보급률과 SNS 이용률을 토대로 2012년에 도입되어 국제적 성공사례를 벤치마킹한 창업이 활발하게 이루어지고 있음.
서울특별시는 2012년 '공유도시 서울'을 선언하고, 「서울특별시 공유촉진 조례」를 제정, 제도적 기반 마련과 공유경제 촉진 정책을 다양한 분야에서 실시하고 있음.
대도시를 중심으로 공유경제 기업들(플랫폼)이 활성화 되어 있음.
북미의 경우, 2010년 기준 개인 간 공유경제의 시장 규모가 약 $260억에 달한다고 추산함 (Botsman and Rogers, 2010).
특히 샌프란시스코 시를 포함한 캘리포니아(California)주가 북미 전역의 공유경제 핵심지역임.
2016-08-24
Copyright*Dr. Cho
15


공유경제 현황
유럽위원회는 2008년 금융위기 후 세계적인 경제악화로 인한 실업률 증가, 산업 침체 등의 문제를 해결하고 지속가능한 경제를 만들기 위하여 EU2020 전략을 발표함.
암스테르담은 2015년 2월, 유럽에서 처음으로 공유도시를 선언하였는데, 암스테르담 거주자의 84%가공유경제에 참여할 의사가 있음을 밝힘.
영국은 리즈와 그레이터 멘체스터에서 시범적으로 공유도시사업을 시행할 예정
2016-08-24
Copyright*Dr. Cho
16

교통분야
(카셰어링)

2016-08-24 Copyright*Dr. Cho 17

카셰어링 도입배경 및 현황

- 교통인프라 공급정책의 한계, 기존 교통서비스의 한계, 교통수단분담률의 고착화와 같은 현재의 교통체계 문제에 대응하기 위함
- 지자체 및 민간업체에 의한 상업적인 서비스가 전국적으로 확산.
- 주요 장점: 개인 자동차 보유수, 통행거리 감소, 주차수요 감소, 환경문제 도움.
- 많은 연구자가 카셰어링 시스템의 성장잠재력을 높게 평가하고 있음.

2016-08-24 Copyright*Dr. Cho 18

카셰어링 도입배경 및 현황

- 문헌상으로는 1948년 스위스 취리히 공동주택의 카셰어 프로그램으로, 공동 주택에 거주하는 여러 사람이 공동으로 차량을 구입하여 공동으로 소유, 관리, 이용하는 서비스.
- 2000년 12대의 리스 자동차로 시작한 미국의 zipcar (Hart 외)의 새로운 형태의 비즈니스 모델이 응용 되어 온 실정.

카셰어링 유형

	유형	한국	해외 마켓	
			북미	유럽
1	B to C (Business to Consumer)	그린카, 쏘카 유카, 한국카셰어링, 씨티카	미국: Zipcar, GoGet, Citycarshare, StreetCar 캐나다: CarShareHFX, Co-Auto	독일: BMW's DirveNow, Volkswagen's Quicar, Car2Go, StattAuto, Car2Go, StattAuto, Cambio 스위스: Mobility CarSharing, 노르웨이: Bikollekivet, 네덜란드: Greenwheels 영국: Whizzgo, CityCarClub France: Autolib
2	Carpool (Ride sharing)	티클	Modo. Goroco, zimride	
3	P to P (Peer to Peer)	불법	DriveMyCar.com RelayRides, Lyft, Flight Cars	France: Buzzcar Netherland: Speelotheken

카셰어링 유형

- 1) 민간기업의 카셰어링 서비스 (예: zipcar, 쏘카)
- 2) 공기업의 카셰어링 서비스 (예: 코레일 네트워크의 유카)
- 3) 기존 렌터카 업체의 카셰어링 서비스 (예: Hertz의 Connect와 on Demand, Enterprise Rent-A-Car의 WeCar, Avis의 Avis On Location, 그린카)
- 4) 전기차 중심의 카셰어링 서비스 (예: Car2go, 독일에서 시작, 독일, 네덜란드, 오스트리아. 프랑스, 미국, 캐나다 등에서 서비스 제공)
- 5) 자동차 업계의 카셰어링 서비스 (예: BMW, Folkswagen)
- 6) 개인 승용차의 공유 (예: RelayRides, Buzzcar등)

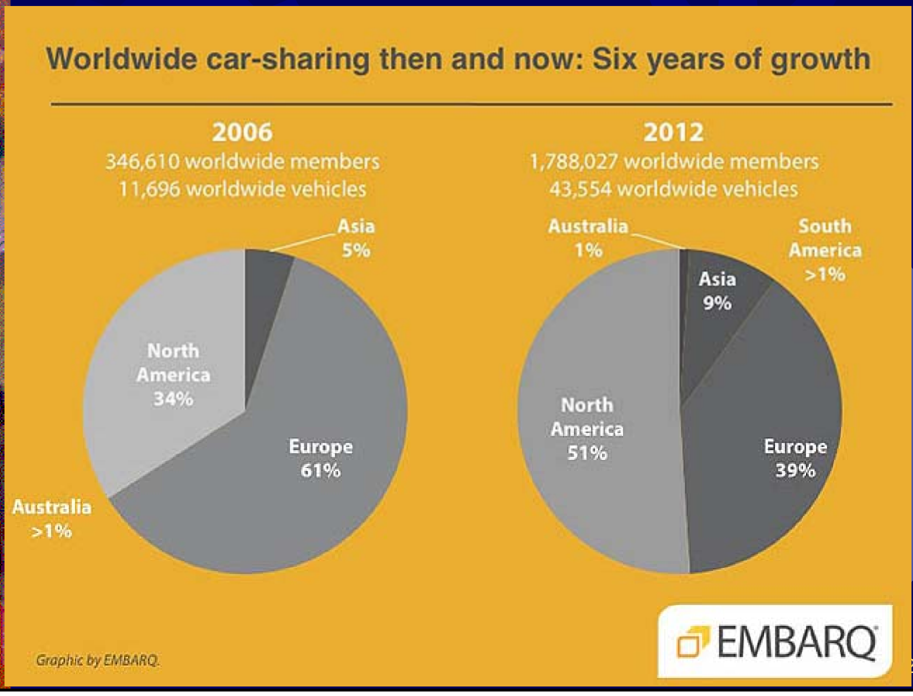
2016-08-24 Copyright*Dr. Cho 21

카셰어링 이용자 특성

- 1) 연령별: 20 ~ 30, 성별: 남성 가입자 수가 많음
- 2) 이용빈도: 월 3.2회.
- 3) 주차공간 부족 및 높은 주차비등 차량 보유가 어려운 경우 (예: 뉴욕시)
- 4) 차량 보유율이 낮거나 차가 드문 지역 (예: 세종시, 대학 캠퍼스 등)
- 5) 필요시에만 이용 (의료기관 방문, 주말 , 출장, 무거운 짐 운반시, 등)
- 6) 대중교통이용이 불편한 경우
- 7) 프라이버시가 필요한 경우

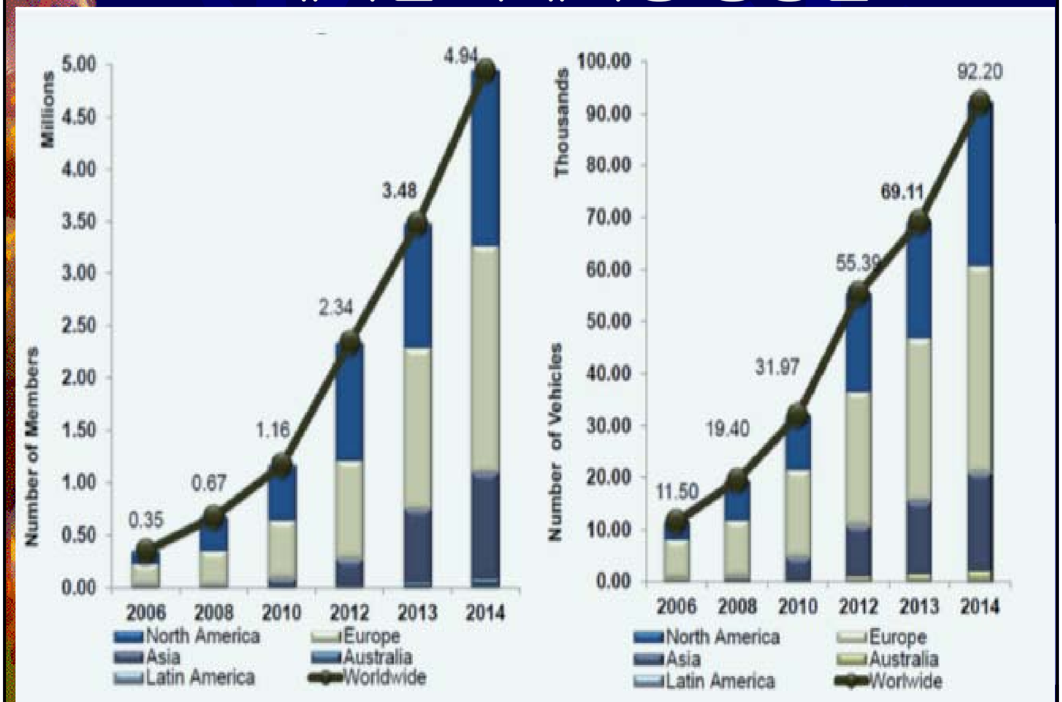
2016-08-24 Copyright*Dr. Cho 22

전세계 카세어링 성장율



23

대륙별 카세어링 성장율



카셰어링 현황 (한국)

	쏘카	그린카
창립년도	2011	2011
매출 (2014)	147 hundred million	140 hundred million
이용자수	850,000	650,000
자동차수	2,900	2,000
카셰어링 가능 지역수	Approximately 1,200	Approximately 1,400
기본요금	1390won	1380won
거리당 요금 (km)	180won	180won
비즈니스 유형	Rooted in Jeju. Partnered with cacaotaxi, Jeonjaland	Operated with Lotte Rent-a-car

카셰어링 기대효과

- **1)** 통행비용 감소효과
- **i)** 통행량 감소효과
  - 기존의 승용차를 매각하고 카셰어링 서비스를 이용하는 경우 통행량 감소
- **ii)** 통행량 감소에 따른 유류사용량 감소 효과
  - 유류사용 감소효과를 경제적 비용으로 환산하면 연간 **841**억원의 효과로 추정 **(**장원재 외**, 2008)**
- **iii)** 온실 가스 배출량 감소효과
  - 교통혼잡 완화로 에너지 소비량 및 온실가스 배출량 감소 (박준식 외, 2013)
  - 전기차나 하이브리드 차량 등 친환경 차량 및 소형차 이용 시.
  - CO2 배출량 감소 (유럽의 경우 평균 11% 정도 적게 연료소비)

2016-08-24 Copyright*Dr. Cho 26

카셰어링 기대효과

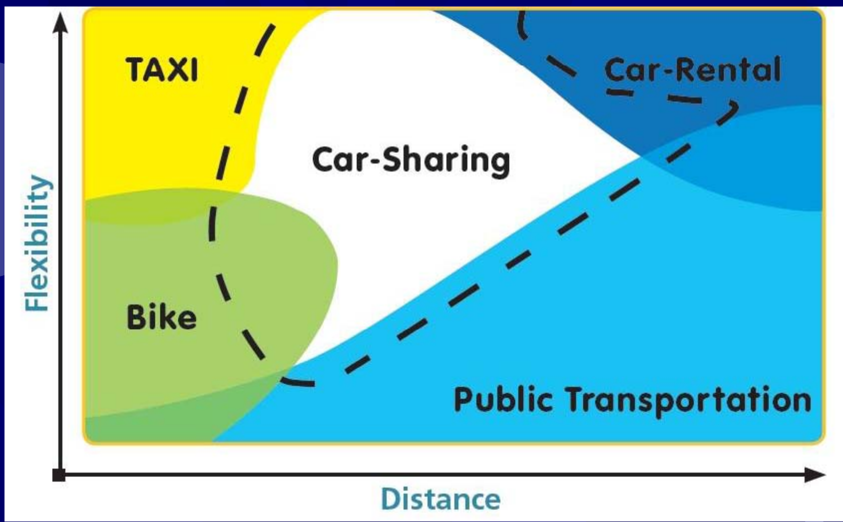
- **2)차량보유 및 주차공간 감소효과**
- **3)** 테크놀로지 발전에 따른 기대효과
  - (e.g., RFID, IoT, 소프트웨어 모니터 기능으로 차량 유지 보수 및 기능 체크 가능)
- 4) SNS 등 이용 온라인 커뮤니티 이용 기대 효과
  - 이용자들간의 커뮤니티 형성으로 이용 증가.
- 5) 대중교통 (예: KTX, 카카오택시)과의 연계에 따른 이용 증가

카셰어링 효과

- 학계 분석:
  - 효용 증가 효과 (Sharing Utility)
  - 거래상 효용 (Transaction utility)
  - 이동성 효용 (Mobility utility)
    - 대중교통 (예: KTX, 카카오택시)과의 연계에 따른 이동성 효용
  - 주차공간에 따른 효용 (storage utility)
  - 사회적 효용 (Social utility)
  - 감정적 효용 (Emotional utility)
  - 기술적 효용 (Technology utility)에 따른 가치 증대로 분석

카셰어링 효과

- 카셰어링은 단거리 및 중거리에서 가장 효과적



Source: Schwartz, Joachim. (1999)

2016-08-24 Copyright*Dr. Cho 29

정책시사점 및 대응방안

- **1.** 기존사업과의 마찰
  - 보유차량 감소로 관련된 세금 감소, 차량 판매량 감소 로 자동차 업계에 미치는 영향 우려 될 수 있음.
  - 각 자동차 업계도 카셰어링을 자체적으로 하고 카셰어링에서 차를 구매함으로 우려 감소.
  - 렌터카 업계와의 마찰은 렌터카가 자체적으로 카셰어링을 하는 경우가 많아 마찰 적을 것으로 예상.

2016-08-24 Copyright*Dr. Cho 30

정책시사점 및 대응방안

- **2. 대중교통과의 연계의 필요성**
  - 철도역 및 중장거리 노선의 버스 등 대중교통과의 연계를 위한 인프라조건이 양호한 지역을 근거지로 하는 카셰어링 거점에 있어 활용도가 두드러지는 것 확인.
  - 지방자치단체 및 전국철도사업자와의 연계를 통한 카셰어링 사업 고려
  - 환경부하저감을 목표로 연계를 시도하는 사업 모델개발 필요 (예: 일본)
  - 대중교통요금통합시스템을 활용하는 정책 마련

정책시사점 및 대응방안

- **3. 지역별교통 수요 분석에 따른 효율성 증대**
  - 수요그룹에 따른 지역별 이용수요 효과 분석에 따라 카셰어링 거점확보 및 교통수요에 대응할 수 있는 전략방안 마련.
  - 수요그룹 확대를 위한 방안 마련.
  - 사회경제적 속성 및 통행목적 등 교통 행동주체의 사회적 속성에 대한 구체적인 조건을 반영한 분석에 따른 정책 마련.

정책시사점 및 대응방안

- **4.** 데이터베이스 활용 및 분석에 따른 효율성 증대 (최현수 외)
  - 인구-사회-경제-교통 등 지역적 특성을 반영하는 데이터베이스 활용 증대.
  - 철도역 발전관련 (예: 발전 빈도, 국토사용비율 등) 데이터베이스 분석
  - 녹색교통 접근성을 역세권별로 정리 (예: 개발특성, 주거, 상업, 업무 등에 대한 토지이용 복합도, 접근성 등)
  - 주변개발계획에 따른 분석
  - 대중교통 운영 특성에 따른 분석

숙박분야

도입배경 및 현황

- 대표적인 미국의 에어비앤비의 경우 컨퍼런스등으로 숙박시설이 부족한 상황에서 출발하여 현지 거주민들이 자신의 거주지를 기반으로 큰 추가 투자 없이도 갖고 있는 집과 주변시설을 활용해 경제적인 도움을 얻을 수 있고, 스스로 가치 중심 모델 (Value-driven-model)로 일컬음 (조선비즈)
- 2014년 2월 한국형 숙박공유 중개사이트 (코자자, 비앤비히어로)가 창업되고, 부산의 시민단체가 “공유경제 창업기업 인큐베이팅센터” 등을 여는 등 우리나라에서도 공유경제 숙박부분 등장하기 시작.

도입배경 및 현황

- 외국의 경우 여러가지 형태의 도시민박업으로 운영되며 제도적, 운영 및 관리 측면, 정부의 지원정책 측면 기준이 마련 되 왔으며, 특히 공유경제 개념으로 2008년 시작된 에어비앤비가 부각되면서 급속히 성장.
- 국내의 경우 현재 공유경제 개념이 적용되면서 숙박부분은 민박업으로 분류, 외국인관광 도시민박업으로서 추진되었으며 게스트 하우스 와 별도의 개념과 기준을 적용 (한국관광공사).

.

도입배경 및 현황

- 공유경제 숙박 플랫폼을 이용하는 게스트는 외국인관광 도시민박업에 따라 외국인으로 제한되어 옴.
- 2016년 공유민박업 (공유숙박업: 가칭) 신설하여 주거중인 주택을 숙박서비스에 제공할 수 있도록 허용 (현재 규제프리존 (부산, 강원,제주) 도입후 전국 확대 추진).
- 도시지역 (전용주거지역 제외)에서 내/외국인을 대상으로 연간 최대 120일간 숙박서비스 제공 가능으로 추진 계획 (연면적 230m² 미만; 대상: 단독/다가구/아파크/연립/다세대주택)

2016-08-24 Copyright*Dr. Cho 37

도입배경 및 현황

- 공유민박업 등록기준: 인적조건, 물적조건 (주민등록이 되어 있는 거주지), 허용기준, 규모제한 등
- 공유 민박업자의 준수사항: 영업가능일수:
  - 연간 120일 이내
  - 서비스 기준: 안전 위생 시설 기준 등 규정
- 공유 민박업자 등에 대한 관리 및 감독
  - 자율규제: 이용자 및 공급자 간 평판조회시스템 (후기, 평점 등) 운영 등
  - 관리 및 감독: 필요시 공유민박 중개자에 대한 자료 요청, 관계기관수시 합동단속, 준수사항 위반시 제재 (과태료 등) 등
- 전국 확대 추진시 개별법률 (공중위생관리법, 관광진흥법)에 산재된 숙박업 규정을 통합, 통합적인 관리체계 마련 (문체부)

2016-08-24 Copyright*Dr. Cho 38

숙박공유 기대효과

- **1) 관광경쟁력 강화**
  - 국가차원의 공유경제 사회적 인식확대 및 관광경쟁력 강화.
  - 전국 확산 및 국내 사업자들의 재투자 등을 통한 시장의 활성화.
  - 공유경제가 관광산업에서 갖는 잠재력을 확인하고, 일반 대중이 공유경제의 개념을 인식하는 계기 마련 필요 (한국 문화 관광 공사)

숙박공유 기대효과

- **2) 일자리 창출 기회**
  - 공유경제는 새로운 패러다임으로 일자리 창출 기회 도움
- **3) 유휴자원 활용**
  - **공공재에 대한 인식 필요.**
- 4) 새로운 경제 패러다임
  - microentrepreneurship (Rifkin, 2014) 에 참여하는 것으로, 새로운 경제 패러다임으로 인식 전환 요구.

정책시사점

- **1)** 제도적 기반 조성을 통한 숙박공유 활성화
  - 공유민박업 (가칭) 확대 관련 이슈:
  - 내국인 민박 허용 문제
  - 신개념 숙박시설 관련 관리시스템 정비 및 인식 개선.
  - 관리역량 확보: 알선기관의 인력 및 권한 증대를 통한 관리역량 강화.
  - 외국의 케이스를 참고하여 사업의 본질 이해 필요.

정책시사점

- **2) Start-up** 비즈니스의 창업환경 개선 요구.
  - 수요자, 공급자, 및 정부와 협력하여 공유경제 활성화에 도움.
  - 공유경제에 대한 의미와 효과성에 대한 이해, 시장현황에 대한 정확한 이해 선행 요구.
  - 외국 플랫폼과 공정한 경쟁을 할 수 있는 창업 환경 개선 필요.
- 3) 세금 관련
  - 공급자에게 일정 금액 이상의 경우 과세 하는 등 방안 마련 (예: 독일)

정책시사점

- **4) 공유경제 숙박을 관련 테크놀로지 개발의 필요성.**
- 공유경제는 테크놀로지를 통한 재발견 (Rifkin, 2014).
- 공유경제 핵심인 접근성과 편리성을 위해 제품 서비스 통합 시스템 개발 (PSS: Product Service System) 의 필요성
- RFID, IoT (Internet of Things), SNS등 테크놀로지 기술 향상의 필요성 (예: 사회 문제 기록 확인, 미등록 공급자 확인의 용이성)
- 5) 공유경제 홍보의 필요성
  - 공유경제에 대한 대중의 인식 향상 필요.

2016-08-24 Copyright*Dr. Cho 43

공유경제와 통계 (교통)

- 기대효과에 따른 분석 (통행비용, 통행량, 차량보유, 주차공간)
- 이용도에 따른 지역적, 사회적 요인 분석
- 사용도에 따른 위험과 신뢰분석.
- 접근성, 일시적, 이용자 참여도, 시장중개 (market mediation), 사회교류에 대한 인식 (political consumerism)에 따른 분석.
- 면허정보제공 및 조회시스템 구축 (2017, 2/4)에 따른 활용도.
- 주차장 확보 문제에 따른 분석.
- 대중교통과의 연계에 따른 효과성 분석.
- 시범도시 및 행복주택, 뉴스테이에 도입 (2016, 4/4)에 따른 효과성.

2016-08-24 Copyright*Dr. Cho 44

공유경제와 통계 (숙박)

- 기대효과 (일자리 창출 등)에 따른 분석.
- 우려요인 (등록 여부, 위생, 소방 등 사회적요인 등)에 따른 분석.
- 기존 민박업과의 형평성, 공유경제의 취지 등에 따른 문제 분석.
- 규제프리존에서의 활용도 및 확대 추진 (2017, 6) 가능성.
- 대상 (단독/다가구/아파크/연립/다세대주택)에 따른 분석.
- 공유경제 인식에 대한 분석.

References

- Socar (2015), In-Depth Interview.
- Greencar (2015), In-Depth Interview.
- 박준식·문지혜, 「카셰어링 수요분석 및 지역별 사업타당성 분석방법 연구 (Estimation of the Demand for Car-Sharing Service)」, 『한국교통연구원, 수시연구』, 24호, 2012.
- 박준식·문지혜, 「카셰어링 서비스의 수요추정 및 도입효과 분석」, 『교통연구』, 20권, 2호
- 장원재·박준식·김동준, 「자동차 공동이용 (Car-Sharing) 시스템 도입 방안 연구」, 『연구총서』, 한국 교통 연구원, 4호, 2008.
- 최현수 ·박준태, 「서울시의 카셰어링 이용도에 대한 지역적 요인 특성분석」, 『』, 20권, 2호
- 김기연, 「공유경제 비즈니스에 대한 한국인의 가치인식 스키마 유형 연구」, 『Journal of KSSSS』, 29, 99-119.

References

- Belk, Russell (2010), “Sharing,” *Journal of Consumer Research*, 36, February, 715-734.
- Cho, Sanggu, Founder of Kozaza (2015), In-depth Interview.
- Cho, Yoonje, Founder of bnbhero (2015), In-depth Interview.
- Rifkin, Jeremy (2014), **“**The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism,” St. Martin’s Griffin.
- Wosskow, Debbie (2014), “Unlocking the Sharing Economy: An Independent Review,” Novermber.
- www.collaborativeconsumption.com
- Botzman, Rachel (2013), “The Sharing Economy Lacks A Shared Definition.”
- Schwartz, Joachim. Presentation at Car-Free Cities Working Group Seminar, London, 1999.

THANK YOU.

2015 인구주택총조사의 패러다임변화

등록센서스 방식 도입

2016. 8. 31.

통계청

통계청

목 차

통계청

Ⅰ. 등록센서스 도입 배경 및 추진 방향

1. 등록센서스 개념
2. 도입 배경
3. 인구주택총조사 환경 분석
4. 인구주택총조사 추진 방향

통계청

Ⅰ. 등록센서스 도입 배경 및 추진방향

1. 등록센서스 개념

등록센서스 :

주민등록부, 건축물대장 등
인구, 가구, 주택과 관련된
가용한 행정자료를 이용하여
현장조사 없이
통계를 생산하는
새로운 인구주택 총조사 방식

※ 다양한 행정자료 활용을 위해 부처간 소통과 협력이 필수

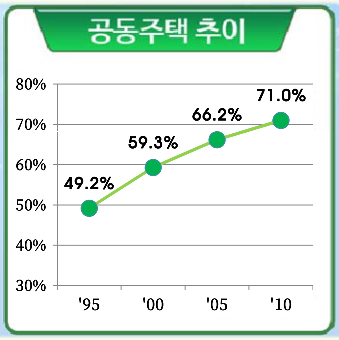
4

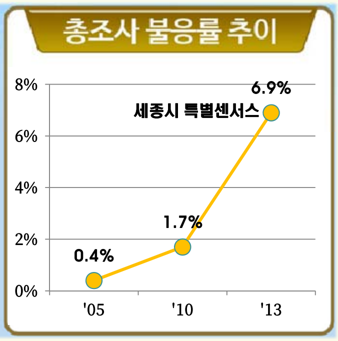
통계청 **Ⅰ. 등록센서스 도입 배경 및 추진방향**

2-1) 도입배경 : 현장조사 환경 악화

현장조사 환경의 지속적인 악화 → 총조사 품질저하 우려

- **(인구, 가구 변화)** 1인가구와 맞벌이가구, 고령인구 증가로 응답자 면접이 어려움
- **(주택 변화)** 외부인 출입을 통제하는 공동주택 증가로 조사원의 접근 곤란
- **(의식 변화)** 사생활 보호의식이 높아짐에 따른 통계조사 불응사례 증가





5

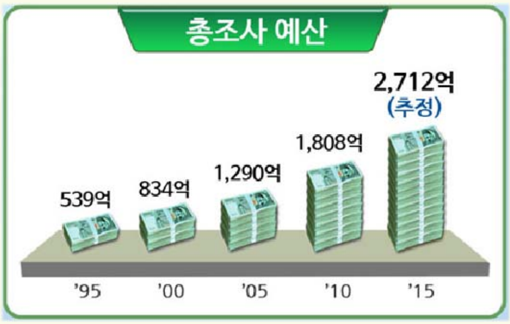
통계청 **Ⅰ. 등록센서스 도입 배경 및 추진방향**

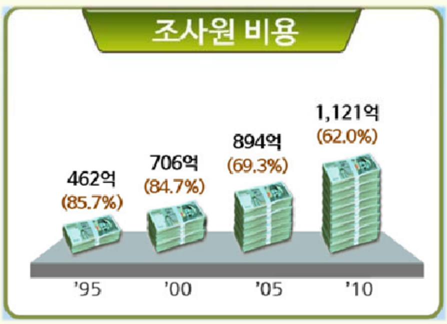
2-2) 도입배경 : 총조사 예산부담 증가

대규모 현장조사(전수조사) 실시로 비용이 크게 증가 ⇒ 조사효율화 요구

- 조사대상 가구수(1인 가구 등) 증가
- 조사원 인건비 상승
- 조사가구 부재 등으로 가구방문 횟수 증가

현장조사 사업비 증가
(매 주기별 50% 수준)





6

통계청 Ⅰ. 등록센서스 도입 배경 및 추진방향

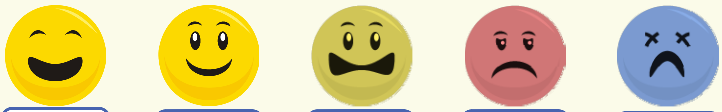
2-3) 도입배경 : 통계 시의성 개선 요구

5년 주기 조사 → 결과의 시의성 한계

- (통계 시의성) 5년 주기 조사를 통해 급변하는 현대 사회상 설명에 한계 존재
- (모집단 노후화) 표본조사를 위한 모집단 정보 제공 시, 5년 동안의 변화 반영 곤란

2010년 인구주택총조사 실시

통계자료 및 모집단 정보 시의성에 대한 만족



2011년 이용자 | 2012년 이용자 | 2013년 이용자 | 2014년 이용자 | 2015년 이용자

7

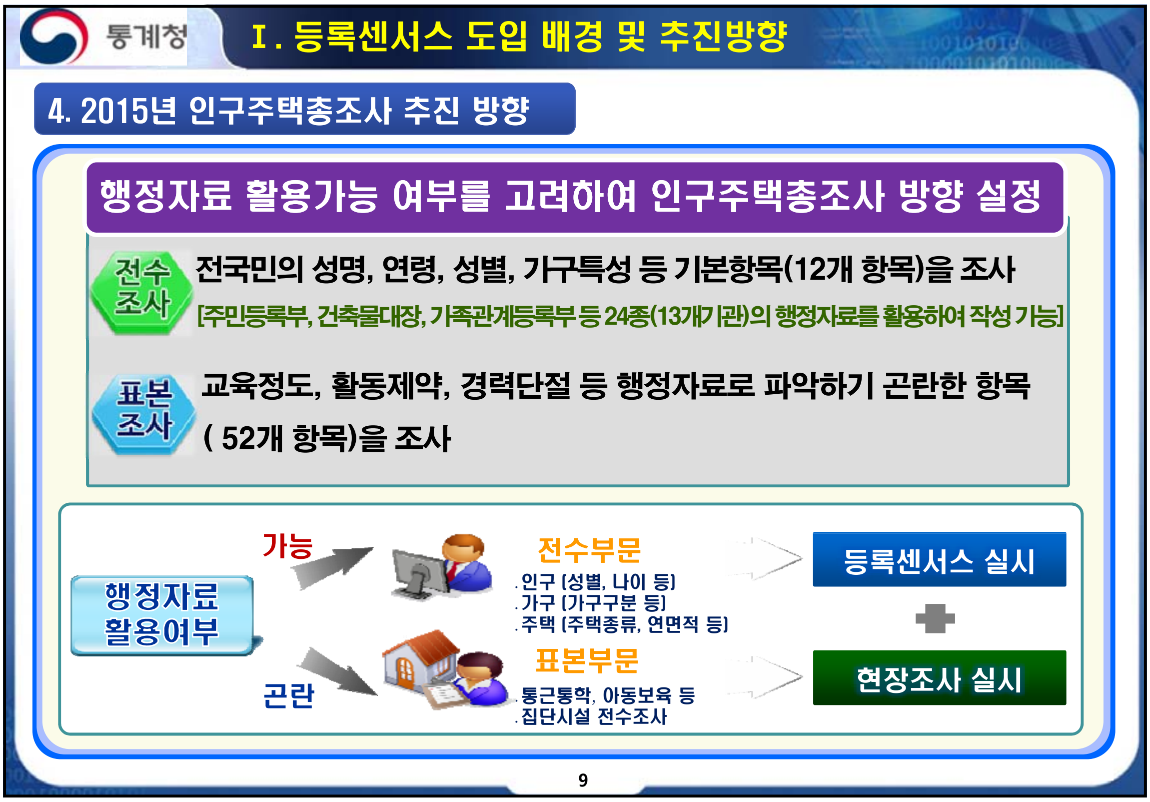
통계청 Ⅰ. 등록센서스 도입 배경 및 추진방향

3. 인구주택총조사 환경 분석

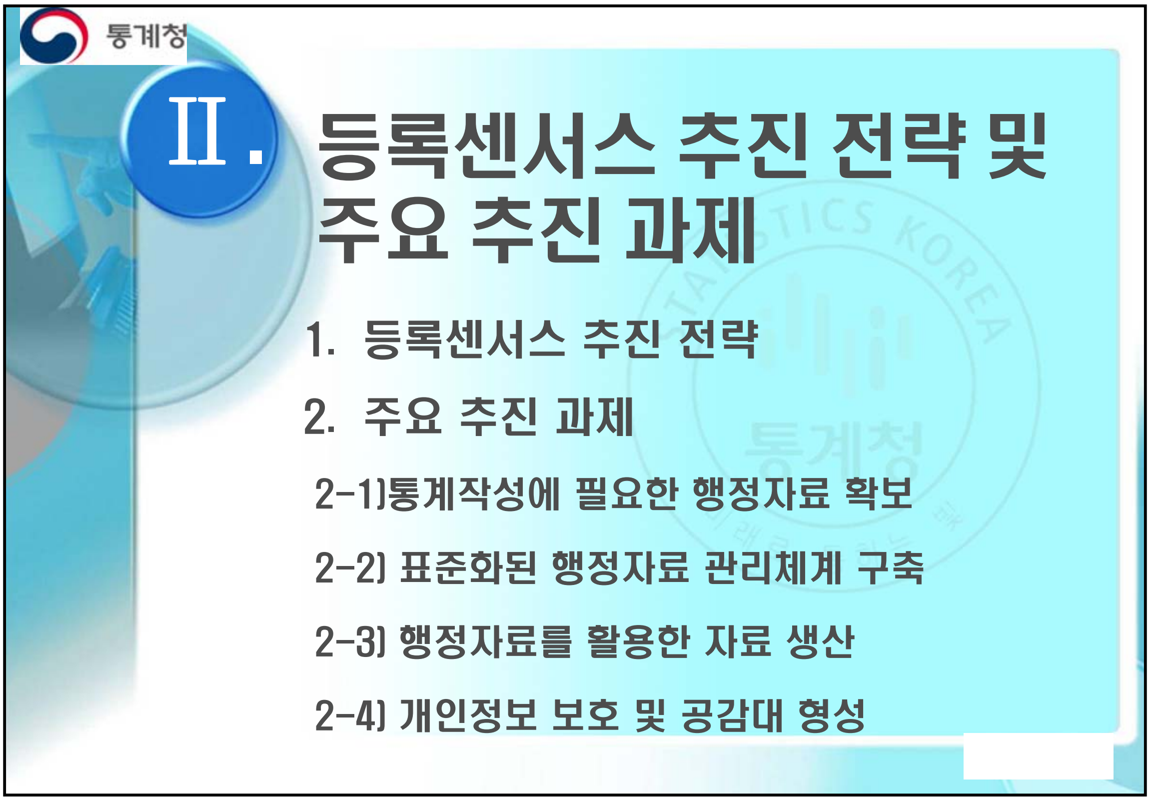
내부환경 분석 / 외부환경 분석	Strengths(강점) 1. 통계작성 활용 가능한 행정자료 지속적 증가 2. 총조사 추진을 위한 전문인력 보유 3. 고품질 통계 작성을 위한 전문성 보유 4. 주택DB에 대한 행정자료 활용 경험	Weaknesses(약점) 1. 인구주택총조사의 고비용 구조 2. 행정자료의 기관별 비표준적 관리 3. 다 부처 행정자료의 안전관리 부담
Opportunities (기회요인) 1. 등록센서스 도입국 및 관심국 증가세 2. 행정자료 전산화 3. 행정자료 공유 및 활용확산 추세 4. 도로명 주소 정착	SO전략 • 통계작성에 필요한 행정자료 확보	WO전략 (통계작성에 활용할 수 있는) • 표준화된 행정자료 관리체계 구축
Threats (위협요인) 1. 인구·가구구조의 변화 2. 개인정보유출사고에 따른 국민의 불안감 3. 5년 주기 통계자료의 시의성 한계 4. 행정자료활용 공감대 부족	ST전략 • 행정자료를 활용한 통계자료생산	WT전략 • 개인정보 보호 및 공감대 형성

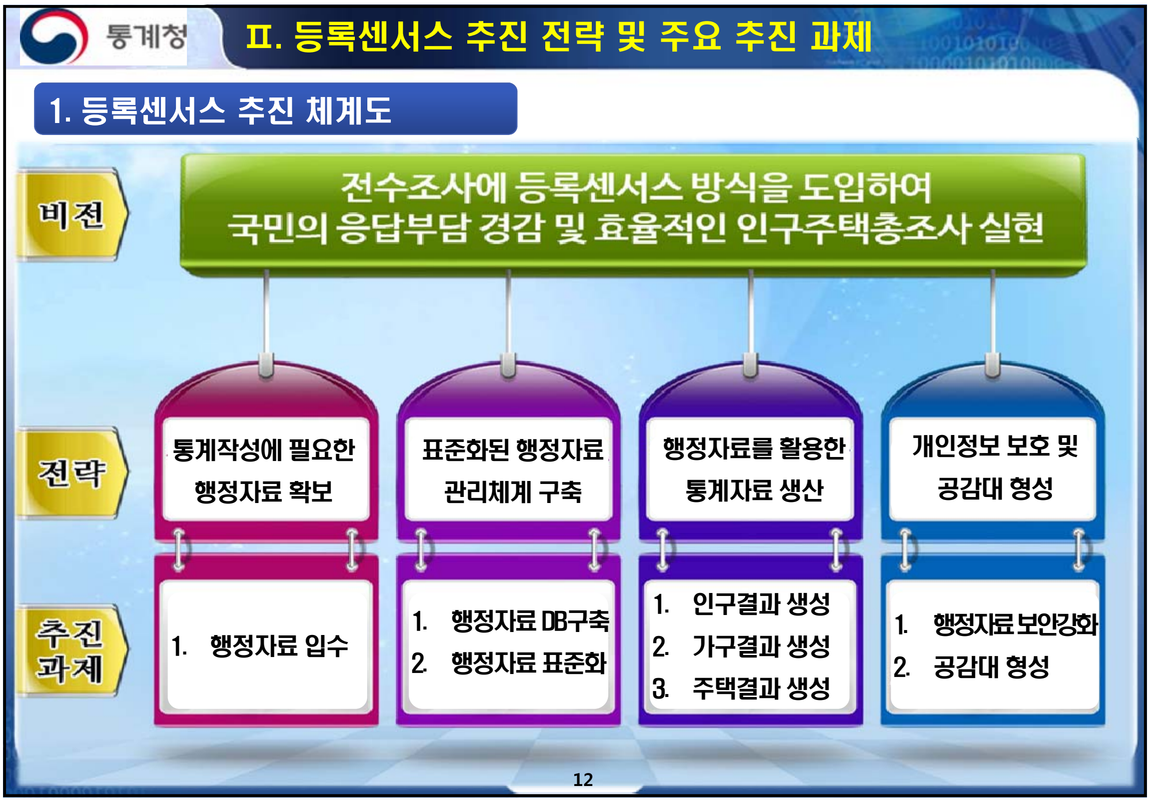
총조사 효율화 [등록센서스도입]

8


통계청
Ⅰ. 등록센서스 도입 배경 및 추진방향
4. 2015년 인구주택총조사 추진 방향
행정자료 활용가능 여부를 고려하여 인구주택총조사 방향 설정
전수조사
전국민의 성명, 연령, 성별, 가구특성 등 기본항목(12개 항목)을 조사
[주민등록부, 건축물대장, 가족관계등록부 등 24종(13개기관)의 행정자료를 활용하여 작성 가능]
표본조사
교육정도, 활동제약, 경력단절 등 행정자료로 파악하기 곤란한 항목 (52개 항목)을 조사
행정자료 활용여부
가능
곤란
전수부문
. 인구 (성별, 나이 등)
. 가구 (가구구분 등)
. 주택 (주택종류, 연면적 등)
표본부문
. 통근통학, 아동보육 등
. 집단시설 전수조사
등록센서스 실시
현장조사 실시
9


통계청
<참 고> 등록센서스 작성 절차
1 자료 입수
2 표준화 코드화
3 자료 연계
4 통계 전환
5 결과 공표
• 대상 : 13개 기관의 24종 행정자료
•기준시점 : 11월 1일 0시
• 자료간 연계키 관리
• 항목 표준화/코드화
• 인구 관련 자료 연계
• 거처 관련 자료 연계
• 센서스제공항목 선정
• 마이크로데이터 작성 집계표 작성
• 품질점검 및 이슈분석
• 결과보고서 공표
개인정보 보호 및 공감대 형성
10


통계청
Ⅱ. 등록센서스 추진 전략 및
주요 추진 과제
1. 등록센서스 추진 전략
2. 주요 추진 과제
2-1)통계작성에 필요한 행정자료 확보
2-2) 표준화된 행정자료 관리체계 구축
2-3) 행정자료를 활용한 자료 생산
2-4) 개인정보 보호 및 공감대 형성


통계청
Ⅱ. 등록센서스 추진 전략 및 주요 추진 과제
1. 등록센서스 추진 체계도
비전
전수조사에 등록센서스 방식을 도입하여
국민의 응답부담 경감 및 효율적인 인구주택총조사 실현
전략
통계작성에 필요한 행정자료 확보
표준화된 행정자료 관리체계 구축
행정자료를 활용한 통계자료 생산
개인정보 보호 및 공감대 형성
추진 과제
1. 행정자료 입수
1. 행정자료 DB구축
2. 행정자료 표준화
1. 인구결과 생성
2. 가구결과 생성
3. 주택결과 생성
1. 행정자료 보안강화
2. 공감대 형성
12

통계청

Ⅱ. 등록센서스 추진 전략 및 주요 추진 과제

2-1) 통계작성에 필요한 행정자료 확보-행정자료 입수

주요 내용: 등록센서스 시산 및 작성에 필요한 13개 기관의 24종 행정자료 입수

추진 경위
- 건축물대장, 주민등록부, 외국인등록부 등 7종 자료 입수(10년)
- 가족관계등록부, 출입국자료, 사회복지시설명부, 학적부 등 14종 자료 입수(11년~13년)
- 가족관계등록부, 사회시설 입소자 명부 등 2014년 기준 21 종 행정자료 입수(14년)

현 실 태
- 등록센서스 시산에 필요한 24종 행정자료 입수 완료 (401개 대학 학적부 포함)
  - 가족관계등록부는 개정된 통계법(제24조의2)에 근거하여 '15년 5월 입수
- 등록센서스 보완에 필요한 추가적인 행정자료 입수 완료(사회시설입소자명부(15.2월) 등)
- 2015년 기준 24종 행정자료 입수(15년)

향후 계획
- 2016년 기준 24종 행정자료 입수(16년~17년)

13

통계청

[참고] 24종 입수 행정자료

13개 공공기관의 협력을 기반으로 24종의 행정자료를 연계, 활용할 예정

행정자료명 (24종)	제공기관 (13개 기관)	관련부서	이용 목적
주민등록부	행정자치부	주민과	인구 · 가구 모집단 생성
도로명주소(국가기초구역) 자료		주소정책과	거처 모집단 생성
외국인등록부	법무부	이민정보과	인구(외국인) 모집단 생성
재외국민거소신고자 명부			인구 보완
외국국적동포거소신고자 명부			인구(외국인) 보완
단기체류자(90일 이상) 명부			인구(외국인) 모집단 생성
국적취득자 명부			인구, 가구 보완
출입국 자료		출입국정보화센터	인구(해외거주자) 보완
교정시설 수용자 명부		보안과	가구 보완
소년원 원생 명부		소년과	가구 보완
가족관계등록부	대법원	가족관계등록과	가구 보완
사회시설 및 이용자 명부	보건복지부	사회서비스자원과	인구, 집단가구 보완

14

통계청

[참고] 24종 입수 행정자료

13개 공공기관의 협력을 기반으로 24종의 행정자료를 연계, 활용할 예정

행정자료명 (24종)	제공기관 (13개 기관)	관련부서	이용 목적
군인 명부	국방부	인력정책	가구 보완
해외주재 공관원 명부	외교부	인사기획관실	인구, 가구 보완
의무경찰 명부	경찰청	경비과	가구 보완
의무해양경찰 명부	국민안전처	해양경비안전총괄과	가구 보완
의무소방대원 명부		소방정책과	가구 보완
전기시설(주택용) 자료	한국전력공사	영업운영부	주택(거주가능가구수) 보완
건축물대장	국토교통부	녹색건축과	주택 보완
주택공시가격자료		부동산평가과	주택 모집단 생성
학적부(대학)	교육부, 약 400개 대학	대학원지원과	인구(실거주지), 가구 보완
기숙시설(대학) 이용자 명부		(학사행정관리자협의회)	인구, 집단가구 보완
고용보험 피보험자 명주	한국고용정보원	고용보험팀	가구 보완
국민연금 사업장가입자 자료	국민연금공단	가입지원실	가구 보완

15

통계청

Ⅱ. 등록센서스 추진 전략 및 주요 추진 과제

2-2) 표준화된 행정자료 관리체계 구축 – 행정자료 DB구축

주요 내용 입수된 행정자료의 활용을 위한 「행정자료 통합DB」 구축

추진 경위
- 등록센서스 시스템 및 행정자료 통합 DB 구축 (10년)
- 입수 행정자료를 연도별로 통합 DB에 적재(11~14년)
- 등록센서스 시스템 기능 고도화 (15년)
  - 건축물 이력관리 기능, 가구주택기초조사 활용 주택DB보완 기능 개발
- 2014년 기준 행정자료 통합 DB 적재 (15년)

현 실 태
- 등록센서스 시스템 및 행정자료 통합관리시스템 개발 완료
  - 행정자료 기반으로 자료 표준화, 자동연계, 전산 내검 및 집계 기능 구현
- 등록센서스 추진을 위한 행정자료 DB구축 완료

향후 계획
- 2015년 기준 행정자료 통합 DB 적재 (16년)

16

통계청

Ⅱ. 등록센서스 추진 전략 및 주요 추진 과제

2-2) 표준화된 행정자료 관리체계 구축 – 행정자료 표준화

주요 내용

서로 다른 행정자료의 항목을 표준화 하여 관리하고,
행정자료간 연계를 위한 일관성 있는 Key (식별번호, 주소) 확보

추진 경위

- 행정자료 주소표준화(주소 분리, 전산표준화, 표준화 수작업, 내검) (10년 ~14년)
- 도로명주소 정보 표준화를 위한 시스템 구축 완료 (12년)

현 실 태

- 기존 행정자료(10~14년) 표준화 및 코드화 완료
  - 표준화 대상 : 새주소(행정동, 도로명코드, 지하여부, 건물본번, 건물부번), 구주소(본·부번지), 대지구분, 건물명, 층구분, 동명, 호명
  - 코드화 대상 : 행정동, 성별, 만나이, 거처종류, 거주기간 등

표준화 및 코드화	대상 행정자료
주소표준화	주민등록부, 건축물대장, 사회시설명부, 주택공시가격자료 등 (10종)
성별코드화	주민등록부, 가족관계등록부, 외국인등록부, 학적부 (4종)

향후 계획

- 2015년 기준 행정자료 주소 표준화 및 코드화(16년. 6월)

17

통계청

[참고] 자료 표준화 및 코드화 사례

표준화 사례

자료A	청운아파트	103동	1712호
자료B	청운APT	103	1712
자료C	청운아파트 103동 1712호		
자료D	청운아파트	103-1712	
자료A	엔에스빌라	가동-B01호	
자료B	NS빌라 가동 지하 101호		
자료C	엔에스빌라	가동	지하101호

↓

자료A	청운아파트	103	1712	20*
자료B	청운아파트	103	1712	20
자료C	청운아파트	103	1712	20
자료D	청운아파트	103	1712	20
자료A	엔에스빌라	가	101	10
자료B	엔에스빌라	가	101	10
자료C	엔에스빌라	가	101	10

※ 10 : 지하, 20 : 지상

코드화 사례

	이름	성별	행정구역
자료A	홍길동	남자	대전광역시 서구 만년동
자료B	홍길동	남	대전광역시 서구 만년동
자료C	홍길동	M	대전광역시 서구 만년동
자료D	홍길동	남성	대전광역시 서구 만년동

↓

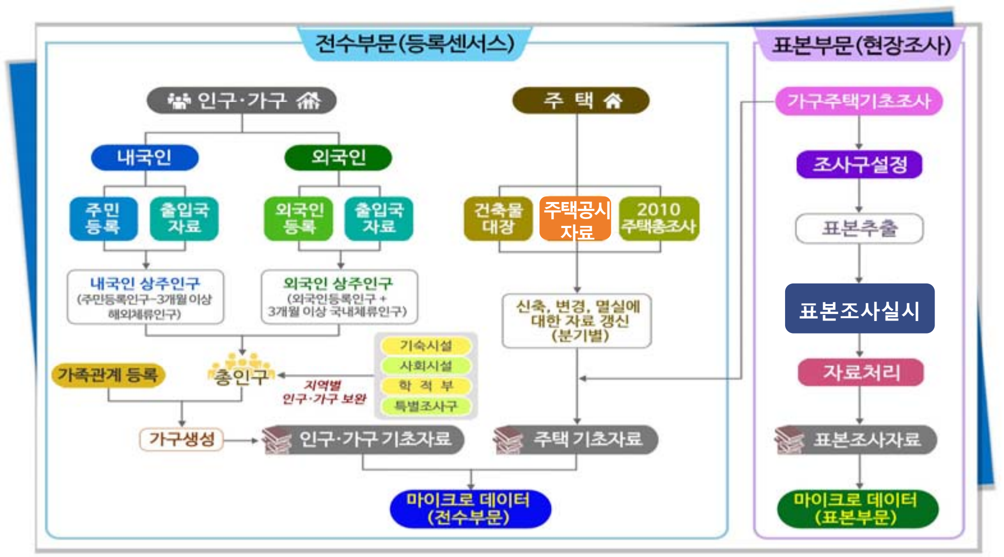
	이름	성별	행정구역
자료A	홍길동	1	00001234
자료B	홍길동	1	00001234
자료C	홍길동	1	00001234
자료D	홍길동	1	00001234

18

Ⅱ. 등록센서스 추진 전략 및 주요 추진 과제

통계청

2-3) 행정자료를 활용한 자료생산 – (개요) 2015 인구주택총조사 작성 체계



Ⅱ. 등록센서스 추진 전략 및 주요 추진 과제

통계청

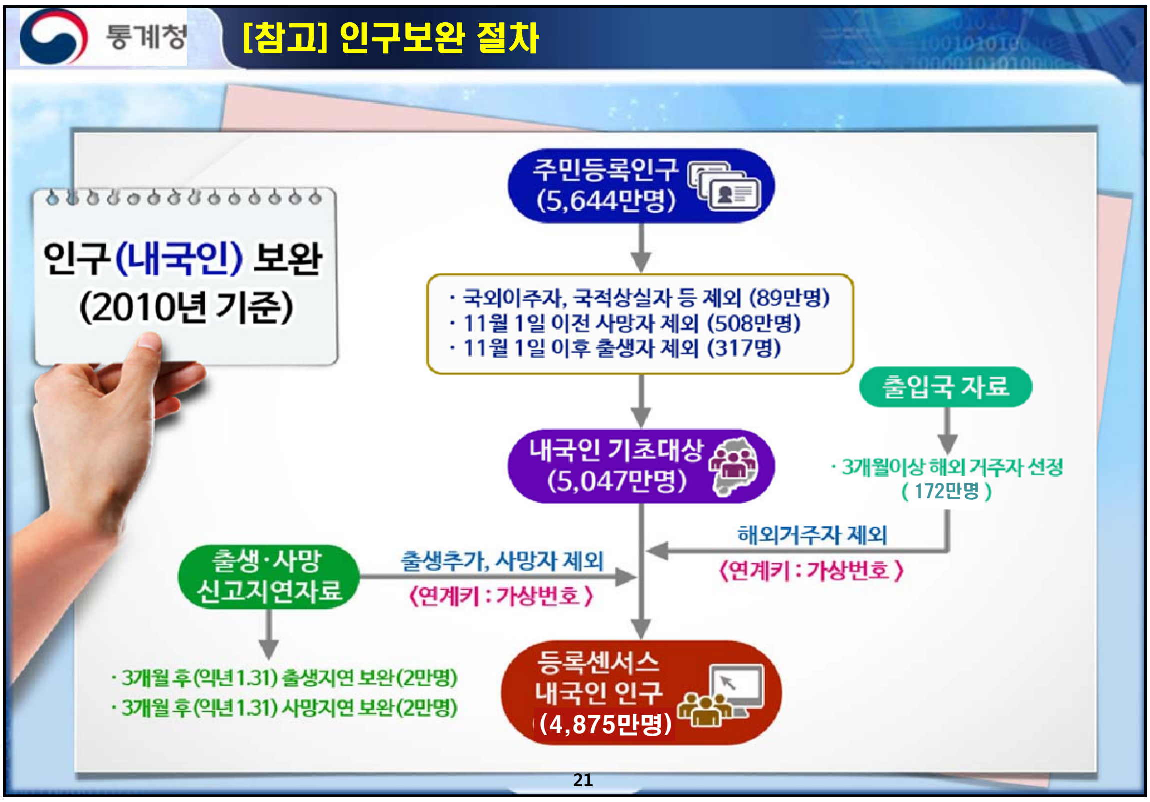
2-3) 행정자료를 활용한 자료생산 – 등록센서스 인구·가구·주택 결과 생성

주요 내용 — **주민등록부, 외국인등록부, 출입국 자료 등 관련 행정자료를 연계하여 포괄범위를 조정하고, 지역별 자료 생산**

추진 경위
- 학계 전문가와 등록센서스 공동연구 및 방법론 개발 (10~13년)
- 2010~2013년 등록센서스 인구, 가구, 주택 결과 생성 (12~14년)

현 실 태
- 2010~2013년 등록센서스 시계열 안정성 검토 및 품질 점검 (15년)
- 2014년 가구주택기초조사 결과를 활용하여 등록센서스 보완 (15년)
- 2010~2014년 등록센서스 시산 결과 작성 (15년)

향후 계획
- 2015년 등록센서스 결과 생성 및 공표 (16년)
- 2016년 등록센서스 결과 추가 공표 항목 검토 (16년)
  * 1년전 거주지, 타지주택 소유 여부, 주인가구 여부


통계청
[참고] 인구보완 절차
인구(내국인) 보완
(2010년 기준)
주민등록인구
(5,644만명)
· 국외이주자, 국적상실자 등 제외 (89만명)
· 11월 1일 이전 사망자 제외 (508만명)
· 11월 1일 이후 출생자 제외 (317명)
출입국 자료
· 3개월이상 해외 거주자 선정
(172만명)
내국인 기초대상
(5,047만명)
해외거주자 제외
〈연계키 : 가상번호〉
출생·사망
신고지연자료
출생추가, 사망자 제외
〈연계키 : 가상번호〉
· 3개월 후(익년 1.31) 출생지연 보완(2만명)
· 3개월 후(익년 1.31) 사망지연 보완(2만명)
등록센서스
내국인 인구
(4,875만명)
21


통계청
[참고] 인구보완 절차
외국인 등록
(90만명)
외국국적동포 거소신고자
(8만명)
인구(외국인) 보완
(2010년 기준)
외국인 기초대상
(98만명)
출입국 자료
· 3개월이상 해외거주 외국인 선정(12만명)
(외국인 등록 9만명, 거소신고자 3만명)
익년 외국인등록자료
· 외국인등록 지연신고자(1만명)
해외거주자 제외
〈연계키 : 성명+생년월일+국적+여권가상번호〉
단기불법 체류자
불법체류자 추가
· 단기불법체류자로 3개월이상
체류한 외국인(9만명)
등록센서스
외국인 인구
(96만명)
22

통계청

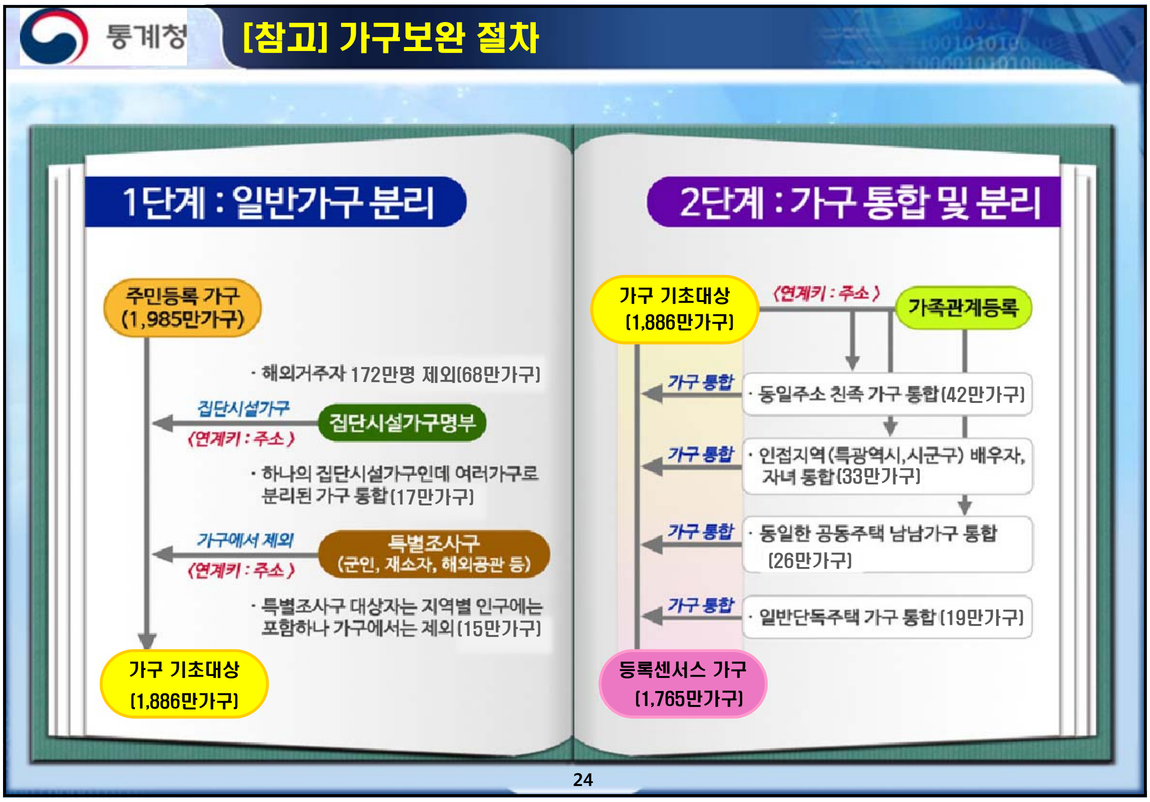
< 등록센서스 시산 결과 - 인구 >

등록센서스 인구 (2010년 비교)

(단위 : 만명, %)

	행정자료 인구(A)	등록센서스 인구(B)	추계인구 (C, 11.1)	총조사 인구(D)	차이				차이율
					(E=A-D)	(F=A-B)	(G=B-D)	(H=B-C)	(H/B*100)
계	5,145	4,971	4,953	4,858	287	174	113	18	0.4
내국인	5,048	4,875	4,855	4,799	248	173	76	19	0.4
남자	2,529	2,440	2,428	2,384	145	89	56	12	0.5
여자	2,518	2,435	2,427	2,415	103	83	20	7	0.3
성비	100.4	100.2	100.0	98.7	1.7	0.2	1.5	0.3	-
외국인	98	96	98	59	39	2	37	-1	-2.1

23



통계청

[참고] 가구통합 사례

주민등록세대와 주소, 가족관계 정보를 이용한 가구 결정

· 가구(취사, 취침 등 생계를 같이하는 생활단위 : 인구주택총조사 규칙) 와 세대(생계를 같이하는 집단 : 주민등록사무편람) 는 동일한 개념임

가구보완 사례 (동일한 거주지에서 세대가 분리된 3세대 → 1가구로 통합)

주민등록 자료				1가구 통합	가족관계 활용(가족관계등록부 연계)			
이름	세대주	세대주와의 관계			이름	성별	나이	가구주와의 관계
이 AA	이 AA	세대주	세대		이 AA	1	41	가구주(남편)
임 BB	임 BB	세대주	세대		임 BB	2	38	부인
이 CC	임 BB	아들			이 CC	1	9	아들
이 DD	임 BB	딸			이 DD	2	8	딸
박 EE	박 EE	세대주	세대		박 EE	2	60	장모

25

통계청

< 등록센서스 시산 결과 –가구 >

등록센서스 가구 2010년 비교

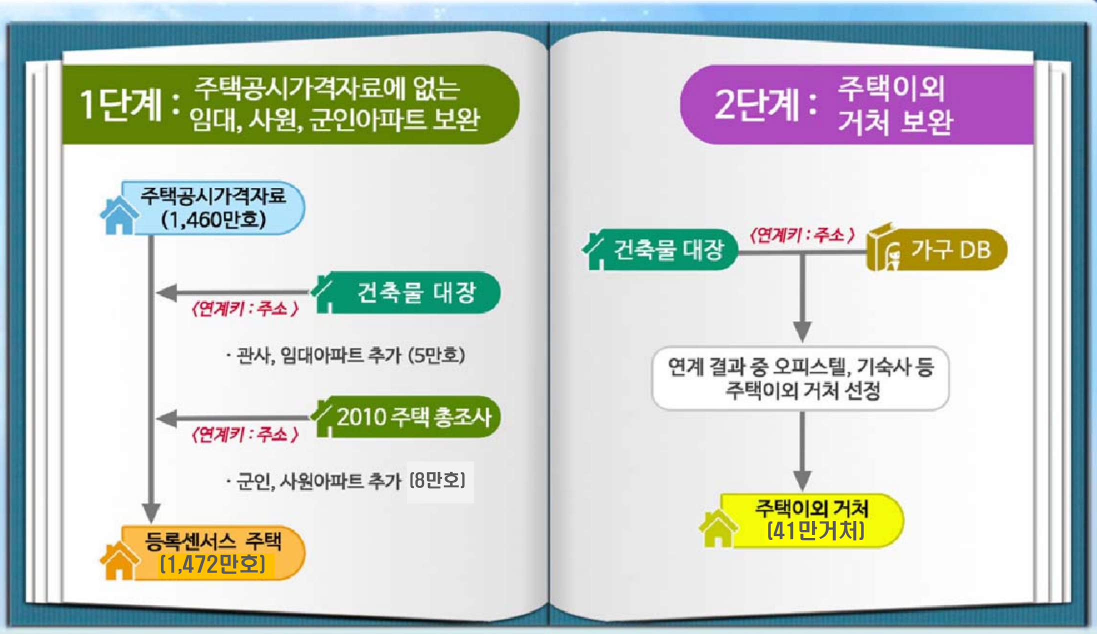
(단위 : 만가구, %)

	주민등록(A)		등록센서스(B)		주계가구(C)		총조사(D)		주계와차이		총조사와차이	
		구성비		구성비		구성비		구성비	(E=B-C)	E/B	(F=B-D)	F/B
계	1,984	100.0	1,765	100.0	1,747	100.0	1,734	100.0	18	1.0	31	1.8
1인가구	662	33.3	422	23.9	422	24.2	414	23.9	0	0	8	1.9
2인가구	378	19.1	433	24.5	426	24.4	421	24.3	7	1.6	12	2.8
3인가구	365	18.4	380	21.5	372	21.3	370	21.3	8	2.1	10	2.6
4인가구	423	21.3	388	22.0	388	22.2	390	22.5	0	0	-2	-0.5
5인이상	157	7.9	142	8.1	138	7.9	140	8.1	4	2.8	2	1.4

26

통계청

[참고] 주택보완 절차



※ 2014년 가구주택기초조사에서 모든 주택에 대해 확인 후 보완 완료

– 보완대상 : 무허가주택, 비닐하우스 등 주택이외 거처, 가구와주택 미연계 주소 등

27

통계청

< 등록센서스 시산 결과 – 주택 >

등록센서스 주택 2010년 비교

(단위 : 만호, %)

구분		주택공시(A)		등록센서스(B)		총조사(C)		총조사와차이	
			구성비		구성비		구성비	(F=B-C)	F/B
계		1,460	100.0	1,472	100.0	1,468	100.0	4	0.3
단독주택		385	26.3	403	27.4	409	27.9	-6	-1.5
	일반단독	300	20.5	283	19.2	285	19.4	-2	-0.7
	다가구	54	3.7	83	5.6	89	6.0	-6	-7.2
	영업겸용	31	2.1	37	2.5	36	2.4	1	2.7
아파트		847	58.0	854	58.0	858	58.4	-4	-0.5
연립		46	3.2	45	3.1	54	3.7	-9	-20
다세대		150	10.3	149	10.1	131	9.0	18	12.1
비거주용		32	2.2	21	1.4	16	1.1	5	23.8

28

통계청 **Ⅱ. 등록센서스 추진 전략 및 주요 추진 과제**

2-4) 개인정보보호 및 공감대 형성 – 행정자료 보안강화

주요 내용

- 주민등록번호 등 개인식별번호는 입수단계에서 암호화 과정 후 통계목적용 가상번호로 변환 · 삭제
- 등록센서스용 행정자료 DB에는 통계목적용 가상번호만 있고 개인식별번호는 없음

추진 경위

- 자료 입수단계부터 대체번호 생성 및 제공방안 마련 (14년)
- 행정자료를 제공 할 때부터 대체번호를 생성할 수 있는 프로그램 및 자료수집시스템 개발 및 지원 (14년~)

현 실 태

- 행정자료간 연계를 위해 별도의 통계목적용 가상번호(10자리) 생성 및 부여
  - 부여된 가상번호는 주민등록번호 또는 외국인등록번호로 복원 불가
- 주민등록번호, 외국인등록번호 등 개인정보가 삭제된 행정자료만 DB로 관리 운영

향후 계획

- 등록센서스 결과공표 이후 입수된 행정자료 DB 삭제(16년~)
  - 입수 행정자료는 삭제되고 등록센서스 DB만 존재(매년 자료 공표 후)

29

통계청 **[참고] 안전한 자료 관리**

주민등록부 / 가족관계 등록부 / 출입국자료 / 건축물대장 소유자정보 / 외국인등록부 / 학적부

주민등록번호 외국인등록번호 → 오류 점검 → DI* 생성 모듈로 암호화 → 통계목적용 가상번호 변환

암호화된 대체번호와 가상번호 매핑표

* DI (중복가입방지정보) : Duplication Information

암호화된 대체번호	통계목적용 가상번호
MC0GCCqGSIb3DQIJAyEAbDSAY…..Cw+3Li4ENiRKJhDHsZXXXXX=	1234567
MC0GCCqGSIb3DQIJAyEAu36el/W…..DKJhl8pUqcN2SeRaXXXXX=	2345678
MC0GCCqGSIb3DQIJAyEAselHV37…..ghViiONXgzIQdw8GXXXXX=	3456789

DI* 생성 모듈로 보안성 높은 단방향 암호화 (복원불가)
· 큐브원

암호화된 대체번호(64자리)를 통계목적용 가상번호(10자리)로 변환

주민/외국인등록번호 즉시 삭제

30

통계청

Ⅱ. 등록센서스 추진 전략 및 주요 추진 과제

2-4) 개인정보보호 및 공감대 형성 – 공감대 형성

주요 내용
- 효율적인 소통·협력을 위해 12개 관계기관과 협의체 운영
- 등록센서스 결과 활용을 위한 대국민 소통 채널 유지

추진 경위
- 등록센서스 설명회 실시, 홍보리플릿 제작 및 배포 (12년~)
- 행정자료 제공기관 대상 등록센서스 협의체 구성 및 운영(14년~)

현 실 태
- 등록센서스 설명회를 매년 실시 중
- 24종 행정자료 관계기관과 자료제공 협의체 구성 및 협의체 회의 운영
- 정부3.0 체험마당(15.4월), 공중파 프로그램(5월), 기획보도(9월) 등

향후 계획
- 정부3.0 체험마당(16.6월, 11월) 참여, 학술대회 기획세션 공동 개최
- 2015년 기준 등록센서스 결과 활용 홍보(16년)
  - KOSIS를 통한 결과 공표 및 MDIS를 통한 마이크로데이터 제공

31

통계청

[참고] 등록센서스 협의체

등록센서스 협의체 구성 목적

❖ 성공적인 「2015 등록센서스」 추진을 위해서는 행정자료의 안정적인 확보가 필요

❖ 자료 입수의 원활함을 위해 자료 제공기관 담당자를 중심으로 협의체를 구성

참여 기관 및 부서

※ 12개 기관 18개 부서

기관명	부서명
행정자치부	주민과
	주소정책과
법무부	이민정보과
	출입국정보화센터
	보안과
	소년과
대법원	가족관계등록과
보건복지부	사회서비스지원과
국방부	인력정책과

기관명	부서명
외교부	인사기획관실
경찰청	경비과
국민안전처	해양경비안전총괄과
	소방정책과
한국전력공사	영업운영부
국토교통부	녹색건축과
	부동산평가과
한국감정원	아파트부동산공시부
교육부(약 400개 대학)	대학학사제도과(학사행정관리자협의회)

32

통계청

Ⅲ. 기대효과

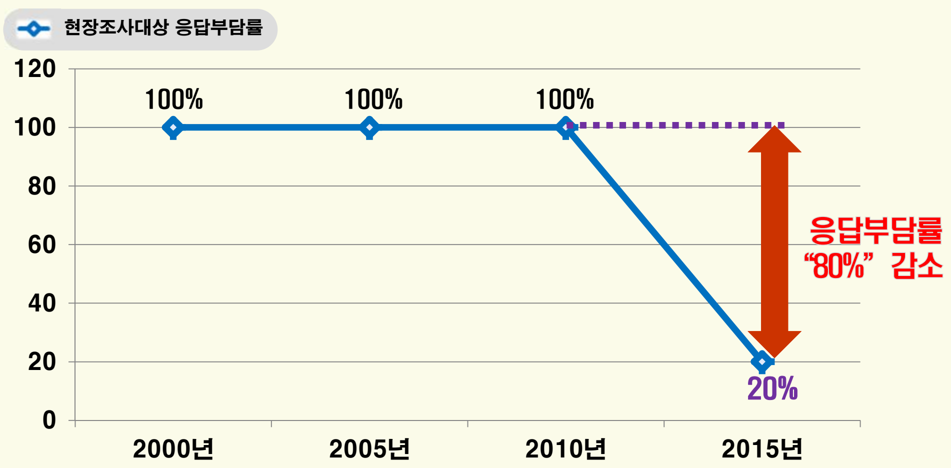
1. 국민의 응답부담 경감
2. 국가통계 품질 제고
3. 비용절감

통계청

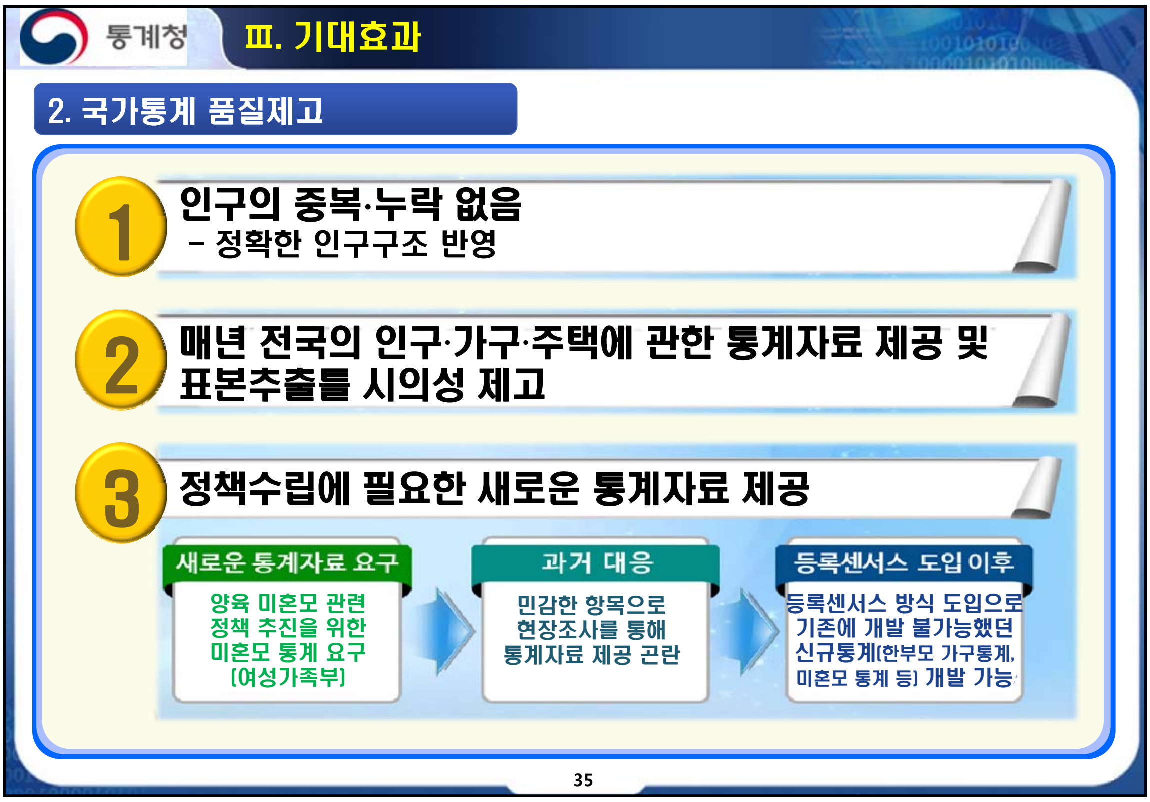
Ⅲ. 기대효과

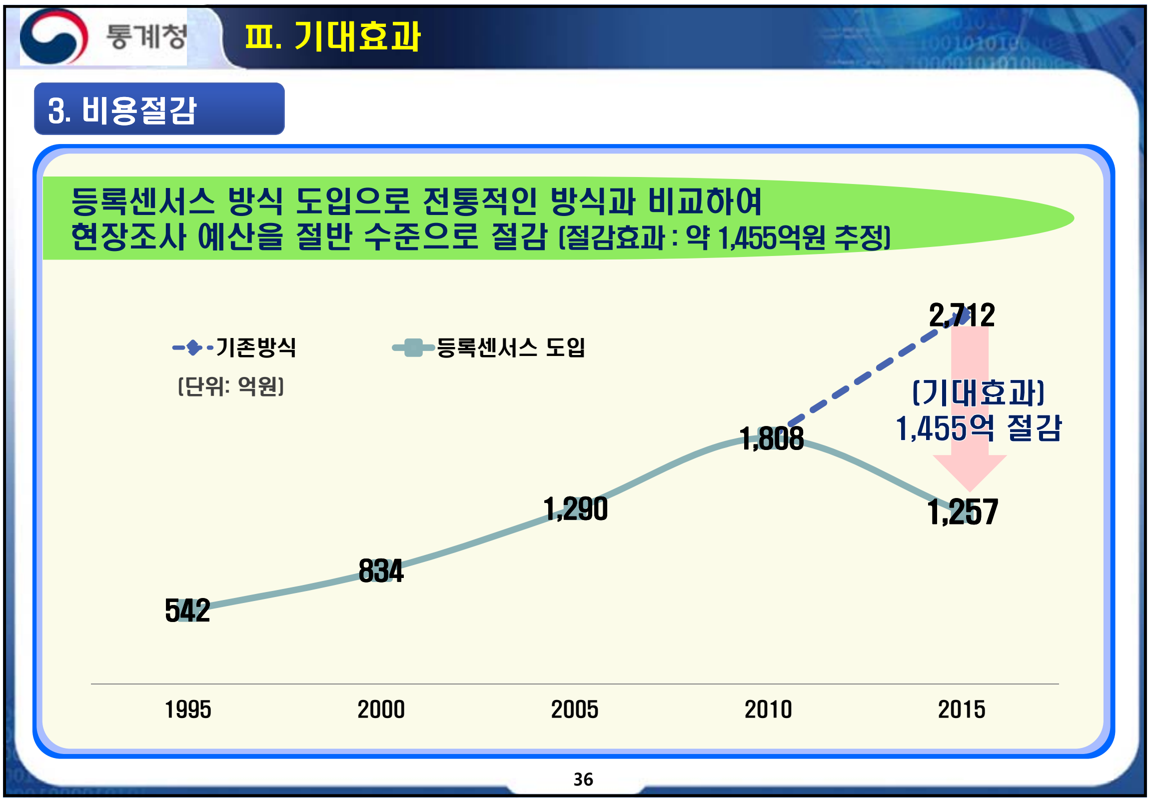
1. 국민의 응답부담 경감

응답부담율 100% → 20% 경감되어
국민의 응답에 대한 부담감 최소화



34


통계청
Ⅲ. 기대효과
2. 국가통계 품질제고
1
인구의 중복·누락 없음
– 정확한 인구구조 반영
2
매년 전국의 인구·가구·주택에 관한 통계자료 제공 및
표본추출틀 시의성 제고
3
정책수립에 필요한 새로운 통계자료 제공
새로운 통계자료 요구
양육 미혼모 관련
정책 추진을 위한
미혼모 통계 요구
(여성가족부)
과거 대응
민감한 항목으로
현장조사를 통해
통계자료 제공 곤란
등록센서스 도입 이후
등록센서스 방식 도입으로
기존에 개발 불가능했던
신규통계(한부모 가구통계,
미혼모 통계 등) 개발 가능
35


통계청
Ⅲ. 기대효과
3. 비용절감
등록센서스 방식 도입으로 전통적인 방식과 비교하여
현장조사 예산을 절반 수준으로 절감 (절감효과 : 약 1,455억원 추정)
기존방식
등록센서스 도입
(단위: 억원)
542
834
1,290
1,808
2,712
1,257
(기대효과)
1,455억 절감
1995
2000
2005
2010
2015
36


감사합니다

통계청
Statistics Korea

기업등록부 구축 및 활용

2016. 8.

경제통계국

목 차

1. 구축배경

2. 추진경과

1) 포괄범위 설정
2) 조사 및 행정자료 결합
3) 통계단위 설정 및 그룹핑
4) 산업분류 대체 및 결정
5) 고유번호 부여 및 관리
6) 기타사항

3. 결과활용

1. 구축배경

정부3.0 이행

조사 및 행정자료를 융합한
새로운 경제통계프레임 구축

→ 핵심가치(협업/공유) 이행

단일 모집단 구축

통계조사별 모집단 및
행정자료별 DB를 통합

→ 경제통계 모집단 일원화

효율성 제고

차기 등록센서스 기반 마련
각종 통계조사 간소화를 통한

→ 응답부담 경감, 소요예산 절감

조사한계 대안 모색

사업체 비밀노출 우려로 불응 증가

고정장소가 없는 다양한 형태의
사업체 포착 한계 극복

2. 추진경과

포괄범위 설정

공식/비공식 및 영리/비영리에 관계없이 조사자료와 행정자료에 의해
수집 가능한 모든 자료 포함

조사자료

등록 및 합법 여부에 관계없이
고정장소가 있는 모든 사업체
정보 제공

▸ 재래시장, 노점 등 소규모
무등록 사업체도 포함

자료융합

행정자료

고정장소가 없거나, 가구 내 등
현장조사 시 포착 곤란 사업자
정보 제공

▸ 전자상거래, 운송업 등
고정장소가 없는 사업체도 포함

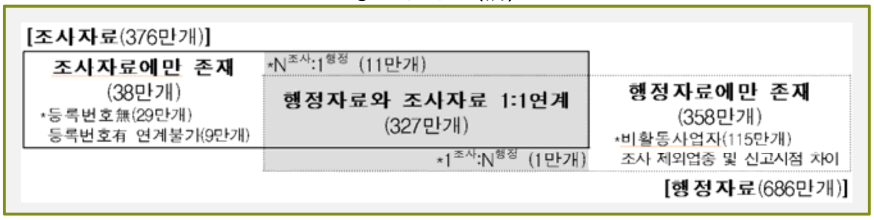
2 조사 및 행정자료의 결합

(구축 前)

- 연계율: 87.0% (조사자료 기준)
- 미연계: 조사자료(38만개), 행정자료(358만개)*

* 매출액과 종사자가 없는 비활동 사업자(115만개)

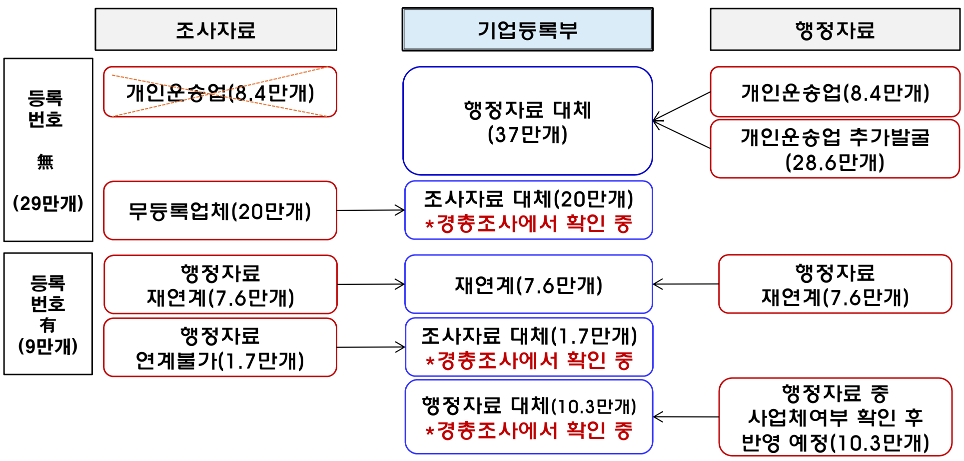
기업등록부 구축(前) 모습



조사/행정 미연계자료(여집합) 축소를 위한 **행정자료 대체방안 및 연계정보 재검토 추진**

(연계 및 검토과정)

- 개인운송업체 행정자료 대체, 등록번호가 없는 사업체의 무등록 여부, 등록번호가 있는 사업체의 재연계 등 연계율 제고 검토



(구축 後)

- 연계율: 87.0%→ 94.2% (조사자료 기준)
- 미연계: 조사자료(38만개→21.7만개) 행정자료(358만개→303만개)

기업등록부 구축(後) 모습

- 물리적 장소가 있는 조사대상 사업체(무등록 포함)
- 물리적 장소가 없는 조사대상 제외 사업자
- 비활동사업자 모두 포괄

3 통계단위 설정 및 그룹핑

(사업자)

조사/행정자료를 연계융합하여 수록한 모든 단위를 별개의 사업자로 구분

(사업체)

일정한 물리적 장소 또는 지역 내의 하나의 단일 또는 주된 경제활동을 독립적으로 수행하는 단위를 별개의 사업체로 구분
* 하나의 장소에 동일인이 여러 개 사업자등록을 한 경우→ 1개 사업체

(기업체)

재화 및 서비스를 생산하는 법적 또는 제도적 최소 경영단위로 자원배분의 의사결정권이 있고, 재무제표를 독립적으로 관리하는 단위를 별개의 기업체로 구분
* 법인등록번호(법인)/개인대체식별번호(개인)을 이용하여 그룹핑

④ 산업분류 대체 및 결정

- 조사자료와 행정자료의 산업분류 일치율은 95%(연계자료의 대분류 기준)

대분류	중분류	소분류	세분류	세세분류
(95%)	(89.6%)	(89.1%)	(88.8%)	(88.4%)

1) 사업체

– (적용) 조사자료의 산업분류로 대체 적용
– (사유) 행정자료는 자진신고에 기반하나, 조사자료는 현장방문 결과이기 때문에 정확성이 상대적으로 높을 것으로 판단하여 대체

* 중앙행정기관 분류 사례: 행정자료(부동산임대업) / 조사자료(공공행정)

2) 기업체

– (적용) 산하 사업체 중 매출액이 큰 산업을 기업체의 주산업으로 결정

통계단위 구분 및 산업분류 결정(예시)

사업자 단위

ID	명칭	(통계단위 구분 정보) 사업자	사업체	기업체	기업집단	지역 코드	산업	(사업자 속성정보) 종사자	매출	수출	수입 …
10001	에스케이하이닉스	1	1	0	0	1123063	장비 도매업	100	300	0	0 …
10002	에스케이하이닉스	1	1	0	0	3102357	연구개발	50	100	0	0…
10003	에스케이하이닉스	1	1	1	0	3121012	전자집적 회로제조	600	800	700	400 …
10004	에스케이하이닉스	1	1	0	0	3304358	전자집적 회로제조	500	700	600	300 …
10005	에스케이하이닉스	1	0	0	0	3304358	부동산 임대업	30	50	0	0 …

⇒ 기업 내 사업자별(5개) 정보 제공

20001	에스케이텔레콤㈜	1	1	1	1	1102055	무선통신	500	1,200	0	0 …

사업체 단위

ID	명칭	사업자	사업체	기업체	기업집단	지역 코드	산업	종사자	매출	수출	수입	…
10001	에스케이하이닉스	1	1	0	0	1123063	장비 도매업	100	300	0	0	…
10002	에스케이하이닉스	1	1	0	0	3102357	연구개발	50	100	0	0	…
10003	에스케이하이닉스	1	1	1	0	3121012	전자집적 회로제조	600	800	700	400	…
10004	에스케이하이닉스	1	1	0	0	3304358	전자집적 회로제조	530	750	600	300	…
10005	에스케이하이닉스	1	0	0	0	3304358	부동산 임대업	30	50	0	0	…

(통계단위 구분 정보: 사업자/사업체/기업체/기업집단) (사업체 속성정보)

⇒ 동일장소의 사업자를 사업체로 그룹핑(2개→1개) 하여, 사업체별(4개) 정보 제공

ID	명칭	사업자	사업체	기업체	기업집단	지역 코드	산업	종사자	매출	수출	수입	…
20001	에스케이텔레콤㈜	1	1	1	1	1102055	무선통신	500	1,200	0	0	…

기업체 단위

ID	명칭	사업자	사업체	기업체	기업집단	지역 코드	산업	종사자	매출	수출	수입	…
10001	에스케이하이닉스	1	1	0	0	1123063	장비 도매업	100	300	0	0	
10002	에스케이하이닉스	1	1	0	0	3102357	연구개발	50	100	0	0	…
10003	에스케이하이닉스	1	1	1	0	3121012	전자집적 회로제조	1,280	1,950	1,300	700	…
10004	에스케이하이닉스	1	1	0	0	3304358	전자집적 회로제조	500	700	600	300	…
10005	에스케이하이닉스	1	0	0	0	3304358	부동산 임대업	30	50	0	0	…

(통계단위 구분 정보: 사업자/사업체/기업체/기업집단) (기업체 속성정보)

⇒ 동일기업내 사업자 및 사업체를 그룹핑하여, 기업체별(1개) 정보 제공

ID	명칭	사업자	사업체	기업체	기업집단	지역 코드	산업	종사자	매출	수출	수입	…
20001	에스케이텔레콤㈜	1	1	1	1	1102055	무선통신	500	1,200	0	0	…

5 고유번호 부여 및 관리

(현황: 조사및행정자료의 고유번호)

조사자료는 별도의 사업체고유번호로 관리
행정자료는 관리사업자등록번호, 법인등록번호로 관리

(변경: 기업등록부의 고유번호)

기업등록부의 고유번호 체계=사업자등록번호 및 체계 유지
17년부터 각종 통계조사의 고유번호=기업등록부의 고유번호로 단일화

[조사자료에만 존재]

[1:1 연계 및 행정자료에만 존재하는 자료]
사업자등록번호→ 기업등록부 고유번호로 부여
(예) 123-12-12345-0000

코드체계: A+시도(2)-개인/법인(2)-연도(2)+일련번호(3)-종된일련번호(4)
(예) A11-00-16001-0000
* 행정자료에 없는 자료이기 때문에 신규 부여에 따른 사업자등록번호와 중복되지 않으면서 구분되도록 설계

6 기타사항

경제총조사 현장조사 추가 확인 사항(45.3만개)

1) 등록번호가 없는 사업체(29만개)

조사자료 중 사업자등록번호가 없는 사업체에 대한 무등록(20만개)
등록번호 오류 여부(9만개) 재확인 중

2) 전자상거래업체(4.9만개)

행정자료에만 있는 전자상거래 사업자 등 가구 이외의 고정장소 유무

* 15만개 사업자 중 소재지 형태 사전 확인 후 선정한 4.9만개 대상

3) 추가 사업체 발굴(11.4만개)

동일인이 동일장소에 여러 개 사업자등록한 일정규모 이상 사업자(1.1만개)*, 미연계 사업자(10.3만개) 재확인

* 신규등록 및 폐업신고, 전년도 조사 시 변동 사항 등을 확인 후 선정한 사업자

사업자등록관리 체계화를 위한 국세청 협업

기업등록부가 제대로 구축/관리되기 위해서는 국세청의 사업자등록 단계부터 통계작성 목적에 부합되도록 신고 받을 필요

→ 제도적 보완을 위해 국세청 협업 추진 중

1) 협의사항: 사업자등록관련 서식개정, 업종분류 세분화, 자료제공 주기단축 등

* (서식) 사업자단위 과세자의 종된 사업장의 생산요소 항목 추가 및 DB화
  소득지급명세서의 종된 사업장일련번호 추가
  (업종분류 세분화) 업종분류와 산업분류 간 다중연계(1업종:N산업) 세분화
  (자료제공) 등록자료 제공주기(분기→월) 단축, 개인사업자의 동일인 정보 추가

2) 현재까지 추진경과

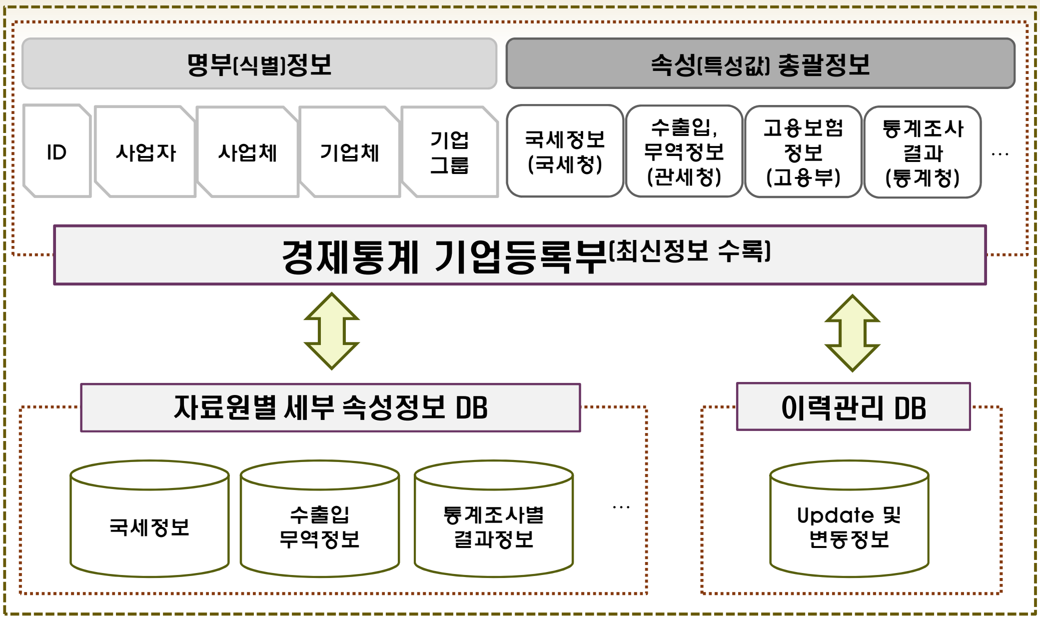
(사업자등록서식 및 업종분류) 업종분류와 산업분류가 1:1로 연계될 수 있도록 사업자등록시스템의 업종 색인*보완 및 세세분류 내용으로 신고되도록 개선

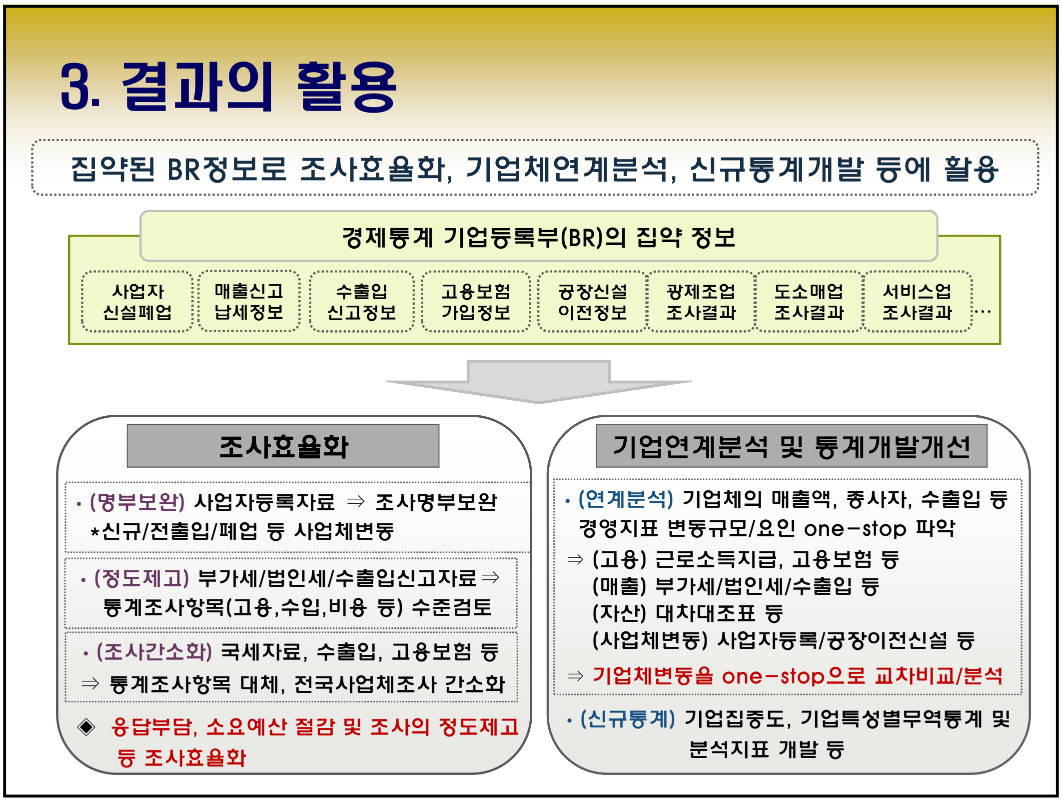
* 통계청에서 추가한 세세분류 종목 내용 반영

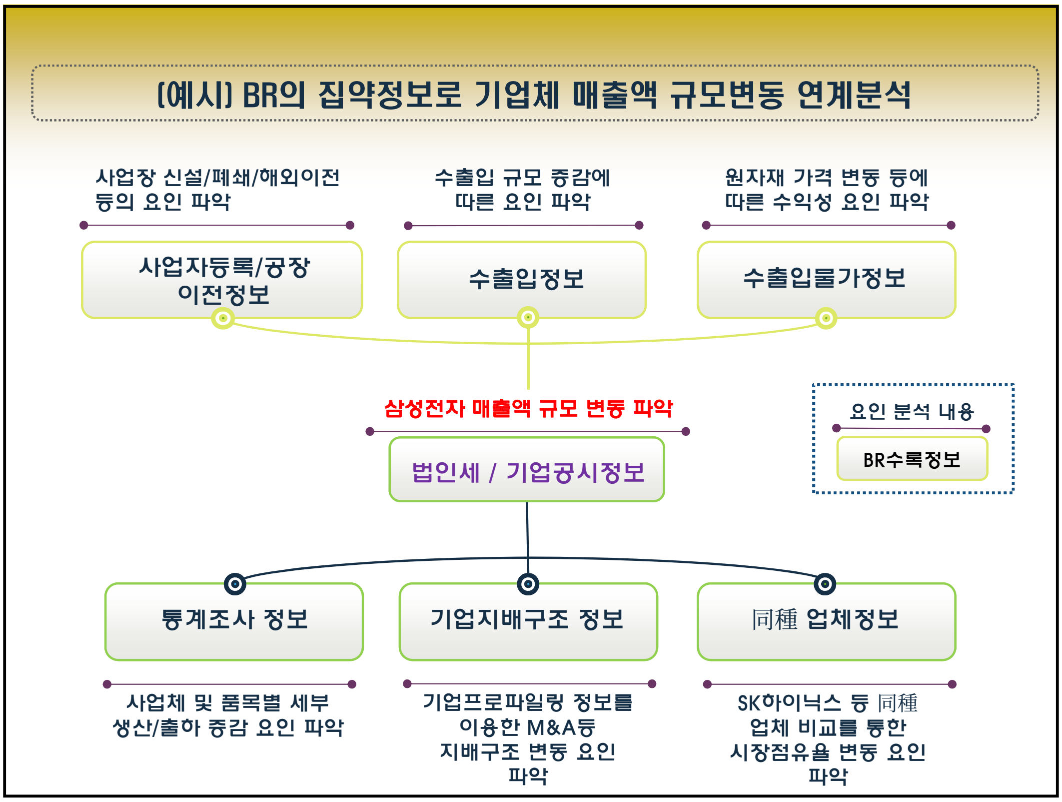
(근로소득지급명세서 서식) 기업의 응답부담을 고려하여 사업자단위 과세사업자의 근로소득 신고부터 우선 소득자별 종된 일련번호 수집

(자료제공) 등록자료 월주기로 입수가능, 개인사업자의 동일인 정보 추가입수

기업등록부의 구축 체계도




3. 결과의 활용
집약된 BR정보로 조사효율화, 기업체연계분석, 신규통계개발 등에 활용
경제통계 기업등록부(BR)의 집약 정보
사업자 신설폐업
매출신고 납세정보
수출입 신고정보
고용보험 가입정보
공장신설 이전정보
광제조업 조사결과
도소매업 조사결과
서비스업 조사결과
…
조사효율화
• (명부보완) 사업자등록자료 ⇒ 조사명부보완
*신규/전출입/폐업 등 사업체변동
• (정도제고) 부가세/법인세/수출입신고자료 ⇒ 통계조사항목(고용,수입,비용 등) 수준검토
• (조사간소화) 국세자료, 수출입, 고용보험 등 ⇒ 통계조사항목 대체, 전국사업체조사 간소화
◈ 응답부담, 소요예산 절감 및 조사의 정도제고 등 조사효율화
기업연계분석 및 통계개발개선
• (연계분석) 기업체의 매출액, 종사자, 수출입 등 경영지표 변동규모/요인 one-stop 파악
⇒ (고용) 근로소득지급, 고용보험 등
(매출) 부가세/법인세/수출입 등
(자산) 대차대조표 등
(사업체변동) 사업자등록/공장이전신설 등
⇒ 기업체변동을 one-stop으로 교차비교/분석
• (신규통계) 기업집중도, 기업특성별무역통계 및 분석지표 개발 등


[예시] BR의 집약정보로 기업체 매출액 규모변동 연계분석
사업장 신설/폐쇄/해외이전 등의 요인 파악
수출입 규모 증감에 따른 요인 파악
원자재 가격 변동 등에 따른 수익성 요인 파악
사업자등록/공장 이전정보
수출입정보
수출입물가정보
삼성전자 매출액 규모 변동 파악
법인세 / 기업공시정보
요인 분석 내용
BR수록정보
통계조사 정보
기업지배구조 정보
同種 업체정보
사업체 및 품목별 세부 생산/출하 증감 요인 파악
기업프로파일링 정보를 이용한 M&A등 지배구조 변동 요인 파악
SK하이닉스 등 同種 업체 비교를 통한 시장점유율 변동 요인 파악


통계청
Statistics Korea
경제통계 기업등록부 구축
경제국
경제통계기획과
경제총조사과
:
허브국
행정자료관리과
행정통계과
:
감사합니다

제 4 세 션

국가통계의 개방과 활용

- SGIS 오픈플랫폼 지자체 정책활용 / 김수진 주무관
- 수도권 도시철도 역 승하차 패턴과 역세권 인구/사회적 특성 분석 / 손홍규 교수
- 통계산업 국내외 현황 및 시사점 / 김지영 교수
- 통계산업 분류 / 이정희 사무관

- 제6회 국가통계발전포럼 -

SGIS 오픈플랫폼 지자체 정책활용

안산시

2016. 8.

발표자 김수진

순서

1. SGIS 오픈플랫폼 소개
2. 안산시 활용사례 (주거지분석맵)
3. 안산시 활용사례 (커뮤니티맵)
4. 향후계획

순서

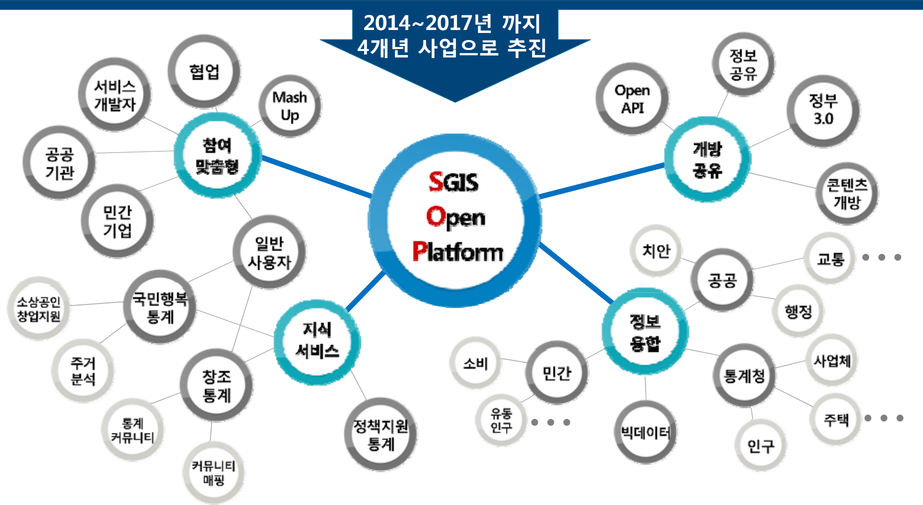
1. SGIS 오픈플랫폼 소개
2. 안산시 활용사례 (주거지분석맵)
3. 안산시 활용사례 (커뮤니티맵)
4. 향후계획

1-1. SGIS는?

인구주택, 사업체 등 센서스자료 중심의
국 가 통 계

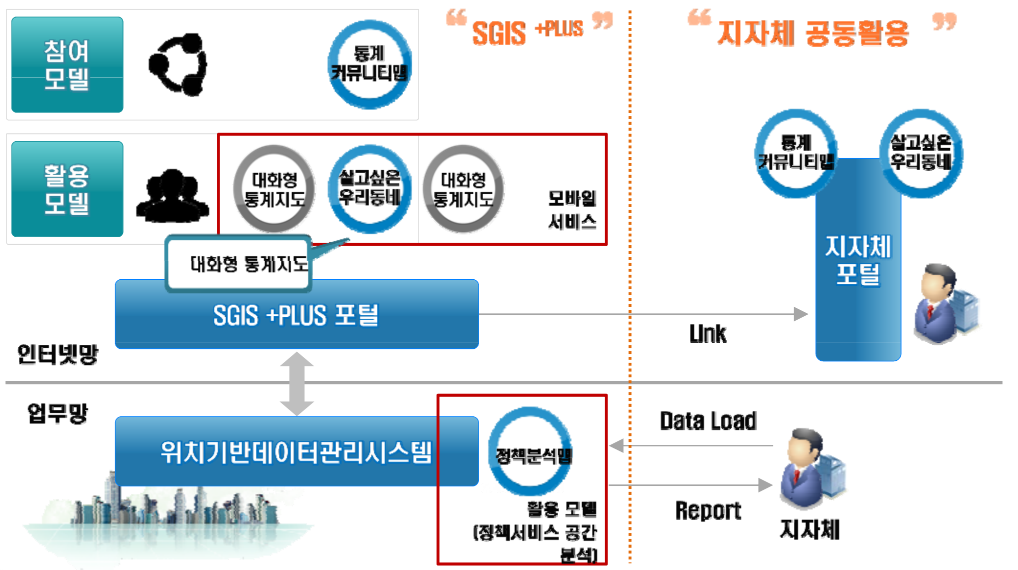
건강, 복지, 주거, 등 의사결정에 필요한
공공 · 민간 데이터

위치기반 통계데이터를 연계, 융합 할 수 있는 **오픈플랫폼**
(http://sop.kostat.go.kr)



1-2. 서비스 개요

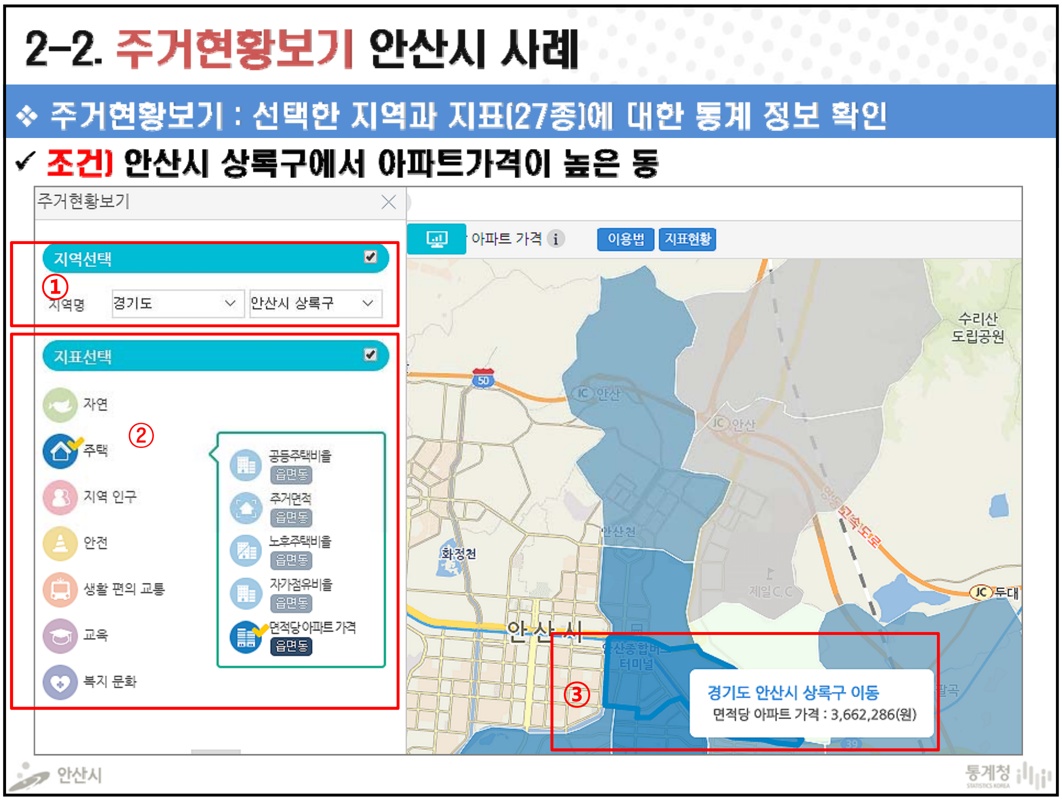
주거 선택, 주민 참여, 협업 의사결정 영역에서 기획된 서비스



순서

1. SGIS 오픈플랫폼 소개
2. **안산시 활용사례 (주거지분석맵)**
3. 안산시 활용사례 (커뮤니티맵)
4. 향후계획


2-1. '살고싶은 우리동네' 주거지분석맵 소개
❖ 자연, 주택, 지역인구, 안전, 생활편의교통, 교육분야 등 27개 지표 및
주거현황보기, 관심지점조회, 추천지역찾기 서비스 제공
주요지표
자연
대기오염도, 생활날씨, 녹지비율
대기오염도
시도
생활날씨
시도
녹지비율
읍면동
주택 및 인구
공동주택, 주거면적, 노후주택비율,자가점유비율, 아파트가격 등
공동주택비율
읍면동
주거면적
읍면동
노후주택비율
읍면동
자가점유비율
읍면동
면적당 아파트 가격
안전
화재안전, 교통 사고안전
화재 안전
시군구
교통사고 안전
시군구
생활편의
편의, 쇼핑, 외식, 대중교통정보
편의시설 수
읍면동
쇼핑시설 수
읍면동
외식시설 수
읍면동
교육
교원비율, 교육기관 및 학원비율
교원 1인당 학생 수
읍면동
고등교육기관 수
읍면동
학원 수
읍면동
복지문화
보육, 병의원, 복지, 문화체육시설
유치원 및 보육시설
읍면동
병의원 및 약국
읍면동
노인복지시설
읍면동
사회복지시설
읍면동
문화체육시설 수
읍면동
안산시
통계청


2-2. 주거현황보기 안산시 사례
❖ 주거현황보기 : 선택한 지역과 지표(27종)에 대한 통계 정보 확인
✓ 조건) 안산시 상록구에서 아파트가격이 높은 동
주거현황보기
지역선택
①
지역명
경기도
안산시 상록구
지표선택
자연
②
주택
지역 인구
안전
생활 편의 교통
교육
복지 문화
공동주택비율
읍면동
주거면적
읍면동
노후주택비율
읍면동
자가점유비율
읍면동
면적당아파트 가격
읍면동
아파트 가격
이용법
지표현황
수리산
도립공원
화정천
안 산 시
③
경기도 안산시 상록구 이동
면적당 아파트 가격 : 3,662,286(원)
안산시
통계청

2-3. 관심지점조회 안산시 사례

❖ 관심지점조회 : 학구도(초중고),생활편의교통(편의시설,쇼핑시설,외식시설,버스정류장,지하철),교육(학원, 대학 등),복지문화(보육시설, 복지시설, 도서관 등)

✓ 조건) 안산시 상록구에서 버스정류장이 집중되어 있는 곳

2-4. '살고싶은 우리동네' 추천지역찾기 소개

❖ 전입을 계획 중인 개인에게 주거 목적에 맞는 적합한 지역을 추천

통계로 찾은
살고싶은 우리동네 이용안내

- 살고싶은 우리동네는 SGIS plus 활용 사례로 사용자 조건에 맞는 주거지역을 추천해 주는 서비스입니다.
- 주거현황보기 선택한 지역(시도, 시군구)과 지표(27종)에 대한 통계정보를 확인할 수 있습니다.
- 추천지역찾기 생활편의 시설 현황, 교육환경, 자연환경 등 27종의 주거지 선정 조건에 해당하는 추천지역리스트 및 통계정보를 확인할 수 있습니다.

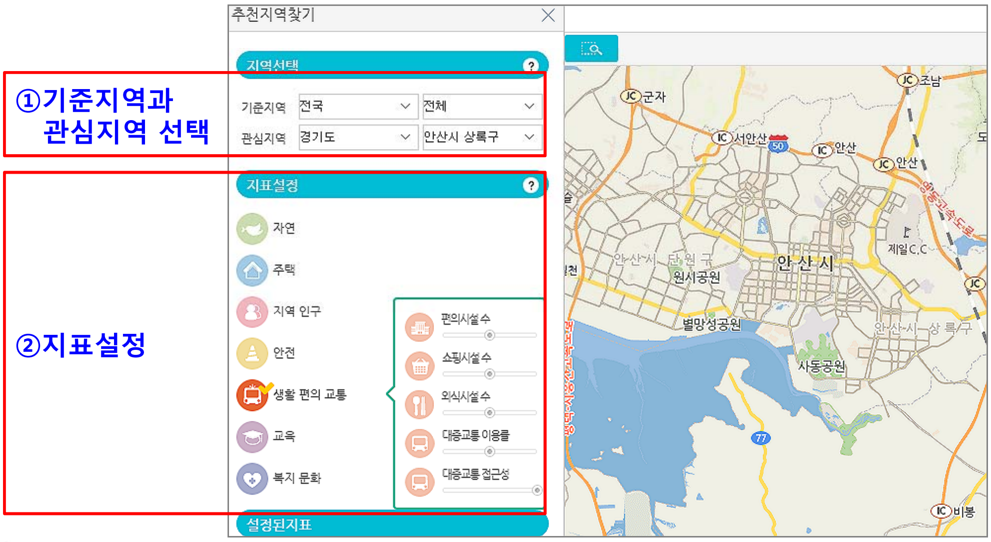
1. 지역설정	2. 지표 설정	3. 추천지역 선택	4. 추천지역 정보 조회	5. 소지역 정보 조회

기준지역과 관심지역을 선택합니다.	주거지와 관련되는 조건 중에서 원하는 세부지표를 설정합니다.	지도상에서 추천된 지역을 선택합니다.	읍면동에서 해당하는 추천지역 정보를 볼 수 있습니다.	추천지역에서 소지역 정보를 조회 할 수 있습니다.

2-5. 추천지역찾기 안산시 사례 ①

❖ 초등생과 유치원생 두 자녀를 둔 가정이 안산시 상록구 전입시 추천 지역

✓ 조건) 유치원과 병원이 가까운 지역이면서 학원과 대중교통 이용이 쉬운곳



안산시 통계청

2-5. 추천지역찾기 안산시 사례 ②

❖ 초등생과 유치원생 두 자녀를 둔 가정이 안산시 상록구 전입시 추천 지역에 지표 설정

✓ 조건) 유치원과 병원이 가까운 지역이면서 학원과 대중교통 이용이 쉬운곳

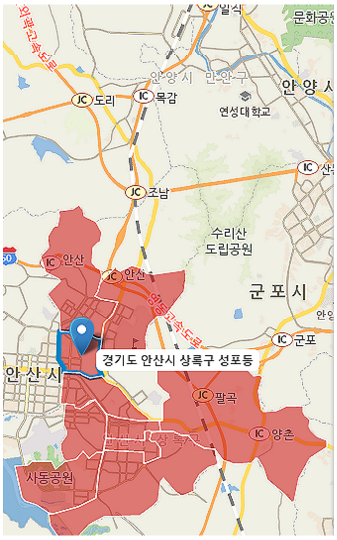
✓ 지표) 대중교통 접근성, 학원수, 유치원, 병의원 선택

안산시 통계청

2-5. 추천지역찾기 안산시 사례 ③

❖ 초등생과 유치원생 두 자녀를 둔 가정이 안산시 상록구 전입시 추천 지역

✓ 결과) 추천지역 리스트에서 선택

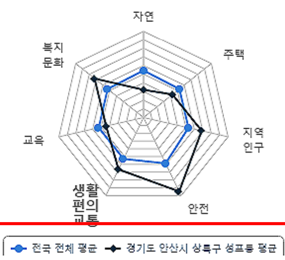


추천지역 리스트

1. 경기두 안산시 상록구 성포동
2. 경기도 안산시 상록구 월피동
3. 경기도 안산시 상록구 본오3동
4. 경기도 안산시 상록구 본오2동
5. 경기도 안산시 상록구 반월동

① 추천지역

지역 종합현황 보기



② 조건별 현황

안산시 통계청

2-6. ‘살고싶은 우리동네’ 활용 사례

❖ 지자체의 추천지역, 주거현황보기, 관심지점조회 활용 예시

추천지역	• 민원콜센터의 주거지 상담 전화 -> 상담원이 주거지분석맵을 활용하여 답변.안내
주거현황보기	• 부서의 업무기획시 객관적 자료로 활용 -> 공동주택비율, 노후주택비율 등
관심지점조회	• 시민관련 업무추진시 활용 -> 행사추진시 대중교통이 편리한 지역 참고 등

안산시 통계청

순서

1. SGIS 오픈플랫폼 소개
2. 안산시 활용사례 (주거지분석맵)
3. **안산시 활용사례 (커뮤니티맵)**
4. 향후계획

3-1. 통계커뮤니티맵 개요

❖ 통계커뮤니티맵은 지역사회 구성원이 직접 참여하여 지역사회의 이슈를 찾아내고 해결하는데, 객관적인 지표인 관련 통계를 제공하고, 이를 기반으로 한 커뮤니티 매핑 기능을 제공하는 서비스

지역주민 참여를 통한 지역적 이슈 개선요구 및 활성화

통계정보 활용	커뮤니티맵 개설/운영	참여를 통한 이슈 등록	이슈 개선 및 활성화
• 인구주택총조사 통계 정보 활용 • 사업체조사 통계정보 활용	• 통계정보를 활용하여 교통 사고, 아동안전 등의 목적의 부합한 커뮤니티맵 개설 및 운영	• 지역주민의 커뮤니티 참여 및 이슈 등록 • 이슈에 대한 댓글 및 부적합 이슈에 대한 신고 기능제공	• 지방자치단체에서 지역주민이 제시한 개선 요구를 수렴 하여 향후 의사결정 에 활용함

안산시 16 통계청

3-2. 안산시 커뮤니티맵 운영 계획

❖ '주민과 함께하는 소통지도' 커뮤니티맵 운영

추진방향

- 안산시민이 바라는 현장의 소리를 효율적으로 소통할 수 있도록 통계커뮤니티맵을 활용

2015년 추진실적

- 통계청SGIS오픈플랫폼 활용 시범과제 선정 : 2015.4.
- 커뮤니티맵 개발 관련 의견 공유 : 2015.5. ~ 12.
- 안산시-통계청간 시범과제 연계 협업 추진 : 2015.12.

2016년 추진일정

- 희망마을만들기 커뮤니티맵 시범 운영 : 2016.4.
- 아동안전지도 커뮤니맵 개설 운영 : 2016.5.

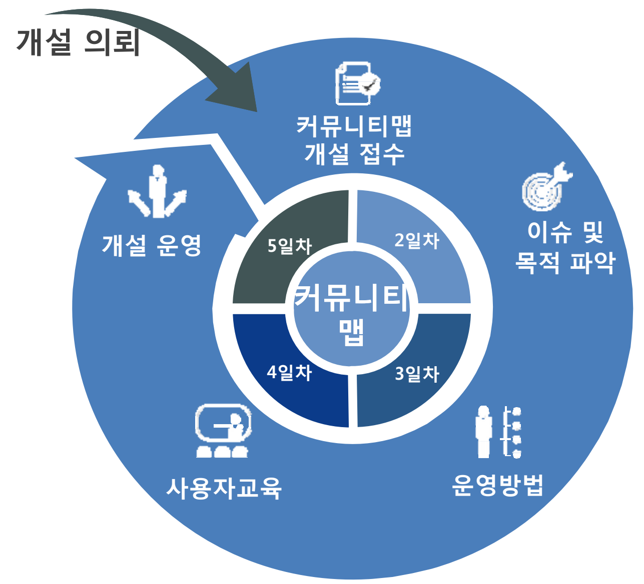
커뮤니티맵 개설 계획

- 시민 소통형 업무 활용을 위한 커뮤니티맵 개설 추진
- 환경, 일자리, 평생학습 등 시민과 밀접한 업무중심으로 부서별 컨설팅 추진

안산시 통계청

3-3. 안산시 커뮤니티맵 운영 절차

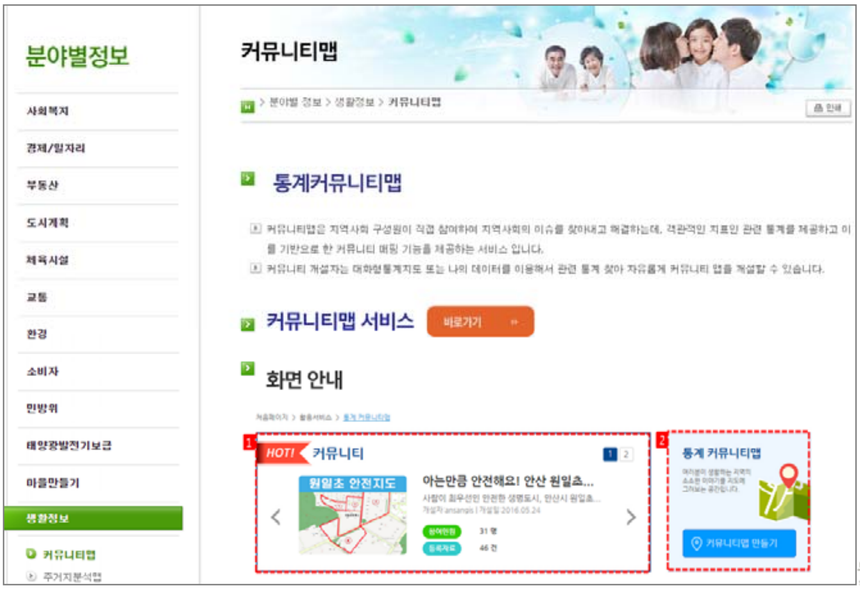
❖ '주민과 함께하는 소통지도' 커뮤니티맵 개설 과정



안산시 통계청

3-4. 통계커뮤니티맵 안산시 운영 사례 ①

❖ 주민들이 안산지역의 특정 주제에 대한 의견 및 정보를 현장에서 수집하고, 이를 온라인 지도(커뮤니티맵)로 만들어 공유
❖ 지역의 이슈를 찾아내고 해결하는데, 객관적인 지표인 관련 통계를 활용

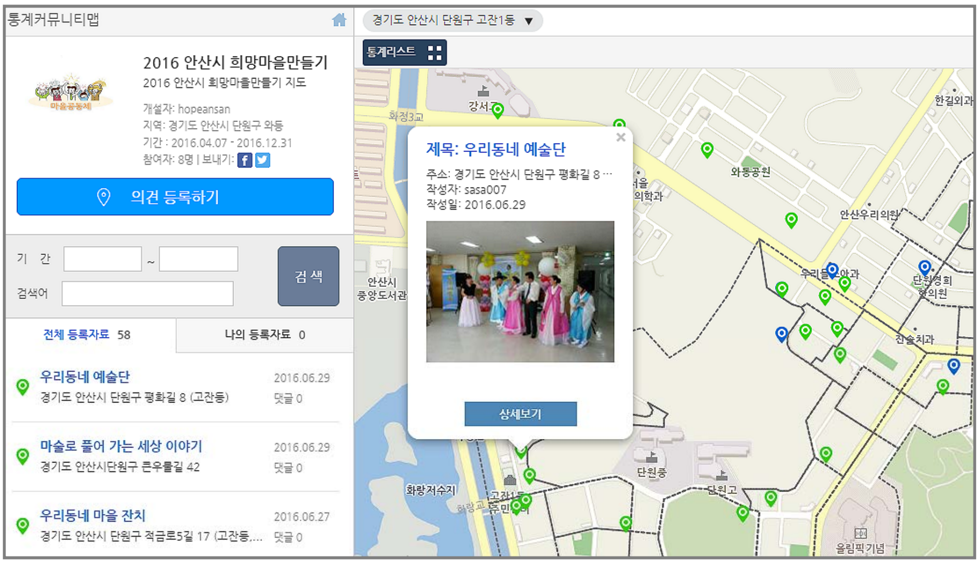


3-4. 통계커뮤니티맵 안산시 운영 사례 ②

❖희망마을만들기,안전지도, 환경커뮤니티맵 등으로 운영

3-5. ‘2016 안산시 희망마을만들기’ 사례 ①

❖ 운영목적 : 세월호 집중 피해지역을 대상으로 희망마을 사업에 대한 추진 사항 및 소식을 공유하고자 운영



안산시 통계청

3-5. ‘2016 안산시 희망마을만들기’ 사례 ②

❖ 활용의미 : 마을 간사들이 마을의 주요 활동을 기록.공유하며, 함께 만들어 가는 공동체 지도 이야기로 활용

안산시 통계청

3-6. ‘아는만큼 안전해요! 안전지도’ 사례 ①

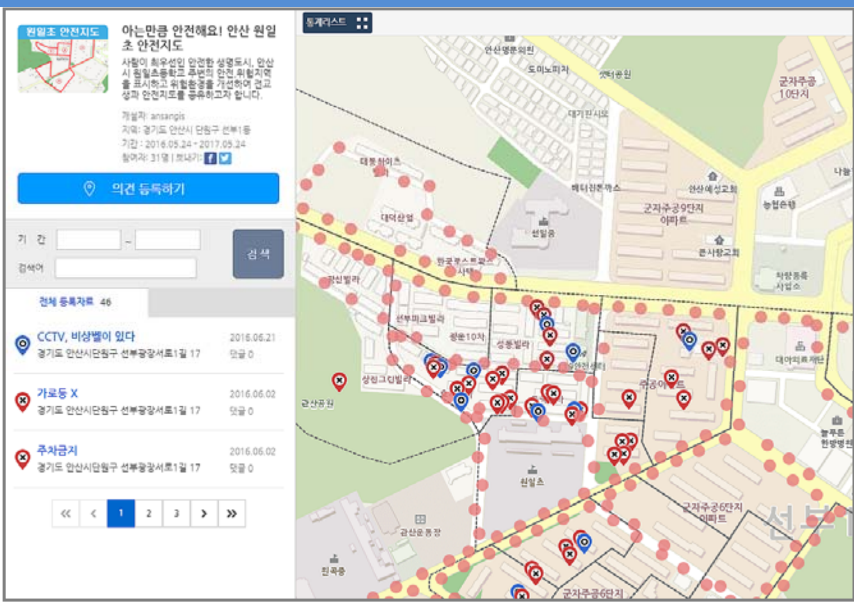
❖ 운영목적 : 초등학교 주변의 범죄피해 예방을 위한 안전지도 커뮤니티맵을 개설하여 이를 공유하고, 안전환경 조성에 활용

추진개요	• 안산시 여성가족과에서 추진중인 아동 안전지도 제작과 관련하여 협업으로 추진
운영 개요	•커뮤니티맵 : ‘아는만큼 안전해요! 안산원일초 안전지도’ •활동대상 : 원일초등학생(아이디 등록부여) •주요내용 : 안전한 곳과 위험한 곳에 대한 조사 활동
운영 결과	•등록자료 : 46건 •주요내용 : 안전지역과 위험지역에 대해 알아보고, 내용과 사진을 등록, 안전지도 제작 배포
향후계획	•커뮤니티맵의 등록자료를 기초로 위험지역의 정보를 관련부서 및 기관에 개선조치토록 협조 요청

안산시 통계청

3-6. ‘아는만큼 안전해요! 안전지도’ 사례 ②

❖ 공간정보 활용 : 안산시공간정보시스템에서 원일초주변의 가로등,보안등에 대한 위치 정보를 추출하여 커뮤니티맵에 표시



안산시 통계청

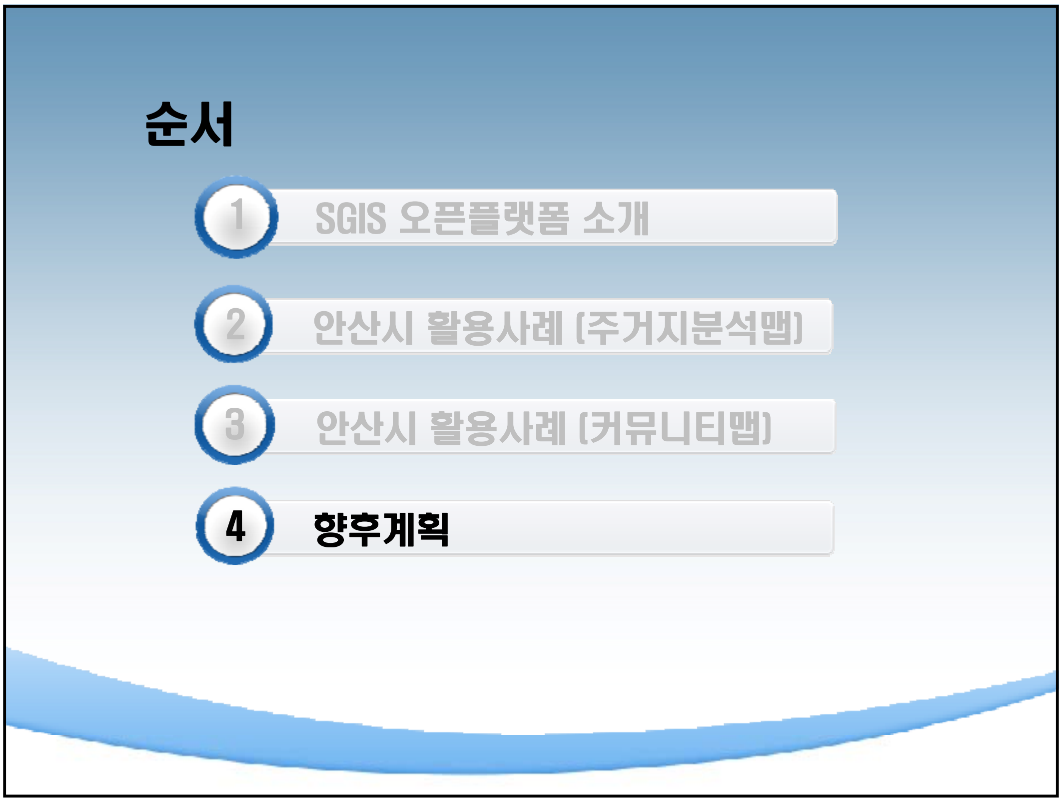
3-6. ‘아는만큼 안전해요! 안전지도’ 사례 ③

❖ 활동내용 : 커뮤니티맵과 같은 안전지도를 인쇄하여 가로등과 보안등의 위치를 확인하고, 안전한 곳과 위험한 곳에 대한 의견 및 사진정보를 등록

3-6. ‘아는만큼 안전해요! 안전지도’ 사례 ④

❖ 의견등록사례 : 위험한 곳에 대한 위치와 의견, 관련 사진 등록


3-6. '아는만큼 안전해요! 안전지도' 사례 ⑤
❖ 커뮤니티맵 결과물에 대한 향후 일정
실행
결과물
확인 점검
조치
커뮤니티맵 완료
결과 분석 및 분류
관련부서 제공
보완 및 개선
안산시
통계청


순서
1 SGIS 오픈플랫폼 소개
2 안산시 활용사례 (주거지분석맵)
3 안산시 활용사례 (커뮤니티맵)
4 향후계획

4-1. 안산시 커뮤니티맵 추가 운영 계획

❖ 시민의 생활과 밀접한 시정에 대한 정보 및 의견 공유를 위해 활용 확대

커뮤니티맵 개설 계획

일자리맵	•일자리정보에 대한 정보와 의견 공유
평생학습맵	•안산시 평생학습센터에 대한 정보와 의견 공유
구정활동맵	•구청에서 추진 중인 사업에 대한 정보제공 및 의견 수렴
환경맵	•안산시와 안산환경재단간 환경.생태정보 공유

안산시 통계청

4-2. 커뮤니티맵 개선 사항 요청

❖ 통계청 SGIS 오픈플랫폼 기능 개선사항 요청

개선 요청

접근성

• 통계커뮤니티맵 인터넷 접속 용이성 강화

최신성

• 맵지도의 최신성 강화

통계활용

• 조건에 맞는 인구통계 제공

기타

• 등록자료 보기 건수 조정
• 등록자료 숨김 기능 제공
• 지역별 검색 제공

안산시 통계청

감사합니다.

2016. 8.

안산시

제6회 국가통계발전포럼

통계청

〈세션 4〉 국가통계의 개방과 활용

수도권 도시철도 역 승하차 패턴과 역세권 인구/사회적 특성 분석

2016. 08. 31

손홍규
연세대학교 사회환경시스템공학부 교수

연세대학교 측량및지형공간정보연구실

Contents

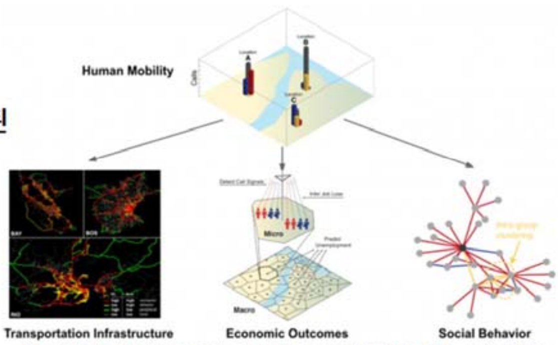
Contents

국가통계발전포럼 Statistics Korea　　3　　GRSLAB YONSEI UNIVERSITY

서론

연구배경

- **사람들의 이동에 대한 이해**
  - 도시 내/사이에서 사람들의 이동은 노선설계, 도시계획, LBS, 마케팅 등 다양한 분야에서 중요한 정보(Zhao *et al.*, 2008)
  - 전통적으로 인구사회학에 의해 이에 대한 분석이 이루어졌지만 이는 거주인구를 기반으로 하여 하루 동안의 인구이동을 반영하지 못하는 제약이 따름(Chen *et al.*, 2009; Goodchild *et al.*, 1993; Janelle and Goodchild, 1983)
- **빅데이터의 활용**
  - 전 세계적으로 AFC(Automated Fare Collection)시스템이 널리 활용되면서 AFC 시스템에서 추출된 데이터가 유용하게 사용되고 있음(Bagchi and White, 2005; Blythe, 2004; Kim *et al.*, 2013)
  - 대중교통 통행 자료는 대중교통 이용객의 흐름을 파악할 수 있는 기초자료로서 여러 연구에 활용됨



Source: Putting big data in its place : understanding cities and human mobility with new data sources (Toole, 2016)

국가통계발전포럼 Statistics Korea　　4　　GRSLAB YONSEI UNIVERSITY

서론

우리나라 수도권(서울 대도시권)

- 국토의 약 12%를 차지하며, 인구의 49.4%가 밀집되어 있음 (MOI, 2015)
- 통근수단으로 약 76.3%가 대중교통 이용 (Seoul survey, 2012)
  - ▶ 수도권에서 **대중교통수단**을 이용하는 것은 매우 **보편적**
- 대중교통 이용자 중 약 64%가 지하철을 이용
  - ▶ 수도권 내 대중교통 통행데이터는 사람들의 이동패턴과 주변 환경에 대해 충분한 정보를 줄 수 있으며 보편성을 지님

연구목표

- **수도권 도시철도 역의 군집 분석**
  - ▶ 수도권의 도시철도 역 승하차 패턴을 이용한 도시철도 역의 군집화
  - ▶ 보편적인 **이동패턴** 파악
- **분류된 각 군집의 특성 분석**
  - ▶ 분류된 각 역의 토지피복 및 인구사회학적 특성 분석
  - ▶ 이동패턴과 **주변환경과의 관계** 파악

국가통계발전포럼 Statistics Korea　5　GRSLAB YONSEI UNIVERSITY

Contents

국가통계발전포럼 Statistics Korea　6　GRSLAB YONSEI UNIVERSITY

연구대상지역

- 수도권
  - 서울, 인천, 경기도를 포함
  - 서울의 인구는 감소하고 수도권의 인구는 증가하는 교외화 현상 발생
  - 2013년을 기준으로 수도권에는 1~9호선, 분당선, 신분당선, 중앙선, 공항철도 1호선, 인천1호선 등 총 25개의 철도노선이 운행 중
    - ▶ 수도권 내 도시철도 역/역세권을 대상으로 함

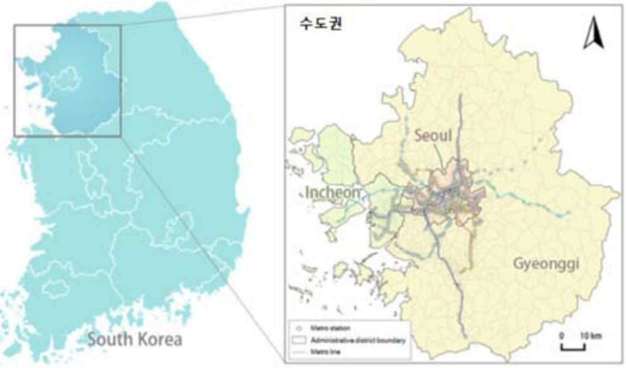
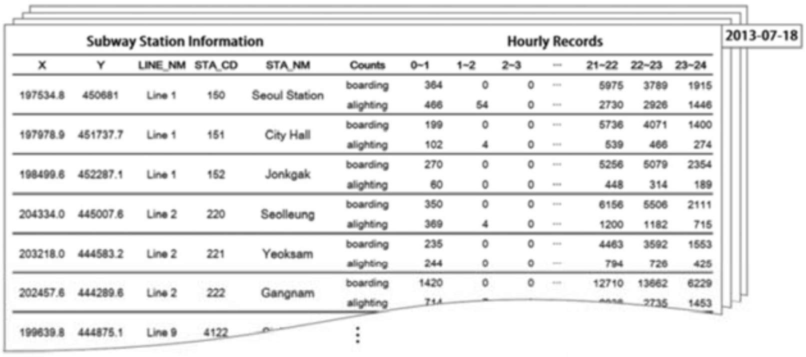


Fig. 1. Study Area and Location of Subway Stations in SCA

분석자료

- 대중교통자료
  - 도시철도 역별/시간대별 승하차 인원 (2013.07.18-2013.07.24)
    - ▶ 서울시 교통정책과 제공



2013-07-18

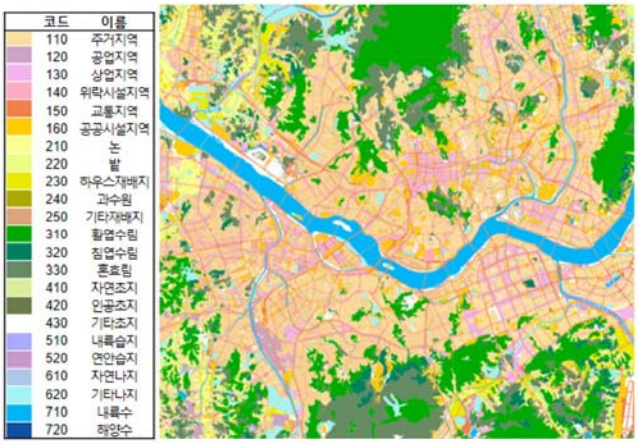
Subway Station Information					Hourly Records							
X	Y	LINE_NM	STA_CD	STA_NM	Counts	0~1	1~2	2~3	···	21~22	22~23	23~24
197534.8	450681	Line 1	150	Seoul Station	boarding	364	0	0	···	5975	3789	1915
					alighting	466	54	0	···	2730	2926	1446
197978.9	451737.7	Line 1	151	City Hall	boarding	199	0	0	···	5736	4071	1400
					alighting	102	4	0	···	539	466	274
198499.6	452287.1	Line 1	152	Jonkgak	boarding	270	0	0	···	5256	5079	2354
					alighting	60	0	0	···	448	314	189
204334.0	445007.6	Line 2	220	Seolleung	boarding	350	0	0	···	6156	5506	2111
					alighting	369	4	0	···	1200	1182	715
203218.0	444583.2	Line 2	221	Yeoksam	boarding	235	0	0	···	4463	3592	1553
					alighting	244	0	0	···	794	726	425
202457.6	444289.6	Line 2	222	Gangnam	boarding	1420	0	0	···	12710	13662	6229
					alighting	714	[illegible]	[illegible]	[illegible]	[illegible]	2735	1453
199639.8	444875.1	Line 9	4122	[illegible]								

⋮

Fig. 2. Configuration of Subway Transit Data Used in the Study

분석자료

- 대중교통자료
- 토지피복도
  - 중분류 토지피복지도 (2012)
    - 환경부 제공 (환경공간정보서비스, https://egis.me.go.kr)



분석자료

- 대중교통자료
- 토지피복도
- 인구통계자료
  - 인구(2010)/가구(2010)/주택(2010)/사업체(2012)
  - 집계구 경계(2012)
    - 통계청 제공 (통계지리정보서비스, http://sgis.kostat.go.kr)

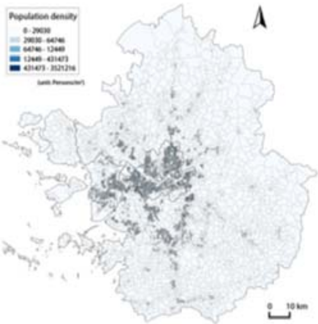


Fig. 3. Population Density of SCA at Census Block Level

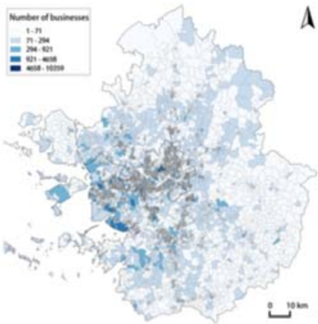
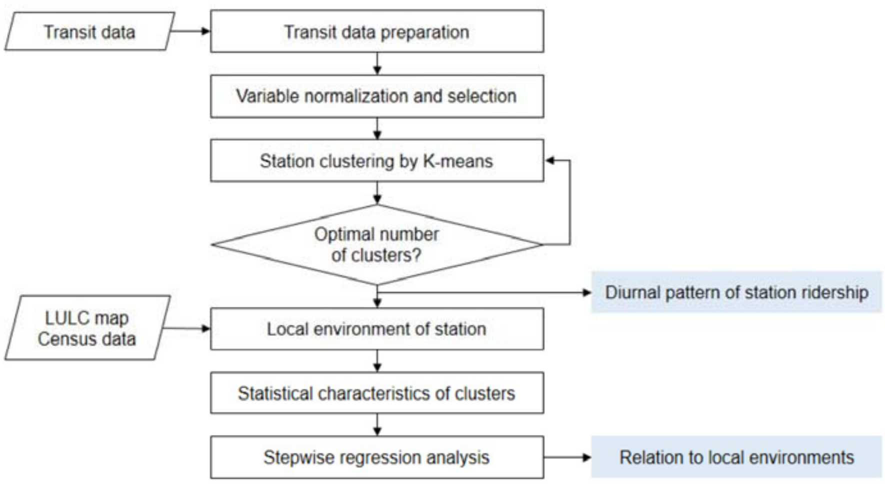


Fig. 4. Number of Businesses of SCA at Census Block Level

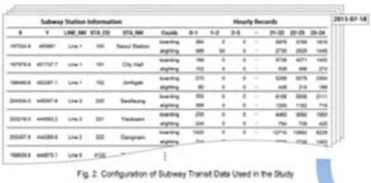
Contents

연구 흐름도



승하차 자료 준비

- **데이터 준비**
  - 환승 가능한 역을 통합하여 하나의 역으로 간주
  - 수도권 내 479개의 도시철도 역이 분석대상



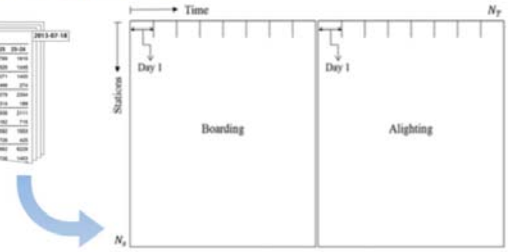
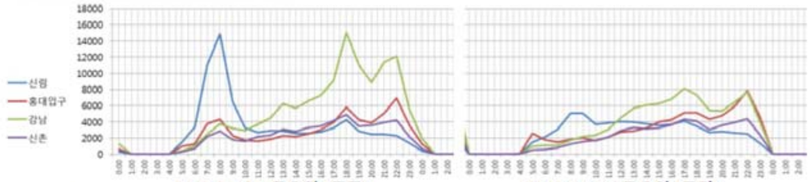


Fig. 6. 2-D matrix form of the Modified Transit Data

데이터 예시: 시간별 승차인원



2013-07-18 목요일 (주중) 2013-07-20 토요일 (주말)

도시철도 역 군집화

- **정규화 및 변수 선택**
  - 각 역의 하루 총 이용객 대비 시간 별 승하차 인원을 사용하여 정규화
  - 의미 있는 값을 선택하여 군집화에 이용하기 위하여 주성분분석 수행
    - ▶ 8시, 9시, 19시(주중) 13, 14시(토요일)의 계수가 높게 나타남

- **K-means 알고리즘을 이용한 역 군집화**
  - K-means 알고리즘은 대표적인 분할 군집화 방법으로 n개의 관측치를 k개의 군집으로 나누기 위하여 각각의 관측치가 각 군집의 평균값 중에서 가장 가까운 값의 군집에 속하게 하는 분류 방법(McQueen, 1967)
    - ▶ 각 역별 주중평균 출/퇴근시간의 승차/하차 인원, 토요일 오후 하차인원을 입력 값으로 사용

- **적정 군집 수 결정**
  - 유효성 측도(Validation index)를 산출하여 적정 군집 수 결정
    - ▶ Xie and Beni index, Dunn index, Average Silhouette 값 비교

주변환경과의 관계 분석

- **역세권 정의**
  - 일반적으로 역의 지배력이 미치는 '지리적 거리'를 의미하며, 철도법과 국토이용 관련 법의 상세계획지침에서 역으로부터 반경 500m 이내로 적용 (김준, 2012)
    - 각 역을 중심으로 **500m 반경 이내** 특성 분석

- **역세권 특성 분석**
  - 기술 통계량 산출
    - 토지피복
    - 인구통계자료
  - 단계적 회귀분석
    - 종속변수: 각 클러스터 해당 여부(y)
    - 독립변수: 인구통계, 토지피복 등의 항목(X)
    - 독립변수를 하나씩 선택하면서 매 단계별로 미리 들어간 변수가 중요한지를 판단하여 최종적으로 설명력이 있는 변수들을 선택하는 방법
    - 총 182개 변수 중, **어떤 변수가 가장 군집의 특성을 잘 설명할 수 있는지**를 선택하기 위하여 활용

$$\underset{479\times 1}{y} = \overset{479\times 182}{X}\beta + \varepsilon,\ \varepsilon \sim N(0, \sigma^2)$$

국가통계발전포럼 Statistics Korea 15 GRSLAB YONSEI UNIVERSITY

Contents

국가통계발전포럼 Statistics Korea 16 GRSLAB YONSEI UNIVERSITY

군집화 결과

▪ 군집별 대표 역

군집	주요 역	역 개수
Cluster 1	시청, 광화문, 서초, 양재, 충무로, 강남구청, 삼성, 공덕, 가산디지털단지 등	49
Cluster 2	서현, 영등포, 김포공항, 한티, 정발산, 안암, 오목교, 강남, 건대입구, 신촌, 이대, 홍대입구, 고속터미널, 이태원, 혜화 등	37
Cluster 3	광흥창, 가락시장, 도곡, 뚝섬, 수서, 대방, 구로, 신설동, 영등포구청, 보라매, 숙대입구, 동막, 신연수 등	81
Cluster 4	오리, 정자, 마장, 송정, 마포구청, 수원, 의왕, 의정부, 대치, 천호, 서울대입구, 강변, 신도림, 노원, 공릉, 이수, 주안, 부평 등	154
Cluster 5	가양, 녹천, 작전, 부개, 행신, 대화, 과천, 구리, 금곡, 역곡, 개봉, 단대오거리, 상록수, 미아, 까치산, 신림, 봉천, 양천구청 등	158

국가통계발전포럼
Statistics Korea
17
GRSLAB

역 군집의 공간적 분포

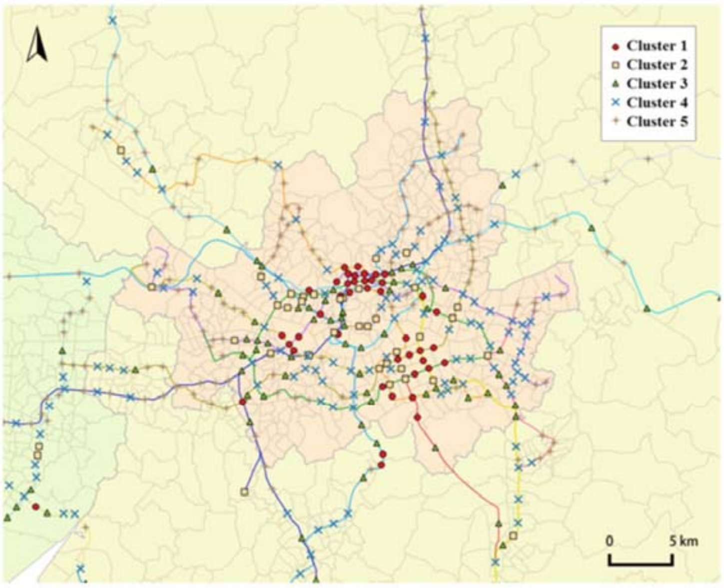
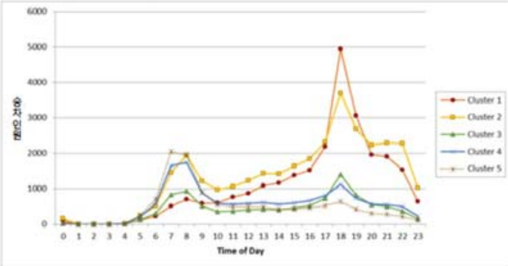


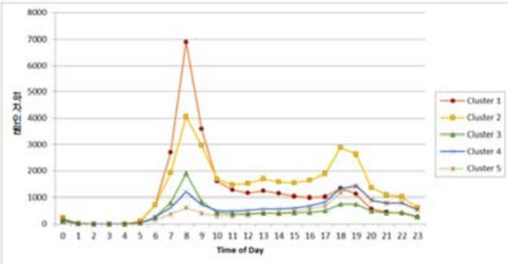
Fig. 9. Spatial Distribution of Clusters

국가통계발전포럼
Statistics Korea
18
GRSLAB

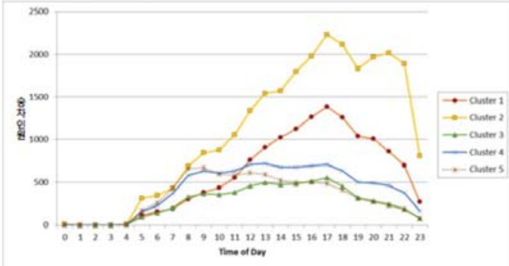
역 군집의 시간별 승하차패턴

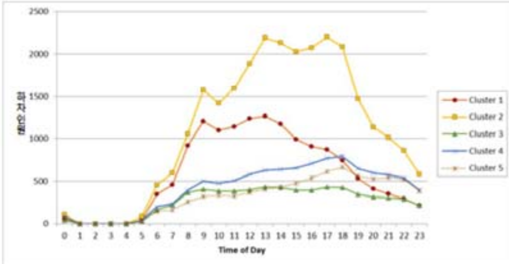
주중(월-금) 패턴





주말(토,일) 패턴





국가통계발전포럼 Statistics Korea

GRSLAB YONSEI UNIVERSITY

역 군집과 주변환경과의 관계

- 주요 토지피복 비율

국가통계발전포럼 Statistics Korea

GRSLAB YONSEI UNIVERSITY

역 군집과 주변환경과의 관계

▪ 기술통계량 분석

Category			Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		ANOVA	
			Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	F	p
인구	인구총괄	노령화지수	286.04	376.2	154.22	138.29	123.76	78.85	95.06	49.86	83.59	32.91	22.35	8.00E-17
	인구총괄	유년부양비	12.06	5.64	13.46	6.33	17.50	5.39	17.94	4.99	20.12	4.63	28.61	4.00E-21
	교육정도별	대학원(석사과정) 계	43.08	23.43	41.55	19.34	31.08	23.13	27.98	18.36	17.72	10.57	27.13	1.00E-36
	교육정도별	대학교(4년제이상) 계	220.06	94.94	227.65	74.95	162.78	68.68	170.57	67.45	130.89	44.24	27.82	4.00E-20
	성/혼인상태별	미혼_여자	121.17	52.49	132.6	48.22	83.47	28.37	93.05	28.54	76.3	18.49	36.07	1.00E-20
가구	방,거실,식당수별	방1개	104.25	81.06	113.08	77.22	53.38	69.22	58.72	58.79	30.41	28.84	26.64	4.00E-26
	세대구성별	1인가구	116.61	81.18	126.15	72.68	66.92	55.87	73.17	50.57	50.31	27.64	28.93	8.00E-20
	세대구성별	비혈연가구	5.71	5.40	6.62	4.63	2.11	4.77	2.28	2.94	0.83	1.96	33.06	2.00E-21
주택	연건평별	40m²~60m² 이하 (호)	16.50	12.14	28.04	37.41	31.95	24.01	38.19	27.47	48.22	32.26	13.11	4.00E-24
	주택유형별	영업용 건물 내 주택	7.55	11.99	3.09	2.95	1.69	2.45	1.72	2.62	0.84	1.18	25.19	4.00E-10
사업체	사업체총괄	총 사업체수	962.05	1150.08	260.32	322.88	129.31	131.48	108.47	108.65	54.77	37.6	53.25	8.00E-19
	사업체수	부동산업 및 임대업	29.28	34.05	8.95	10.71	3.69	2.86	3.83	3.53	1.52	1.52	56.58	5.00E-09
	사업체수	전문,과학 및 기술서비스업	59.97	74.35	17.51	37.89	4.14	4.5	2.94	5.32	0.63	1.86	52.10	1.00E-38
	종사자수	도매 및 소매업	1391.17	1838.59	417.19	608.3	183.03	242.25	117.66	131.59	53.73	71.67	54.97	5.00E-36
토지피복	토지피복비율	주거지역	15.02	14.09	27.63	15.35	25.94	15.04	33.95	14.92	40.47	17.09	28.20	1.00E-37
	토지피복비율	상업지역	35.76	21.91	29.21	16.26	18.86	14.71	17.66	10.75	12.86	8.01	34.21	7.00E-21

국가통계발전포럼 Statistics Korea

21

GRSLAB YONSEI UNIVERSITY

역 군집과 주변환경과의 관계

▪ 회귀분석 결과

Main Variables by Stepwise Regression

클러스터	대분류	소분류	선택된 변수	B	SE	Beta	p
Cluster 1	총괄	인구총괄	노령화지수	-0.0006	0.0170	-0.0938	0.0000
	인구	성/연령별 인구	10세이상~14세이하	-0.0111	0.0289	-0.1346	0.0000
	인구	교육정도별 인구	대학원(박사과정) 계	0.0054	0.0106	0.0486	0.0000
	사업체	종사자수/ 9차 산업분류	도매 및 소매업	0.0001	0.0194	0.0878	0.0000
	사업체	종사자수/ 9차 산업분류	교육 서비스업	0.0008	0.0116	0.0910	0.0000
	사업체	종사자수/ 9차 산업분류	예술, 스포츠 및 여가관련 서비스업	0.0007	0.0110	0.0477	0.0000
Cluster 2	인구	성/연령별 인구	20세이상~24세이하_여자	0.0035	0.0137	0.0506	0.0002
	인구	교육정도별 인구	대학원(박사과정) 계	0.0037	0.0116	0.0334	0.0043
	가구	방,거실,식당수별 가구	식당2개이상	1.0937	0.0118	0.0512	0.0000
	가구	세대구성별 가구	비혈연가구	0.0110	0.0140	0.0425	0.0026
	토지피복	토지피복비율	상업지역	0.0023	0.0122	0.0334	0.0062
Cluster 3	총괄	인구총괄	노령화지수	0.0008	0.0344	0.1114	0.0013
	주택	연건평별 주택	20㎡이하(호)	0.0255	0.0161	0.0411	0.0113
	사업체	사업체수/ 9차 산업분류	숙박 및 음식점업	-0.0029	0.0337	-0.1738	0.0000
	토지피복	토지피복비율	주거지역	-0.0027	0.0177	-0.0482	0.0068
	토지피복	토지피복비율	공업지역	0.0165	0.0168	0.0908	0.0000
Cluster 4	총괄	인구총괄	총인구	0.0005	0.0215	0.0503	0.0199
	인구	성/연령별 인구	65세이상~69세이하	-0.0153	0.0266	-0.1050	0.0001
	사업체	사업체수/ 9차 산업분류	출판, 영상, 방송통신 및 정보서비스업	-0.0030	0.0215	-0.0470	0.0296
Cluster 5	인구	성/연령별 인구	10세이상~14세이하	0.0048	0.0221	0.0587	0.0081
	인구	교육정도별 인구	대학원(박사과정) 계	-0.0168	0.0188	-0.1498	0.0000
	가구	방,거실,식당수별 가구	방1개	-0.0018	0.0196	-0.1115	0.0000
	토지피복	토지피복비율	주거지역	0.0091	0.0202	0.1607	0.0000
	토지피복	토지피복비율	산림지역	0.0087	0.0193	0.0870	0.0000

국가통계발전포럼 Statistics Korea

22

GRSLAB YONSEI UNIVERSITY

Contents

국가통계발전포럼 Statistics Korea

GRSLAB YONSEI UNIVERSITY

결론

연구의 의의

- 우리나라 수도권 도시철도 역을 대상으로 승/하차패턴 군집화
- 역 이용패턴(이동패턴)과 주변환경과의 관계를 분석
  - 토지피복, 주거인구 수, 사업체 수 등과 같은 인자들로 도시철도 역의 승/하차 패턴 설명 가능함
- 도시철도 역 군집의 공간적 분포가 수도권 내 교외화 현상을 반영
- 주중, 주말의 승/하차 분포 차이를 명확히 하여 역 군집 기능 구별

시사점

- 공공데이터로만 이루어진 연구
- 본 연구 방법론은 토지개발, 도시계획 등 관련 분야의 정책 결정에 도움을 줄 수 있음

국가통계발전포럼 Statistics Korea

GRSLAB YONSEI UNIVERSITY

제6회 국가통계발전포럼

감사합니다

GRSLAB

통계산업 국내외 현황 및 시사점

2016. 8. 31.

통계청
Statistics Korea

수원대학교 경영학과 교수 김 지 영

차 례

Ⅰ. 통계산업 정의 및 분류(안)

1. 통계의 정의

해외 + 국내

OECD	개체들의 집단에 대한 수량적 데이터 (numerical data relating to aggregate of individuals)
통계법 제3조	통계라함은 통계작성기관이 특정의 집단이나 대상 등에 관하여 직접 또는 다른 기관이나 법인 또는 단체 등에 위임 · 위탁하여 작성하는 수량적 정보를 말한다
제안	**통계란 자연 현상 혹은 경제 및 사회 행위 등에 의해 발생하는 특정의 집단이나 대상들에 관한 수량적 정보 .**

Ⅰ. 통계산업 정의 및 분류(안)

각종 산업진흥법들의 입법 사례를 기준으로 살펴본 통계산업의 정의: 통계산업이란 통계의 생산, 가공, 전달, 축적, 이용 활동에 종사하는 생산단위의 집합과 이들 생산단위들의 활동을 지원하는 산업으로 정의 가능

통계산업

협의 정의

- 통계생산업 : 통계의 생산활동에 종사
- 통계분석업 : 통계가공 및 분석 활동
- 통계정보서비스업 : 통계자료를 수요자에게 제공
- 통계산업지원업 : 생산, 분석, 유통을 직접 혹은 간접적으로 지원

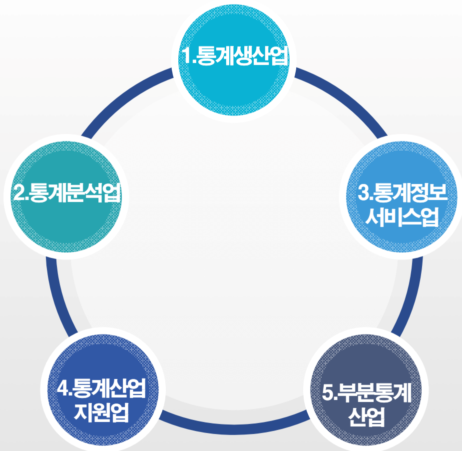
광의 정의

협의의 통계산업 + 부분통계산업

부분 통계산업: 개별산업의 전체 생산활동 중에서 통계서비스(통계의 생산, 분석 및 유통)의 생산 활동이 부분적 또는 보조적으로 수행되는 산업

Ⅰ.통계산업 정의 및 분류(안)

❖ 한국표준산업분류(KSIC)와의 관계



1. 통계생산업 : 통계청, 한국은행, 닐슨컴퍼니코리아, 월드리서치, 무역협회
84(공공행정, 국방 및 사회보장 행정), 71400(시장조사 및 여론조사업), 941(산업 및 전문가 단체)

2. 통계분석업 : 한국경영통계연구소, 한국능률협회컨설팅, 한국리서치, 새롬정보통계
71531(경영컨설팅업), 71400(시장조사 및 여론조사업), 63999(기타 정보 서비스업)

3. 통계정보서비업 : 통계청, 네이버, 국민은행
63991(데이터베이스 및 온라인정보 제공업), 84(공공행정, 국방 및 사회보장 행정), 64(금융업), 65(보험 및 연금업), 66(금융 및 보험관련 서비스업), 63120(포털 및 기타 인터넷 정보 매개 서비스업)

4. 통계산업지원업 : SAS코리아, 통계개발원및교육원
63111(자료처리업), 58222(응용소프트웨어 개발 및 공급업), 70119(기타 자연과학 연구개발업), 70201(경제학 연구개발업), 84(공공행정, 국방 및 사회보장 행정), 85(교육서비스업), 71400(시장조사 및 여론조사업), 7512(인력공급업)

5. 부분통계산업: 기상청, 삼성생명, KCB, 아산병원, 동아제약
84(공공행정, 국방 및 사회보장 행정), 75993(신용조사 및 추심대행업), 65100(보험업), 66120(증권 및 선물 중개업), 86(보건업), 21(의료용 물질 및 의약품 제조업) 등

Ⅱ. 국내 통계산업 현황 및 과제

1. 국내 통계산업 규모 추정

<표> 업종별 통계산업 산출액

단위 : 억원

	2000년	2005년	2008년	2010년
통 계 생 산 업	2,566	5,471	7,536	9,337
통계정보서비스업	101	988	1,881	2,173
통계산업지원업	307	682	769	973
통 계 산 업(협의)	2,974	7,141	10,186	12,483
GDP(명목)	6,032,360	8,652,409	10,264,518	11,732,749
GDP대비 비중(%)	0.05	0.08	0.10	0.11

➡ 통계산업의 산출액 규모를 명목 GDP와 대비하여 보면,
2000년 0.05%에서 2010년 0.11%로 약 2배 정도 상승

Ⅱ. 국내 통계산업 현황 및 과제

1. 국내 통계산업 규모 추정

<표> 업종별 통계산업 종사자

단위 : 명, %

	2000년	2005년	2008년	2010년
통 계 생 산 업	7,351	9,222	10,704	11,152
통계정보서비스업	364	388	465	543
통계산업지원업	303	414	323	373
통 계 산 업(A)	8,018	10,024	11,492	12,068
총취업자(B)	21,156,000	22,856,000	23,577,000	23,829,000
총취업자 대비 비중(A/B)	0.038	0.044	0.049	0.051

총취업자 대비 통계산업 종사자수 비중을 보면
2000년 0.038%에서 2010년 0.051%로 증가

Ⅱ. 국내 통계산업 현황 및 과제

2. 국내 통계산업시장 환경 및 특성

환경

- 국내 DB산업 시장 매출액은 2011년 10.4조원으로 과거 4년간 평균 9.7%의 성장률 기록
- 세계 빅데이터 시장은 2010년 32억 달러로 2015년까지 연 39.4% 성장한 것으로 추정

특성

- IT 산업의 흐름과의 큰 연관성
- 제조업이 뛰어나고 시민들의 학습수준이 높기 때문에 데이터분석 기반 양호
- SW투자 미비로 인해 전반적 통계산업의 중요한 기술 및 솔루션은 외국업체들이 독점

Ⅱ. 국내 통계산업 현황 및 과제

3. 통계산업 활성화 과제

데이터통합 및 공유 ▶ 데이터 품질관리 ▶ 데이터분석 및 보안 기법 연구 ▶ 인식변화 ▶ 개인정보보호

통합 및 공유	- 전체적으로 보면 상당한 규모의 통계가 있지만 각각 분리 관리되고 있기 때문에 통합분석에 장애 - 정부의 각 조직간 정보 단절 상태 완화 필요
품질관리	- 공공정보로 활용되는 DB는 품질을 높일 수 있는 재정적 지원과 인증 도입 필요 - 통계주체가 데이터의 품질을 자율적으로 관리하도록 규제하고 유능한 인력 양성을 추진
분석/보안연구	- 데이터 분석에 대한 연구와 연구에 대한 정부 지원 필요 - 보안기법 강화방안 연구
인식변화	- 데이터에 대한 가치를 인식하고 정당한 대가를 지불하여야 한다는 의식 홍보 필요
개인정보보호	- 통계산업 발전을 위해서는 개인정보보호 환경이 지금보다 개선될 필요

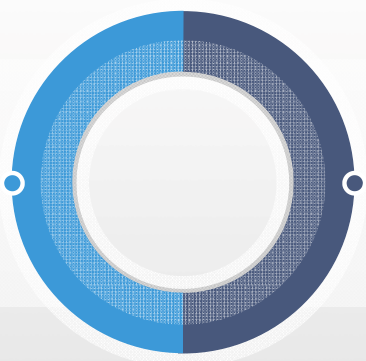
Ⅲ. 해외 통계산업 현황 및 시사점

1. 미국

(1) 공공, 민간분야별 통계

공공분야

1) 연방정부 : 미국 상무성에 속해있는 'Bureau of the Census'가 통계생산기구로서 독보적
2) 지방정부 : 주정부 단위에서는 통계인을 고용하여 조사통계를 생산하며, 시정부도 표본조사 실시



민간분야

표본조사회사와 제약회사, 시장조사회사, 여론조사회사, 통계소프트웨어 패키지 회사, 각종 산업협회 등

Ⅲ. 해외 통계산업 현황 및 시사점

1. 미국

(2) 통계산업의 경제적 비중

- Census Bureau의 2010년 census 예산은 130억불, 외부기관 등의 표본조사를 대행하기 위한 용역수입이 2억 5천만 달러를 상회
- 2010년 SAS 매출액은 24억 3천만 달러, 시장조사회사인 Nielson은 50억 달러, Kantar그룹 20억 달러, IMS Health 20억 달러 등이었으며, WESTAT은 4.3억 달러(Wikipedia)

(3) 통계산업 발전과정과 추세

- 최초의 공식통계(전국단위): 1790 인구센서스, 센서스를 위하여 Hollerith(punch card) 개발, UNIVAC 개발 유도 및 구매
- 정부(NIH) 지원으로 첫 통계 software BMDP 개발
- 정부부처는 정책 수립 및 점검을 위해 통계 필요, 각종 통계국 등을 설립하고 예산을 확보하여 민간업체를 선정함으로써 통계산업을 활성화

(4) 통계를 이용한 사업 모델

- 미국에서는 많은 회사들이 통계자료를 이용하여 다른 통계를 생산
  대표적인 예 : WESTAT

Ⅲ. 해외 통계산업 현황 및 시사점

2. 영국

(1) 영국의 통계제도와 특징

- 국가통계위원회, 통계청, 정부통계기구(GSS) 등으로 구성
- 분산화된 특징을 띠고 있으며, 정부통계기구를 통해 유기적 관계 유지

(2) 영국의 통계산업의 규모

- 영국 통계산업은 GDP의 약 0.17%를 차지하고, 민간부분이 공공부문에 비하여 큰 규모를 차지

< 영국 공공 및 민간통계산업 규모 추정 >

단위 : 백만파운드, %

	2006	2007	2008	2009	2010	2011
공공통계	192.8	172.1	208.3	207.3	320.2	345.6
민간통계	1,992.0	2,038.0	2,164.0	2,076.0	2,060.0	2,171.2
합계	2,184.8	2,210.1	2,372.3	2,283.3	2,380.2	2,516.8
공공/민간비율(%)	9.68	8.44	9.63	10.00	15.55	15.92
GDP 비율	0.164	0.157	0.165	0.164	0.163	0.167

Ⅲ. 해외 통계산업 현황 및 시사점

2. 영국

(3) 영국 통계산업의 특징

- 영국의 리서치업체는 민간 통계산업에서 큰 비중을 차지
- TNS Research Int'l 등 세계적인 리서치업계 다수 보유
  * 영국 업체들은 일찍이 세계 각국에 지사를 설치하고, 이들 국가들에 대한 종합적인 네트워크를 구축하고 다양한 조사를 실시
  * 예를 들어, TNS Research Int'l는 전세계 90개국에 자회사 혹은 지사를 보유
  * 2010년의 조사 결과, 영국 리서치 업체의 75% 이상이 국제적인 조사업무를 수행
  * 영국내에 이미 세계적인 조사협회를 보유하고 운영
- 조사(survey) 방법이 온라인 중심으로 재편되고 있고 대기업 중심으로 재편
- 조사의 기술적인 측면보다 정성적인 면을 강조

Ⅲ. 해외 통계산업 현황 및 시사점

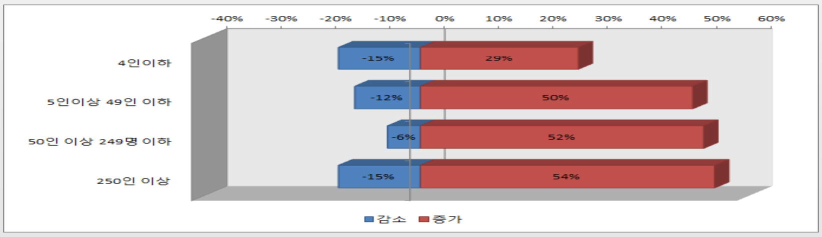
2. 영국

(3) 영국 통계산업의 특징

전세계 TNS Research International의 지사 및 자회사수

지역	중동 및 아프리카	아시아 및 태평량	유럽	남미	북미	전체
지사수	24	16	37	11	2	90

매출이 증가한 업체와 감소한 업체의 비율



Ⅲ. 해외 통계산업 현황 및 시사점

3. 일본

(1) 일본의 통계제도와 특징

- 분산형 통계작성체제를 채택하고 있으며, 각 부처별로 통계조직 운영
- 지방자치단체 통계조직이 거의 모든 국가통계의 조사기능 수행

(2) 일본 통계산업의 규모

- 일본통계산업은 GDP의 약 0.04%를 차지하고, 민간부분이 공공부문에 비하여 5배 큰 규모를 차지

< 일본 정부의 통계예산 >

단위 : 10억엔

	2006	2007	2008	2009	2010	2011
공공통계	32.893	30.056	34.945	38.035	88.844	31.08
민간통계	168.80	175.00	176.60	167.20	169.90	173.10
합계	201.69	205.06	211.55	205.24	258.74	204.18
공공/민간비율(%)	19.5	17.2	19.8	22.7	52.3	18.0
GDP 비율	0.040	0.040	0.042	0.044	0.054	0.044

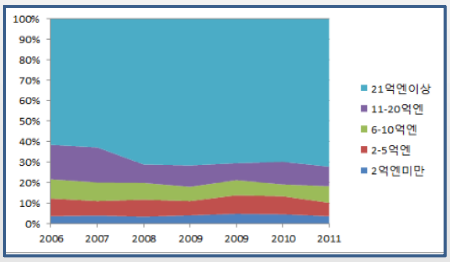
Ⅲ. 해외 통계산업 현황 및 시사점

3. 일본

(3) 일본의 민간 통계산업

- 일본의 리서치업계는 아시아에서 가장 규모가 크고 발전된 모습
- 매출규모가 큰 기업이 전체매출의 대부분을 차지하는 독과점 형태가 뚜렷하게 관찰
- 해외 수주의 규모가 매우 작음
- 전자적 방식을 통한 조사방법의 비중이 크게 상승

< 매출 규모별 리서치업 현황 >



< 조사업체의 조사방법 추이 >

	2004	2006	2008	2010	2011
전체	100.0	100.0	100.0	100.0	100.0
패널조사	33.5	29.0	35.6	34.6	34.2
임시조사	58.5	65.8	56.8	59.8	58.8
인터넷조사	12.0	19.1	19.9	23.9	23.7
기존방법조사	46.5	46.7	36.9	35.9	35.1
기타	8.0	5.2	7.6	5.6	7.0

Ⅲ. 해외 통계산업 현황 및 시사점

4. 독일

(1) 독일 통계제도와 특징

- 모든 통계활동이 연방통계법 및 개별 통계의 근거 법령에 따라 실시되며, 통계작성은 연방통계청과 주정부 통계청에서 주로 담당하는 집중형 통계제도
- 중앙통계청은 통계를 기획하고 실제 통계조사는 모두 지방정부가 실시, 집계

(2) 독일의 통계산업의 규모

- 독일의 공공통계산업(중앙정부 및 지방정부)의 예산은 2011년 474.6백만 유로로 GDP의 0.018% 규모
- 한편 독일의 민간통계산업은 2011년 19억 9백만 유로로 명목 GDP 2조 5708억 유로 약 0.074%에 해당
- 결국 민간 및 공공산업의 규모는 GDP의 0.092%로 추정
- 독일 민간통계산업은 경제위기에도 불구하고, 2011년까지 꾸준한 증가세를 유지

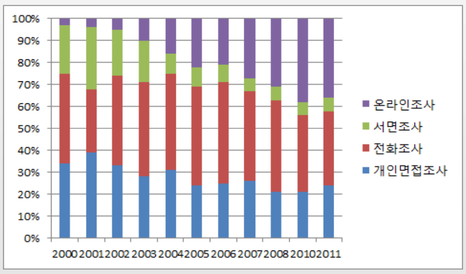
Ⅲ. 해외 통계산업 현황 및 시사점

4. 독일

(3) 독일의 민간통계산업

- 독일의 통계조사 방법에서는 온라인 조사방법이 크게 증가
- 업종별로는 소비재산업이 가장 큰 비중을 차지

< 조사업체의 조사방법 추이 >



<독일 리서치 업계의 업종별 매출>

	조사업종별	2005	2007	2008	2010	2011
1	소비재 산업	57	58	59	55	68
2	자본재 산업	2	2	2	2	1
3	제약 및 의학	12	14	14	14	8
4	정보통신	4	7	7		6
5	운송 물류	1	1	1	1	1

Ⅲ. 해외 통계산업 현황 및 시사점

5. 해외통계산업 구조 및 현황의 시사점

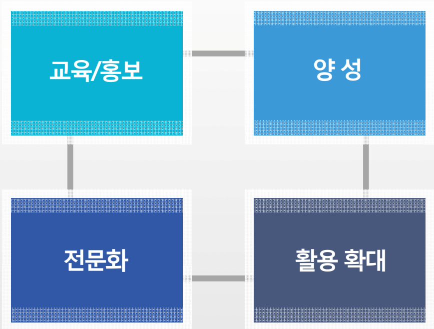
1	국가통계의 제도적인 측면에서 매우 다른 모습을 하고 있어 이를 단순 비교하기 어려운 구조
2	통계산업은 민간과 정부가 공동으로 참여하는 사업이면서도 선진국에서는 공공부문 보다는 민간부문의 비중이 더욱 커지는 산업
3	통계산업은 고부가가치 산업으로 선진국이 매우 발전한 산업
4	통계산업은 자신이 직접 최종소비재를 생산하는 산업이 아니고, 다른 산업을 지원하는 산업으로 해당국가 기업의 발전과 함께 발전
5	통계산업은 서비스산업이기 때문에 다른 국가에 진출하기 쉽지 않음
6	통계산업은 최근 발달된 IT기술을 활용하여 업무 방식을 크게 바꾼 산업
7	통계산업의 관련 기업들은 점차로 대형화되는 특징

Ⅲ. 해외 통계산업 현황 및 시사점

5. 해외통계산업 구조 및 현황의 시사점

국가차원에서 통계의 필요성을 교육, 홍보
미국인들은 대체적으로 표본조사에 잘 응해주며, Census Bureau의 설문조사에 응답할 때에는 국가에 공헌한다는 자부심을 갖고 임함

우수한 통계인 양성
- '공인전문통계사' 제도를 2010년부터 시행
- 미국통계학회는 매년 우수 통계논문 대회를 개최
- 학부생들 대상으로 Undergraduate Statistics Project Competition을 2년마다 개최
- 통계교육을 위해 '통계대사'제도 운영



대학생과 전문인 교육 및 연구
연구계약 혹은 연구그랜트(Research Contract or Grant)의 적절한 활용

PUM(Public Use Microdata)를 작성하여 활용 범위 확대
- Data Center에서 자료 사용 기회 확충

Ⅳ. 통계산업 활성화 방안

1. 기본 방향

1. 통계산업진흥법 제정
2. 통계자료 활용 환경 개선
3. 통계교육 및 홍보활동 강화
4. 통계산업시장 환경 개선
5. 통계품질 확보방안 모색
6. 통계산업 영위 중소기업지원
7. 통계산업 인력 육성
8. 통계 R&D 촉진
9. 통계 국제협력 강화

→ 통계산업 활성화의 장단기 방안 수립

Ⅳ. 통계산업 활성화 방안

2. 단기 과제

1

- 통계산업활성화 추진 담당부서의 지정
- '통계산업백서' 의 작성 추진
- 통계산업관련 연구사업을 추진하여 향후 연구결과를 연구논문집에 기고

2

- 통계교육원은 '통계대사'를 지명하여 전국적으로 통계교육을 실시하는 계획을 수립, 추진
- 마이크로데이터 활용에 대한 정기적인 컨퍼런스를 실시하여 통계활용에 대한 기업 및 일반에 대한 인식을 제고
- 통계산업 전 분야별 통계전문가를 양성하기 위한 온라인 통계전문자격증 연수제도를 실시

Ⅳ. 통계산업 활성화 방안

2. 단기 과제

3

- 마이크로데이터의 이용자 편의를 위한 자료 제공 방안 추진
- 신청방법, 자료제공규정, 이용절차 등을 쉽고 자세하게 설명
- 공공통계 활용도 제고를 위한 '주문형 통계서비스제도'를 추진
- 통계산업 재하청업체의 수익성 보장을 위한 규정 마련

4

- 국가통계위원회는 시도 및 시군구 단위별로 통계전문인력의 정원 확보
- 통계청은 중앙행정부처와 지자체가 민간부문에 용역 발주 가능한 대상통계 또는 기준을 정하고 용역사업시행기준 등을 정하여 민간부문의 참여 유도
- 통계산업 활성화를 위한 기술지원사업 추진을 위해 산,학,연을 연계한 통계산업 기술지원단을 구성하여 유기적, 체계적 사업 추진 방안 강구

Ⅳ. 통계산업 활성화 방안

3. 중장기 과제

1

- 통계산업 영역을 설정하고 통계산업진흥법(가칭) 제정을 추진
- 마이크로데이터 집중대상기관을 확대, 연계
- 기존 자료의 비밀보호장치 등을 통한 활용
- 통계산업 선도를 위한 '공인통계사' 출범 방안 추진

2

- 품질관리과를 통계품질관리원으로 격상하여 통계품질인증제도를 운영토록함
- 통계산업 영위기업을 위한 '통계작성원칙' ,'통계품질인증기준', '통계감리준칙'을 제정하여 법제화하는 방안 추진
- 금융지원을 위하여 '통계산업진흥기금'의 정부출연 추진과 운용 방안 수립

Ⅳ. 통계산업 활성화 방안

3. 중장기 과제

3

- 지식서비스산업에 통계산업 세부업종이 모두 포함될 수 있도록 요청
- 산업은행, 기업은행의 지원대상에 통계산업을 포함시킬 수 있도록 협조 요청
- 통계교육원의 서울분원 설치 추진
- 아시아통계센터(ASC)의 국내 설치 추진

Q & A

통계산업 분류

통계개발원

이정희

목 차

Ⅰ 통계산업이란 ?

Ⅱ 통계산업분류 및 특성

Ⅲ 통계산업분류 작성 의의

통계란?

▣ '특정 모집단의 집단적 현상을 묘사하는 수량적이고 질적인 정보로서 종합적이고 대표적인 정보' (UN)

▣ '개체들의 집단에 관한 수량적 자료' (OECD)

➢ 사회집단 또는 자연집단의 현상을 알아보기 쉽게 일정한 체계에 따라 숫자로 나타낸 것을 의미하며
개별 개체들 각각의 수량적 정보는 통계라고 볼 수 없음

1. 통계산업이란?

통계산업 정의

(일반적 정의) 통계자료를 생산하고, 생산된 통계자료를 의미있는 정보로 만들며, 다른 자료들과 연계 · 가공 · 분석을 한 후 통계 서비스를 제공하는 산업과 이러한 활동을 지원하는 산업

(제안) 정부나 기업 등에서 생산하는 통계 중 상품으로서 거래되는 통계 산출물이 아닌 자체 활용 목적의 통계 산출물을 생산하거나 기타 통계산출물을 활용하는 활동

통계산업 특징

통계산업의 산출물 및 관련 서비스는 공공 및 민간 부문에서 최종재로서 소비되는 것이 아니고 정책형성, 기업경영 등의 의사결정과정에서 중간재 형태로 소비되어 생산성 파악이 곤란하며 산출물의 가치 측정이 어려운 특성이 있음.

2. 통계산업분류 및 특성

통계산업 분류 개요

(목적) 기존 개념의 통계산업 및 새롭게 대두되고 있는 통계산업의 범위를 명확히 정의하여 통계산업을 효율적으로 육성지원하기 위한 기초 자료

(구성) 통계산업분류는 4개의 대분류
20개의 중분류, 27개의 소분류로 구성

(체계) 대분류(1자리 숫자), 중분류(2자리 숫자)
소분류(4자리 숫자)

통계산업 분류기준

(분류기준) 통계산업은 다양한 서비스활동으로 구성되었으며 제공하는 서비스 종류에 따라 분류

- ❖ 통계를 생산하는 활동
- ❖ 통계와 관련된 정보서비스 활동
- ❖ 통계생산을 지원하는 활동
- ❖ 통계를 활용하는 활동

통계산업분류

통계 산업 정의: 통계를 생산하고, 통계자료를 의미있는 정보로 만들며 다른 자료들과 연계 · 가공 · 분석 등을 통해 관련 서비스를 제공하는 사업과 이러한 활동을 지원하는 사업 및 이를 활용하는 산업

대분류	통계 생산업	통계정보 서비스업	통계생산 지원업	통계활용산업
대분류 설 명	• 조사자료, 신고, 등록, 행정자료 등으로 통계 작성 • 통계적 분석기법을 적용하여 모델링화 하거나 통계적 분석처리	• 통계자료 수집, 처리 자료제공 등 통계프로그램을 개발 • 개인정보 보호관련 소프트웨어 개발 • 통계정보를 온라인 등 전자매체로 제공	• 통계관련 연구 및 교육기관 • 통계서적류 출판, 인쇄, 판매업 • 통계인력 공급등 통계지원 활동	• 정부나 기업 등에서 상품으로서 거래되는 통계산출물이 아닌 자체 활용 목적의 통계산출물을 생산하여 활용하는 활동

대분류1 통계생산업

조사자료, 행정자료, 수집자료 등을 가지고 통계적 기법을 활용하여 각종 통계를 생산하는 활동 및 통계적 분석기법을 적용하여 통계 모델링화하거나 분석 등 통계처리하는 산업 활동

11.통계 작성업

12. 통계모델링업

13. 기타 통계 분석 및 가공업

대분류1 통계생산업

11 통계작성업

1101 조사통계 작성업

1102 조사통계 이외의 통계 작성업

1200 통계모델링업

1300 기타 통계 분석 및 가공업

대분류1 통계생산업 특징

❖ 면접조사나 전화조사, 인터넷조사 등 각종 조사방법으로 통계를 생산하는 활동, 행정자료를 가공하여 통계를 생산하는 활동뿐만 아니라 최근 이슈가 되고 있는 빅데이터, 데이터 마이닝 등을 활용하여 통계를 생산하는 활동도 본 분류에 포함

대분류2 통계정보서비스업

통계자료를 수집, 처리, 자료제공 하기 위한 통계관련 프로그램 및 통계생산관리 과정에 수반되는 개인정보 보호관련 소프트웨어를 개발하거나 공급하는 활동 및 통계 시스템을 구축하는 활동

21.통계 자료처리업

22.통계 소프트웨어 개발 및 공급업

23.통계 정보시스템 구축서비스업

24.통계 데이터 베이스업

대분류2 통계정보서비스업

2100 통계자료 처리업

22 통계소프트웨어 개발 및 공급업

- 2201 통계패키지나 통계소프트웨어 개발 및 공급업(범용)
- 2202 통계프로그래밍 서비스업(주문형)
- 2203 빅데이터 수집, 처리관련 프로그램 개발 및 공급업
- 2204 개인정보보호관련 소프트웨어 개발 및 공급업

2300 통계정보시스템 및 통계솔루션 구축 서비스업

24 통계데이타 베이스업

- 2401 통계데이타베이스업
- 2402 통계지리정보서비스업

대분류2 통계정보서비스업 특징

통계산업 산출물은 일반적으로 디지털화가 용이한 수량적 정보로서 정보재의 특성을 가짐.

이러한 정보재의 특성은 짧은 시간에 대량 전파가 가능하여 품질수준 검증 및 개인정보 보호를 위한 제도적 장치들이 사전에 마련되어야 함

이에 따라 통계관련 개인정보보호 프로그램 개발 활동 및 최근에 큰 이슈가 되고 있는 빅데이터 관련 프로그램 개발 활동을 별도로 세분하였음

대분류3 통계생산지원업

통계생산 활동을 지원하는 활동

31 통계연구 개발업

32 통계교육서비

33~35 통계서적류 출판,인쇄,판매업

36.통계인력공급

37.통계협회 및 단체

대분류3 통계생산 지원업

3100 통계연구개발업
32 통계교육서비스업
3201 통계 전문대학
3202 통계관련 대학교 및 대학원
3203 기타 통계관련 교육기관

3300~3500 통계서적류 출판, 인쇄, 판매업
3600 통계인력 공급업
3700 통계협회 및 단체

대분류4 통계활용산업

정부나 기업 등에서 생산하는 통계 중 상품으로서 거래되는 통계 산출물이 아닌 자체 활용 목적의 통계 산출물을 생산하여 활용하는 활동

통계산출물 중 기업경영 과정이나 의사 결정 과정에 주로 활용되는각종 통계들은 중간재적 성격을 가지며, 생산성 파악이 어려워 산출물의 가치 측정을 시도하지 않았음

대분류4 통계활용산업

이에 따라 통계활용산업 규모 측정을 위한 대안을 제시한다면 아래 항목 중 어느 한 항목에라도 포함된다면 통계를 활용하고 있다고 가정하고 경상경비 지출액을 통계활용에 대한 규모로 대체함

① 통계부서나 통계담당 직원이 있는 경우
② 통계예산이 일정규모 이상 있는 경우
③ 통계시스템이나 통계프로그램을 활용하는 경우

3. 통계산업분류 작성 의의

통계산업분류 작성 의의

- 통계산출물이 정보재 성격인점을 감안하여 통계 관련 개인정보보호 관련업을 별도로 분류하였고

  최근 이슈가 되고 있는 빅데이터, 데이터 마이닝 등을 활용하여 통계를 생산하는 활동도 별도로 분류하였음

통계산업분류 작성 의의

- 특히 대분류4 통계활용 산업은 산업규모 측정이 어려운 측면이 있으나 매출액 측정이 아닌 통계작성에 따른 경상경비 지출액등의 자료로 대체하는 방안도 제안함
- 매출액 측정이 아닌 경상경비 지출액 규모 측정이라는 면에서 일반적인 경제통계와는 좀 다를 수 있으나 국가통계 차원에서 규모 파악은 꼭 필요한 분야라고 판단되어 과감히 추가하였음

통계산업분류 작성 의의

- 비록 현 상태는 실험적인 부분도 있으나 더 많은 연구를 통해 국가 차원에서 통계산업에 대한 규모 파악은 꼭 필요하다고 생각함

- 마지막으로 본 연구 자료가 『통계산업진흥법』의 법제화 및 통계산업을 육성 지원 할 수 있는 기초자료로서 도움이 될 것으로 기대됨

참조자료

통계산업 활성화 추진방안 연구(2012, 통계진흥원)

-김지영 교수님외

감사합니다