제11회
국가통계발전포럼

The 11th National Statistics Development Forum

데이터 시대와 국가통계의 역할

2021년 8월 27일(금) | 롯데시티호텔 대전

통계청

제11회 국가통계발전포럼
The 11th National Statistics Development Forum

목차 Contents

프로그램 Program

행사명 제11회 국가통계발전포럼
주 제 데이터 시대와 국가통계의 역할
일 시 2021년 8월 27일(금), 10:30-15:30
장 소 롯데시티호텔 대전
주 관 통계청

시간	세부 프로그램		연사	세션 주제
10:00 - 10:30	등 록			
10:30 - 10:40	전체세션	개회사	**류근관** 통계청장	
10:40 - 10:50		축사	**윤후덕** 기획재정위원장	
10:50 - 11:20		기조연설	**김용대** 서울대학교 교수	데이터의 변신과 국가의 역할
11:20 - 12:00		패널토론	**류근관** 통계청장	
			김용대 서울대학교 교수	
			이태수 한국보건사회연구원장	
			고길곤 서울대학교 교수	
12:00 - 13:00	오 찬			
13:00 - 13:30	전문세션	세션 1	**송용수** 서울대학교 교수	동형암호 기법을 활용한 개인정보보호
13:30 - 14:00		세션 2	**최슬기** KDI 교수	저출산 고령사회 대응 인구통계 발전방안
14:00 - 14:30		세션 3	**온누리** 통계청 사무관	국가통계 관리체계 개편안
14:30 - 15:00		세션 4	**김기태** UPS주식회사 대표	데이터 프라이버시 국제동향과 과제
15:00 - 15:30		세션 5	**한상태** 통계학회 부회장	보건소 모바일 헬스케어 상담 알고리즘 개발

MEMO

제11회 국가통계발전포럼
The 11th National Statistics Development Forum

기조연설

데이터의 변신과 국가의 역할

김용대 서울대학교 교수

기조연설

: 데이터의 변신과 국가의 역할

김용대
서울대학교 교수

학력

서울대학교 계산 통계학과
서울대학교 통계학과 석사
오하이오 주립대 통계학과 박사

경력

(현) 서울대학교 데이터사이언스대학원 교수
(현) 한국데이터마이닝학회장
(전) 미국 국립보건원 연구원

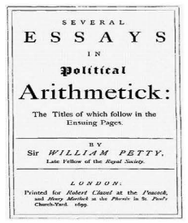
데이터의 변신과 국가의 역할

서울대학교
통계학과/데이터사이언스학과
김용대

목차

1. 데이터 변신

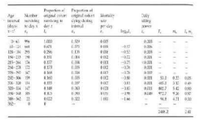
데이터의 탄생


SEVERAL
ESSAYS
IN
Political
Arithmetick:
The Titles of which follow in the Ensuing Pages.
BY
Sir WILLIAM PETTY,
Late Fellow of the Royal Society.
LONDON:
Printed for Robert Clavel at the Peacock, and Henry Mortlock at the Phenix in St. Paul's Church-Yard. 1690

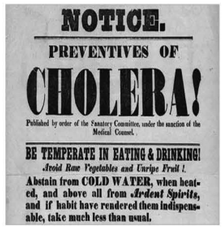
- 통치를 위한 도구
- 세금, 징집 등이 주 목표
- 17세기 정치산술: 국가통치를 위한 통계
- 인구총조사

민간 데이터

그란트 (17세기)



국민을 위한 데이터



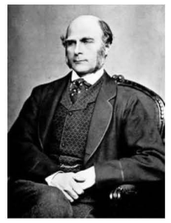
- 18세기 프랑스 대혁명 이후
- 국민이 주도
- 사회통계 개념 강화 (경제, 실업, 보건 등으로 확장)
- 정치적 중립성

사회학을 위한 데이터

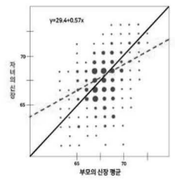
케틀러 (19세기)

평균인간
= 정상인간
= 우생학

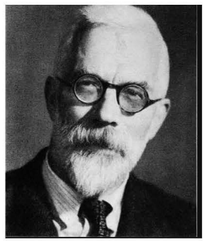
유전학을 위한 데이터



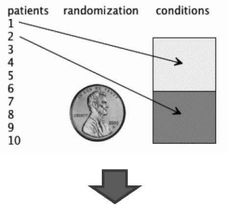
골턴 (19세기)



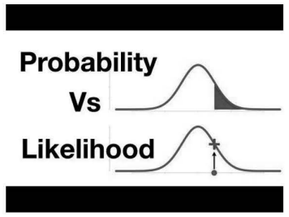
통계학의 등장



RA Fisher (19~20세기)



인과관계 규명 가능



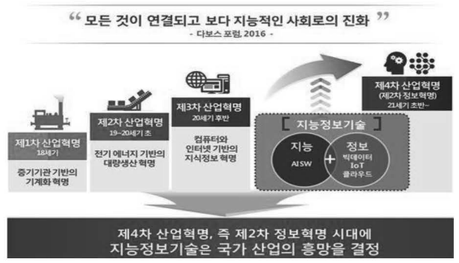
과학을 위한 데이터 분석 시대 개막

빅데이터 시대로 진입

데이터 기반 산업혁명의 시대

2. 4차산업혁명과 데이터

4차 산업혁명이란?



기업들이 제조업/서비스업과 정보통신기술(ICT)을 융합해 작업 경쟁력을 제고하는 차세대 산업혁명

4차 산업혁명이란?

글로벌 기업 **시가총액 순위**
2020. 1.9 기준

순위	거래소	회사명	시총(억달러)	2019.1.2 순위
1	사우디	아람코	18,200	(2019.12 상장)
2	미국	애플	13,300	3
3	미국	마이크로소프트	12,200	1
4	미국	알파벳	9,688	4
5	미국	아마존	9,380	2
6	미국	페이스북	6,138	6
7	미국	알리바바	5,848	8
8	미국	버크셔헤서웨이	5,527	5
9	홍콩	텐센트	4,797	7
10	미국	JP모건체이스	4,295	10
⋮				
15	한국	삼성전자	3,017	#23

자료/ 블룸버그 연합뉴스

장성구 기자 20200112
트위터 @yonhap_graphics 페이스북 tuney.kr/LeYN1

환경부가 적발한 배기가스 불법조작 차량 목록
(단위: 대)

제작사	차명	국내 판매량
아우디폭스바겐	아우디 A6 40 TDI 콰트로	3720
	아우디 A6 50 TDI 콰트로	403
	아우디 A7 50 TDI 콰트로	2533
	폭스바겐 투아렉 V6 3.0 TDI BMT	672
	폭스바겐 투아렉 3.0 TDI 4 모션	0
포르쉐	카이엔	2933

※ 국내 판매량은 환경부 추산. 폭스바겐 투아렉 3.0 TDI 4 모션은 국내 배출가스 인증을 받은 뒤 판매는 하지 않음 지료: 환경부

4차 산업혁명의 시대적 배경

인류의 역사

인간의 시대
키워드 : 생로병사, 행복, 인문학, 엔터테인먼트, 소통, 공감

4차 산업혁명의 시대적 배경

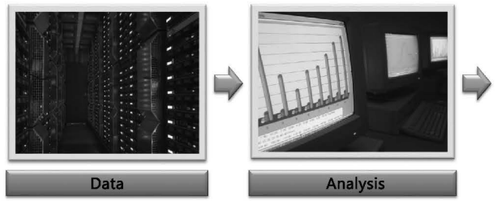
요즘 뜨는 기술

4차산업혁명과 빅데이터

- **데이터로부터 인간을 이해하는 기술**

데이터의 중요성

- 데이터로부터 새롭게 얻을 수 있는 지식의 양

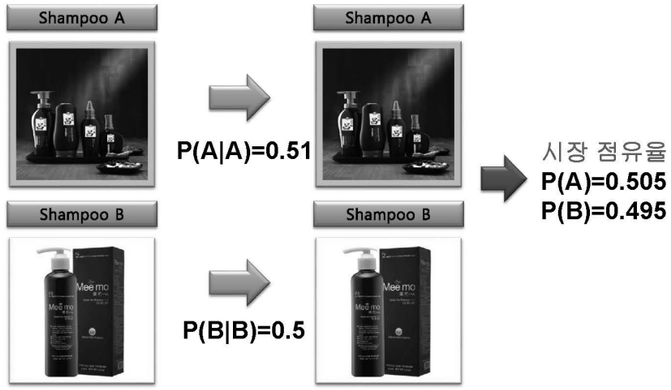


δ

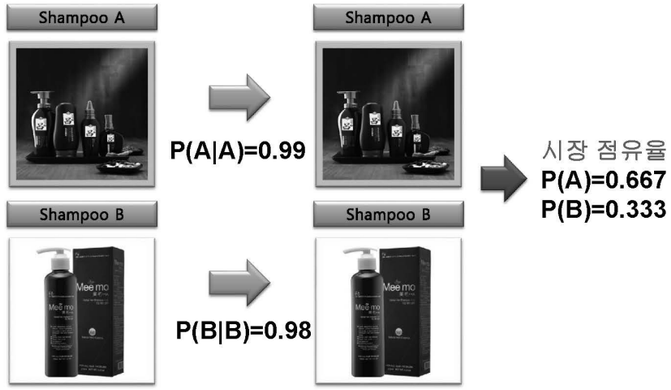
✓매우 작은
새로운 지식의 양

데이터의 중요성

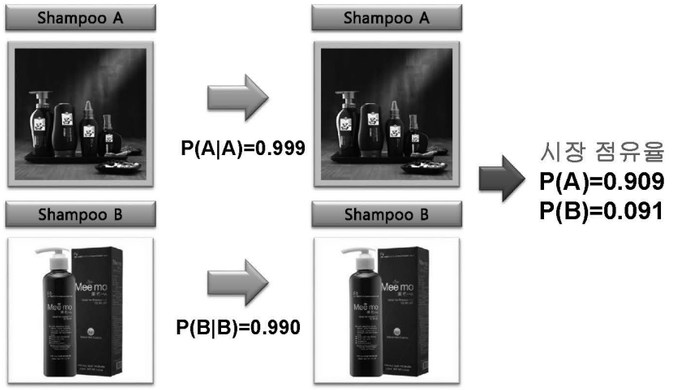
데이터의 중요성



데이터의 중요성



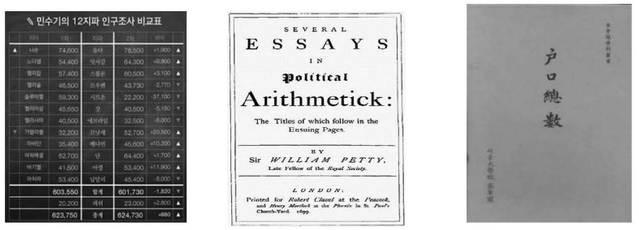
데이터의 중요성



3. 데이터 시대에서 국가의 역할

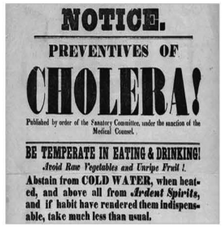
데이터의 활약상

국가통계의 어제



- 통치를 위한 도구
- 세금, 징집 등이 주 목표
- 정치산술: 17~ 18세기 유럽

국가통계의 어제



- 18세기 프랑스 대혁명 이후
- 국민이 주도
- 사회통계 개념 강화 (실업, 보건 등으로 확장)
- 정치적 중립성

통계청

국가통계의 오늘

STATISTICS KOREA 통계청 미래로 통하는 창

United States Census Bureau

- 공식 주요 행정기구
- 국민의 요구를 국가가 수행
- 시장경제의 파수꾼
- 사회적 갈등의 중재자

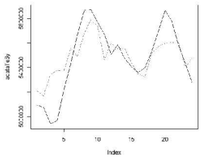
통계청

국가통계의 내일

- 데이터 경제의 콘트럴 타워
- 국가산업발전의 인프라

통계청

축산물이력제 데이터 기반 우유 생산 예측



통계청

두가지 제언

데이터 검색시스템

DATA 공공데이터포털 .GO.KR

KDX 한국데이터거래소

요청하신 콘텐츠 **검색 결과가 없습니다.**

검색 결과에 대한 문의는 고객센터로 연락 주시기 바랍니다.

[정보] AM 09:02 ~ PM 06:00 / 공휴일 및 주말휴무

개인정보보호 콘트럴 타워

'데이터 3법' 주요 내용

개인정보보호법 개정안	정보통신망법 개정안	신용정보법 개정안
-개인정보 관리·감독 기능을 개인정보보호위원회로 일원화 -가명정보(정보 제어)를 제공·서비스 개발에 활용	-온라인상 개인정보 관리·감독 권한을 개인정보보호위원회로 이관	-기명 정보 금융 분야 빅데이터 분석·활용 -기명 정보 주체 동의 없이 이용·제공 허용

(자료: 국회 각 상임위원회)

이루다AI 개인정보 유출 피해자 집단소송

통계청

MEMO

MEMO

제11회 국가통계발전포럼

The 11th National Statistics Development Forum

패널토론

류근관 통계청장

김용대 서울대학교 교수

이태수 한국보건사회연구원장

고길곤 서울대학교 교수

패널토론

류근관
통계청장

학력

1990	미국 스탠퍼드대 경제학 박사
1989	미국 스탠퍼드대 대학원 통계학과 졸
1985	서울대학교 대학원 경제학과 졸
1983	서울대학교 경제학과 졸

경력

2020. 12 ~ 現	통계청장
2004	서울대학교 사회과학대학 경제학부 교수
2004	국제통화기금 방문교수
2002	서울대 기업경쟁력연구센터 연구위원
2001, 2003	일본 오사카대 방문교수
2000	홍콩 과학기술대 방문교수
1998, 1999, 2000, 2002	미국 스탠퍼드 경제학과 방문교수
1998~2000	금융발전심의위원회 은행분과 위원
1995~2004	서울대 경제학부 조교수, 부교수
1990~1995	미국 UCLA 대학교 경제학과 조교수

김용대
서울대학교 교수

학력

서울대학교 계산 통계학과
서울대학교 통계학과 석사
오하이오 주립대 통계학과 박사

경력

(현) 서울대학교 데이터사이언스대학원 교수
(현) 한국데이터마이닝학회장
(전) 미국 국립보건원 연구원

이태수
한국보건사회연구원장

학력

연세대학교 경제학과 학사
연세대학교 경제학과 석사/박사

경력

(현) 한국보건사회연구원 원장 (2021~)
(전) 꽃동네대학교 사회복지학부 교수 (1991~2021)
(전) 균형발전위원회 위원 (2017~2019)
(전) 한국보건복지인력개발원 원장 (2004~2007)

고길곤
서울대학교 교수

학력

연세대학교 응용통계학 학사
서울대학교 행정대학원 정책학 석사
University of Pittsburgh 박사

경력

(현) 서울대학교 행정대학원 조교수, 부교수, 정교수 (2011~)
(전) National University of Singapore 조교수 (2007~2011)
(전) Asian Journal of Political Science 편집장

관련연구

질문주도 근거기반정책 방법론
빅데이터 플랫폼 재정사업 심층평가

MEMO

MEMO

제11회 국가통계발전포럼
The 11th National Statistics Development Forum

세션 1

동형암호 기법을 활용한 개인정보보호

송용수 서울대학교 교수

세션 1

동형암호 기법을 활용한 개인정보보호

송용수
서울대학교 교수

학력

서울대학교 수리과학부 학사/박사

경력

(현) 서울대학교 컴퓨터공학부 조교수 (2021~)
(전) Microsoft Research 수석연구원 (2019~2021)
(전) UC San Diego 박사 후 연구원 (2018)

Homomorphic encryption:

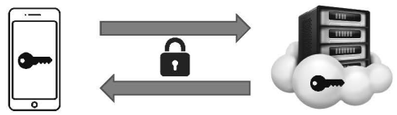
current status and future perspectives

송용수 (서울대학교 컴퓨터공학부)

국가통계발전포럼
August 27, 2021

Secure Computation

- 고전 암호학
  - 암복호화, MAC, 디지털 서명, …
  - 데이터 프라이버시, 무결성, 인증



- 현대 암호학
  - 추가적인 기능성을 제공
  - 데이터 보호와 활용성 보존
  - 데이터 뿐만 아니라 계산과정의 보호

Need for Advanced Cryptography

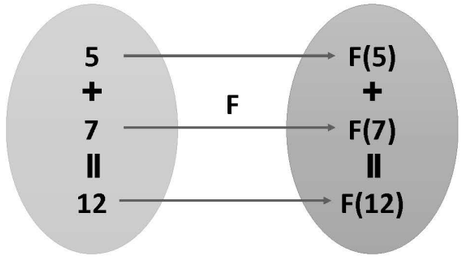
- 기존의 접근방법
  - 데이터 익명화, 랜덤화, 압축, …
  - 데이터 프라이버시와 정보량 사이의 trade-off
  - 필연적으로 정보 손실을 야기 = 데이터 활용성(기능성) 저하
- 서비스를 받기 위해서는 개인정보를 제공해야 함
- 다수 기관의 데이터 결합, 분석을 위해서는 신뢰할 수 있는 기관이 필요함
- 데이터 관련 정책
  - Health Insurance Portability and Accountability Act (HIPAA)
  - General Data Protection Regulation (GDPR)
  - 데이터 3법

Primitives

- 속성 기반 암호
- 검색 가능 암호
- 다중선형함수
- 함수 암호
- …
- 차분 프라이버시 (Differential privacy)
- 영지식 증명 (Zero-knowledge proof)
- 다자간 계산 (Multi-party computation)
- **동형암호 (Homomorphic encryption)**

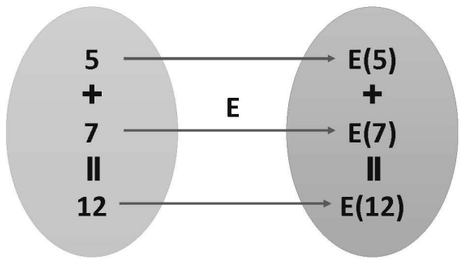
Thm (Informal): Anything that can be done **with trusted authority** can also be done **without!**

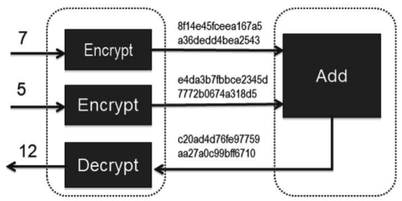
동형사상 (Homomorphism)



F : 동형사상 (w.r.t. 덧셈, 곱셈)

동형암호 (Homomorphic Encryption)

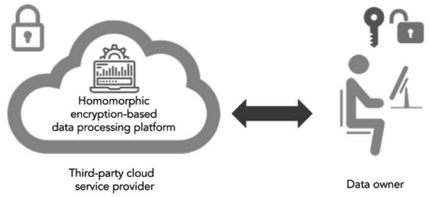




E : 동형암호 (w.r.t. 덧셈, 곱셈)

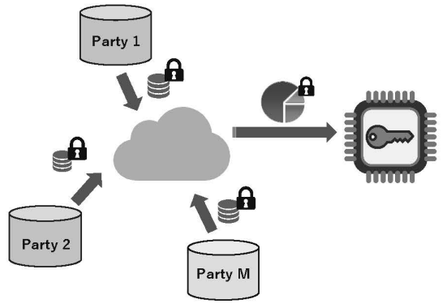
암호화된 데이터 상에서 복호화 없이 계산을 수행할 수 있는 암호화 기술

Use Cases of HE

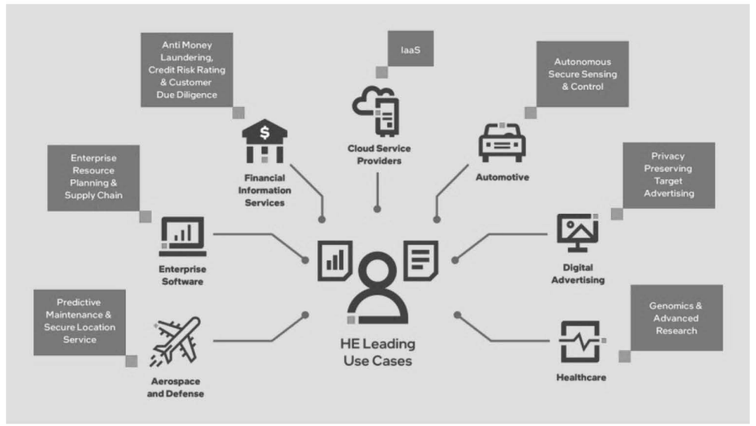


데이터 아웃소싱
(개인맞춤형 서비스)

다기관 데이터 결합/분석



Use Cases of HE



Main Challenges

- 성능 (Performance)
- 일반성 (Generality)
- 사용성 (Usability)

History of Homomorphic Encryption

- 2009 (Gentry): 최초의 동형암호
- 성능개선
  - 2011: PoC 구현 (30분 / gate)
  - 2012: BGV/BFV 스킴
  - 2013: IBM's HElib (4분 / 30K gates)
  - 2014 - 2016: 알고리즘 / 구현 최적화
  - 2016: TFHE 스킴
  - 2017: CKKS 스킴
  - 2019: HW 고속화
- 현재
  - Can be very efficient for lower-depth circuits

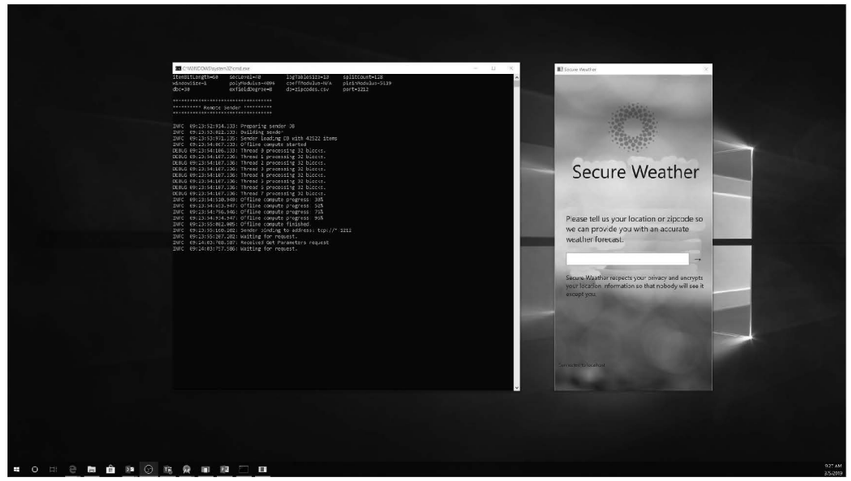
State of the art of HE schemes

*No clear winner – 서로 다른 장단점을 가진 3개의 스킴

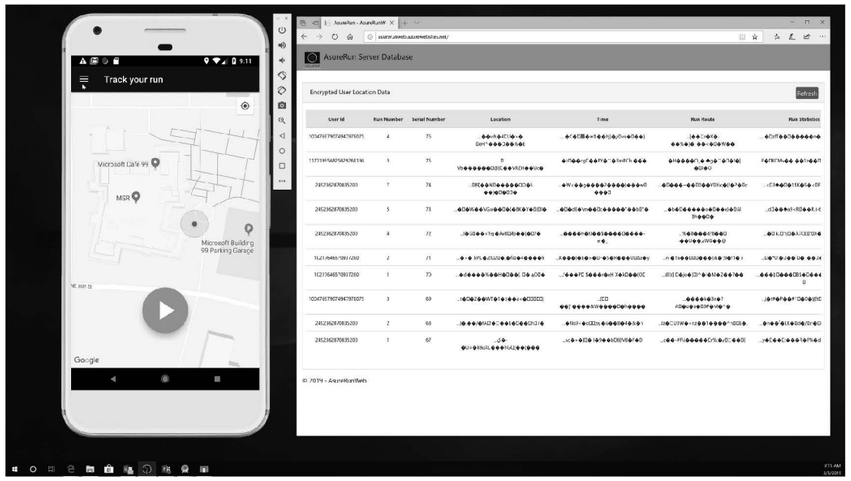
- **BGV, BFV** - 유한체 위에서의 벡터(병렬)화된 modulo 연산 (similar performance with CKKS)
- **TFHE** – 개별 비트 연산에 최적화 (13ms per gate)
- **CKKS** – 실수 상에서의 벡터화된 근사연산 (8192차원 벡터의 곱셈: 10~50ms, 숫자당 1~6 μs)

- 간단한 통계연산: very fast (20 ms ~ 0.5 s)
- 데이터베이스 검색/출력 : fast (1 ~ 10 s)
- 인공신경망 추론과정: acceptable (1 s ~ 10 mins)
- 간단한 기계학습 훈련과정: doable (10 mins ~ hours)
- 복잡한 인공신경망 훈련과정: somewhat slow (hours ~ days)

Demo1: Secure Weather

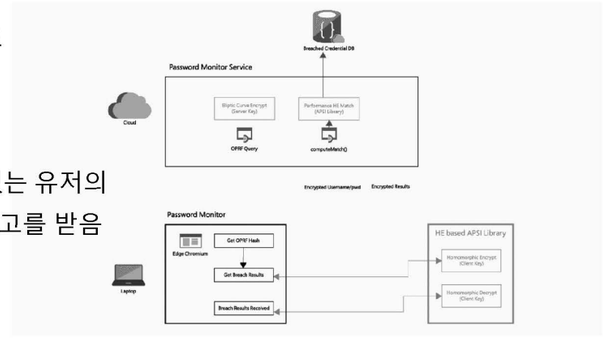


Demo2: AsureRun



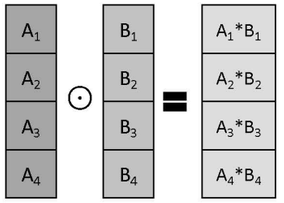
Password monitoring

- Microsoft 는 유출된 비밀번호 데이터베이스를 가지고 있음
- Edge 브라우저에 보관되어 있는 유저의 비밀번호가 유출되면 보안 경고를 받음



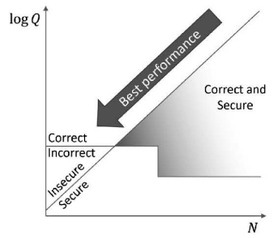
Not generally usable yet

- 기본적인 산술 연산만을 지원
  - 덧셈, 곱셈
  - 다른 연산의 어려움 (e.g. if, max/min, ReLU)
- 제한된 depth
  - 암호문은 유한한 '수명'을 가짐
  - 복잡한 '재부팅' 과정을 통해 복구 가능
- 병렬화
  - 하나의 암호문에 다수의 평문을 보관
  - 벡터(병렬) 연산 (in a SIMD manner)
  - Dimension : 2^{12} - 2^{15}
  - 벡터 순환연산 지원



Background Knowledge

- Distributed computing is already complex enough
  - "advanced crypto" adds secrecy considerations
- 좋은 성능을 위해서는 극한의 최적화가 필요함
  - 단순한 구현은 사용 불가능한 결과를 낼 수 있음
  - 응용 레벨에서의 약간의 변화가 구현/최적화 방법에 큰 차이를 야기함
- Simplicity / Usability vs Performance
- 대부분의 동형암호는 라이브러리는 편의성 보다는 성능 최적화에 치우쳐 있음

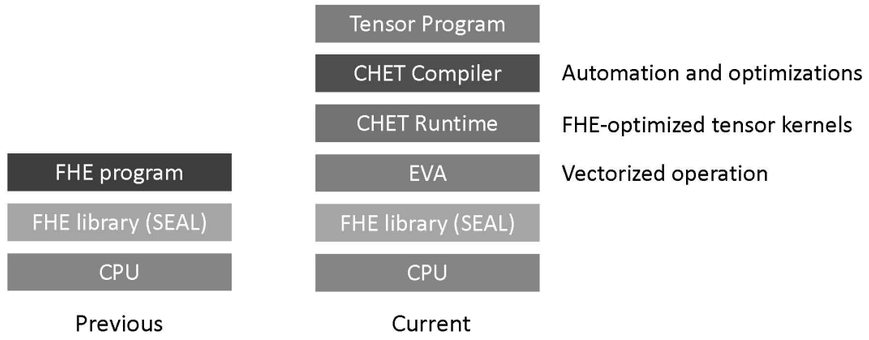


Example source code (SEAL)

```
size_t poly_modulus_degree = 8192;
parms.set_poly_modulus_degree(poly_modulus_degree);
parms.set_coeff_modulus(CoeffModulus::Create(poly_modulus_degree, { 60, 40, 40, 60 }));

/*
We choose the initial scale to be 2^40. At the last level, this leaves us
60-40=20 bits of precision before the decimal point, and enough (roughly
10-20 bits) of precision after the decimal point. Since our intermediate
primes are 40 bits (in fact, they are very close to 2^40), we can achieve
scale stabilization as described above.
*/
double scale = pow(2.0, 40);

SEALContext context(parms);
print_parameters(context);
cout << endl;

KeyGenerator keygen(context);
auto secret_key = keygen.secret_key();
PublicKey public_key;
keygen.create_public_key(public_key);
RelinKeys relin_keys;
keygen.create_relin_keys(relin_keys);
GaloisKeys gal_keys;
keygen.create_galois_keys(gal_keys);
Encryptor encryptor(context, public_key);
Evaluator evaluator(context);
Decryptor decryptor(context, secret_key);

CKKSEncoder encoder(context);
size_t slot_count = encoder.slot_count();
cout << "Number of slots: " << slot_count << endl;

vector<double> input;
input.reserve(slot_count);
double curr_point = 0;
double step_size = 1.0 / (static_cast<double>(slot_count) - 1);
for (size_t i = 0; i < slot_count; i++)
{
    input.push_back(curr_point);
    curr_point += step_size;
}
cout << "Input vector: " << endl;
print_vector(input, 3, 7);

cout << "Evaluating polynomial PI*x^3 + 0.4x + 1 ..." << endl;

/*
Plaintext plain_coeff3, plain_coeff1, plain_coeff0;
encoder.encode(3.14159265, scale, plain_coeff3);
encoder.encode(0.4, scale, plain_coeff1);
encoder.encode(1.0, scale, plain_coeff0);

Plaintext x_plain;
print_line(__LINE__);
cout << "Encode input vectors." << endl;
encoder.encode(input, scale, x_plain);
Ciphertext x1_encrypted;
encryptor.encrypt(x_plain, x1_encrypted);

/*
To compute x^3 we first compute x^2 and relinearize. However, the scale has
now grown to 2^80.
*/
Ciphertext x3_encrypted;
print_line(__LINE__);
cout << "Compute x^2 and relinearize:" << endl;
evaluator.square(x1_encrypted, x3_encrypted);
evaluator.relinearize_inplace(x3_encrypted, relin_keys);
cout << "    + Scale of x^2 before rescale: " << log2(x3_encrypted.scale()) << " bits" << endl;

/*
Now rescale; in addition to a modulus switch, the scale is reduced down by
a factor equal to the prime that was switched away (40-bit prime). Hence, the
new scale should be close to 2^40. Note, however, that the scale is not equal
to 2^40: this is because the 40-bit prime is only close to 2^40.
*/
print_line(__LINE__);
cout << "Rescale x^2." << endl;
evaluator.rescale_to_next_inplace(x3_encrypted);
cout << "    + Scale of x^2 after rescale: " << log2(x3_encrypted.scale()) << " bits" << endl;

/*
Now x3_encrypted is at a different level than x1_encrypted, which prevents us
from multiplying them to compute x^3. We could simply switch x1_encrypted to
the next parameters in the modulus switching chain. However, since we still
need to multiply the x^3 term with PI (plain_coeff3), we instead compute PI*x
first and multiply that with x^2 to obtain PI*x^3. To this end, we compute
PI*x and rescale it back from scale 2^80 to something close to 2^40.
*/
```

+100 lines to compute a cubic polynomial

User-friendly API and compilers



Conclusion

- 동형암호 기술의 성능은 다양한 응용분야에 적용할 수 있는 수준
  - 다른 기술과 접목될 수 있음
- 분야 진입장벽, 사용성 등 해결해야 할 이슈가 남아있음
- 암호 기술의 빠른 상용화를 위해서는 다양한 분야의 지식/전문가가 필요함

MEMO

제11회 국가통계발전포럼
The 11th National Statistics Development Forum

세션 2

저출산 고령사회 대응 인구통계 발전방안

최슬기 KDI 교수

세션 2

저출산 고령사회 대응 인구통계 발전방안

최슬기
KDI 교수

학력

서울대학교 사회학과 학사
미국 노스캐롤라이나대 사회학 석사/박사

경력

(현) KDI국제정책대학원 부교수 (2012~)
(현) 통계청 장래인구추계 자문위원
(전) 저출산고령사회위원회 인구재정분과 자문위원

저출산 고령사회 대응 통계 발전방안

2021. 8. 27.

최슬기

KDI SCHOOL
KDI School of Public Policy and Management

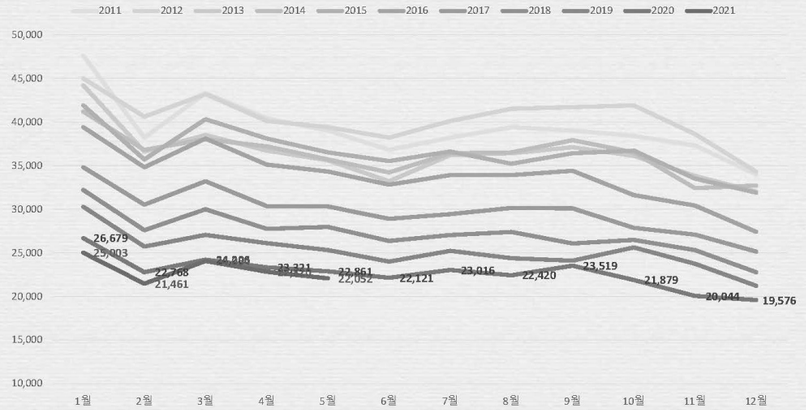
KDI School of Public Policy and Management KDI SCHOOL

연도별 출생아수와 출산율

연도	출생아수	합계출산율
2002	492,111	1.178
2003	490,543	1.191
2004	472,761	1.164
2005	435,031	1.085
2006	448,153	1.132
2007	493,189	1.259
2008	465,892	1.192
2009	444,849	1.149
2010	470,171	1.226
2011	471,265	1.244
2012	484,550	1.297
2013	436,455	1.187
2014	435,435	1.205
2015	438,420	1.239
2016	406,243	1.172
2017	357,771	1.052
2018	326,822	0.977
2019	302,676	0.918
2020(p)	276,000	0.84

KDI School of Public Policy and Management KDI SCHOOL

월별 출생아수 (통계청, 월별인구동향)



KDI School of Public Policy and Management KDI SCHOOL

『함께 일하고 함께 돌보는 사회』 더 촘촘하게 만들겠습니다

- '제4차 저출산 · 고령사회 기본계획(2021~2025)' 수립 -

• 0~1세 영아수당 신설, 영아기 집중투자 / 3+3 육아휴직제 도입, 아빠 육아휴직 문화 정착 / 다자녀가구 지원기준, 2자녀로 단계적 확대 / 인구변화 대응 사회 혁신 / 가족지원 투자 지속 확대 및 저출산·고령사회 투자예산 재구조화 등

• 신중년의 계속 고용 지원, 기초연금 확대 등 다층소득보장체계 강화, 지역사회 통합돌봄 전국 확대 등 활기차고 건강한 고령화 지원

- 1차 기본계획 2006~2010
- 2차 기본계획 2011~2015
- 3차 기본계획 2016~2020
- 패러다임 전환 2017, 재구조화 2018
- 4차 기본계획 2021~2025

KDI School of Public Policy and Management KDI SCHOOL

우리에게 필요한 정책

- 저출산대책: 원하는 만큼 자녀를 낳을 수 있는 사회
  - ➢ 이상자녀수와 실제 자녀수 간의 격차 줄이기
- 저출생대책: 인구구조 변화, 인구감소에 적응
  - ➢ 미래를 예측하고, 이를 미리 준비

KDI School of Public Policy and Management KDI SCHOOL

효과성과 효율성 있는 정책 필요

- 의도하고 있는 정책목표를 효과적으로 달성했는가? 얼마나 효율적이었나?
- 2018년 재구조화의 문제점

[정책 변화 방향]

	지금까지는	앞으로는
① 목표	출산율·출생아 수	2040 세대 삶의 질
② 접근방식	출산장려 캠페인 →국가주도 인식 개선	제도·구조 개혁 →개인의 합리적 선택
③ 정책 대상	육아기 부모, 저소득 위주	청년, 아동, 여성 행복, 서민, 중산층
④ 정책 주안점	보육	주거, 워라밸 강화, 모든 출생 존중
⑤ 실천 전략	새로운 제도	제도 활용 문턱 완화, 실천에 중점

➢ *정책 성과는 어떻게 측정할 것인가*

KDI School of Public Policy and Management KDI SCHOOL

저출산 관련 기본 통계 지표 산출

- 합계출산율
- 유배우출산율
- 혼인율
- 조정합계출산율
- 코호트완결출산율
- 이주배경인구출산율
- …

KDI School of Public Policy and Management KDI SCHOOL

조사자료와 행정자료 공개

- 2020년 중앙부처 저출산 정책 예산 약 40조원
  - 출산 양육비 부담 최소화 4.1조원 (아동수당에 3조원, 예방접종 지원 0.5조원, …)
  - 아이와 함께 하는 시간 최대화 1.4조원 (아빠육아휴직 보너스제 강화, 배우자출산휴가 확대에 1.2천억, …)
  - 촘촘하고 안전한 돌봄체계 13조원 (보육료지원 등 6.2조, 국공립유치원 확대 4조, 지역아동센터 내실화 4천억, …)
  - 모든 아동 존중과 포용적 가족문화 조성 4천억원 (저소득 한부모 가족 지원에 3.5천억원)
  - 2040 세대 안정적 삶의 기반 조성 21조원 (고용보험 사각지대 지원 1.1조, **청년주택 공급 5.3조, 청년 임차가구 주거비 지원 9.4조, 신혼부부 맞춤형 주택공급 3.3조**, 고교무상교육과 저소득층 교육급여 1.5조, …)

❖ ***정책 성과를 어떻게 측정할 것인지 Output과 Outcome을 구분하여 지표 제시 필요***

✓ *Input* *Process Evaluation*
✓ *Output* *Output Assessment*
✓ *Outcome* *Impact Assessment*

KDI School of Public Policy and Management KDI SCHOOL

제4차 저출산 고령사회기본계획 중 청년가구 대상 공급 확대 및 임차지원 정책

< 기본계획 과제 >

○ (청년맞춤형 임대 24만호 공급)청년 행복주택 및 매입·전세임대주택 등 다양한 공공임대주택을 청년층이 선호하는 곳에 저렴하게 제공

○ 임차주택 거주 청년층의 주거비 부담경감 위해 전·월세 청년금융지원 강화

– 공실 오피스·상가 주거용도 전환 용자 지원 신설

성과지표 (가중치)	성격	성과목표 구분	'21	'22	'23	'24	'25	측정산식/ 측정방법	자료수집 방법/출처	목표치 설정 근거	비고
청년맞춤형 행복주택 공급(만호) (0.4)	정량	목표	1.0	1.2	1.5	1.5	1.5	공공임대주택 건설형 공급물량 중 청년 등 젊은층에게 공급된 물량 (준공)	공공임대주택 공급실적 보고자료	주거복지 로드맵 및 기본계획상 공급목표	
		실적									
1.청년 매입임대 주택 공급(천호) (0.2)	정량/ 결과	목표	14.5	14.5	9.5	9.5	10	공급실적	공공주택 사업자별 공급실적 취합자료	주거복지 로드맵상 공급계획 반영	최종성과 =(지표① *0.5)+(지표②*0.5)
		실적	-	-	-	-	-				
2.청년 전세임대 주택 공급(천호) (0.2)	정량/ 결과	목표	10.5	9	9	9	9	공급실적	공공주택 사업자별 공급실적 취합자료	주거복지 로드맵상 공급계획 반영	
		실적	-	-	-	-	-				
3.공공지원 민간임대	정량	목표	1.4	1.4	1.4	1.4	1.4	공공지원민간임대 주택 중 청년층에 공급예정 물량 (부지확보)	공공지원 민간임대 주택 사업자별 공급실적 취합자료	주거복지 로드맵상 공급계획 반영	
		실적									
전·월세자금 수혜가구(호) (0.2)	결과	목표	80,000	80,000	80,000	80,000	80,000	주택도시기금 수요자 대출 통계구입 및 전월세대출	주택도시 보증공사	주거복지 로드맵2.0 (20.3.20) 지원목표	
		실적									

- 저출산 정책으로서 효과성 검증은?
- Output뿐 아니라 Outcome은 평가되고 있는가?

KDI School of Public Policy and Management KDI SCHOOL

제4차 저출산 고령사회기본계획 중 신혼부부 아동양육가구 주거지원 확대

< 기본계획 과제 >

○ 신혼부부 및 "만 6세 이하의 자녀가 있는 가구"에 오랫동안 안심하고 거주할 수 있는 공공주택 및 금융지원 75.4만 가구 지원

– 신혼희망타운 분양 8.4만호 본격 공급, 맞춤형 특화건설 임대단지 27만호 등 공공주택 35.4만가구 공급('21~'25)

□ 성과지표

성과지표 (가중치)	성격	성과목표 구분	'21	'22	'23	'24	'25	측정산식/ 측정방법	자료수집 방법/출처	목표치 설정 근거	비고
1.신혼부부 전세임대 주택 공급(천호) (0.5)	정량/ 결과	목표	11	13	11	11	11	공급실적	공공주택 사업자별 공급실적 취합자료	주거복지 로드맵상 공급계획 반영	최종성과 =(지표① *0.5)+(지표②*0.5)
		실적	-	-	-	-	-				
2.신혼부부 매입임대 주택 공급(천호) (0.5)	정량/ 결과	목표	15	16	10	10	10	공급실적	공공주택 사업자별 공급실적 취합자료	주거복지 로드맵상 공급계획 반영	
		실적	-	-	-	-	-				

- 저출산 정책으로서 효과성 검증은?
- Output뿐 아니라 Outcome은 평가되고 있는가?

KDI School of Public Policy and Management KDI SCHOOL

장래인구추계

- 미래 예측력 높이기
  - *Projection(추계)은 조건부 시산일 뿐 미래 예측이 아니다?*
  - 추계간 간격 줄이기
  - 장기추계
- 불확실성에 대한 적극적 고려
  - 확률론적 추계
  - 추계의 불확실성과 확실성 구분
- 시나리오별 이해도 제고

KDI School of Public Policy and Management KDI SCHOOL

KOSIS에서 30개 시나리오 제공

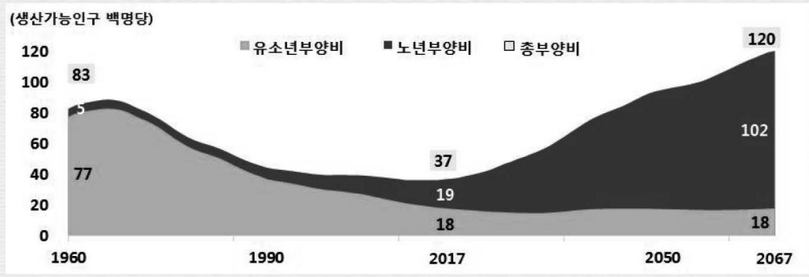
구분	시나리오 명칭	가정설정 수준		
		출산율	기대수명	국제이동
기본시나리오	중위 추계(기본 추계)	중위	중위	중위
	고위 추계(최대인구 추계)	고위	고위	고위
	저위 추계(최소인구 추계)	저위	저위	저위
조합시나리오	최고 출산율 추계	고위	중위	중위
	최저 출산율 추계	저위	중위	중위
	최고 기대수명 추계	중위	고위	중위
	최저 기대수명 추계	중위	저위	중위
	최대 국제순이동 추계	중위	중위	고위
	최소 국제순이동 추계	중위	중위	저위
	…	…	…	…
특별시나리오	국제무(zero)이동추계	중위	중위	무(zero)이동
	출산율현수준(2018년출산율 지속) 추계	2018년 출산율 지속	중위	중위
	출산율 OECD 평균 추계	OECD국가출산율 평균반등속도	중위	중위

KDI School of Public Policy and Management KDI SCHOOL

시나리오별 추계는,
추계 가정과 결과가
연계되어 제공되어야
이해와 활용도 제고 가능.
이것이 가능하도록 보다
Interactive하게 제시 필요

KDI School of Public Policy and Management KDI SCHOOL

생산연령인구와 부양비



- 다양한 부양비 지표 활용
  - 국민이전계정 (NTA; National Transfer Account)
  - 장래고령인구부양비 (POADR; Prospective Old Age Dependency Ratio)
  - 인지능력-조정 부양비 (CADR, Cognitive Ability Adjusted Dependency Ratio)
  - …

KDI School of Public Policy and Management KDI SCHOOL

감사합니다

MEMO

제11회 국가통계발전포럼
The 11th National Statistics Development Forum

세션 3

국가통계 관리체계 개편안

온누리 통계청 사무관

세션 3

국가통계 관리체계 개편안

온누리
통계청 사무관

학력

고려대학교 경제학과 학사

경력

통계청 통계조정과 및 경제통계심사조정과

국가통계 관리체계 개편 기본계획

통계청 Statistics Korea

경제통계심사조정과 (2021.08)

목 차

8/12/2021

I. 추진 배경

□ 4차 산업혁명으로 데이터 축적 가속화,
새로운 데이터를 활용한 통계작성 시도 증가

□ 국가승인통계 종수는 지속적으로 증가함에 불구하고,
승인요건을 갖추지 못한 다양한 형태의 통계도 증가

* 승인 통계 종수 : (06년 12월) 498종 → (10년 12월) 836종 → (20년 12월) **1,238종**
* 미승인 통계 현황: (16년) 73종 → (17년) 82종 → (18년) 86종 → (19년) 114종

⇒ 새로운 데이터 활용 및 다양한 형태의 통계에 대한 효율적 관리방안 마련 필요

8/12/2021

Ⅱ. 현황 및 문제점

1. 현황

현재 국가통계는

① **작성기관**이 생산하는 (통계법 제15조)

② **통계법 적용 대상이 되는 통계** 중 (통계법 제3조제1호, 시행령 제2조)

③ **승인요건을 충족**하는 통계에 대해 (통계법 제18조)

④ 국가통계로 승인하여 **공표하고 품질을 관리** (통계법 제9~11조, 제27조)

8/12/2021

<참고 1> 관련 법령

5

- **작성기관 요건 (통계법 제15조, 시행령 제18조)**
  ⇒ 중앙행정기관, 지자체는 당연지정, 공공기관이나 민간의 경우 ①민법이나 그 밖의 법률에 따라 설립된 법인일 것, ②통계작성 관련 조직, 인력과 예산확보(또는 확보예정),
  ③(구체적이고 실현가능한) 통계의 작성·보급에 관한 계획을 갖추어야 함

- **통계법 적용 대상 통계 (통계법 제3조 제1호)**
  ⇒ 통계작성기관이 정부정책의 수립·평가 또는 경제·사회현상의 연구·분석 등에 활용할 목적으로 산업·물가·인구·주택·문화·환경 등 특정의 집단이나 대상 등에 관하여 직접 또는 다른 기관이나 법인 또는 단체 등에 위임·위탁하여 작성하는 수량적 정보

- **통계법 적용 대상이 아닌 수량적 정보 (통계법 시행령 제2조)**
  ⇒ ①내부사용 목적 작성, ②시험적 작성, ③개인적인 학술연구, ④일상적인 업무수행 과정에서 보고·제출, ⑤주관적 인식·의견 조사 ⑥공공의 이익을 목적으로 작성한다고 보기 어려운 경우

- **승인요건 (통계법 제18조)**
  ⇒ ①기승인 통계와 유사하거나 중복되지 않을 것, ②통계의 신뢰성이 확보될 것,
  ③공공의 이익을 목적으로 할 것

8/12/2021

2. 문제점

6

1) 새로운 통계 작성·활용 미흡

○ 빅데이터 등 **새로운 데이터를 활용**하거나 **새로운 방식으로 작성**한 통계는 **기존 승인요건을 충족하기 어려움**

○ **승인요건을 충족하지 못하거나**, 각 정책부처가 가지고 있는 **행정자료 등의 수량적 정보**는 공표시 **통계법 위반 논란이 발생**

○ 2017년 연구용역 결과 457개 **공공기관에서 보유한 데이터 중 승인통계로 활용되는 것은 일부에 불과(16.8%)**

⇒ 빅데이터 등 새로운 통계에 대한 관리 및 정책부서 행정자료 양성화 필요

8/12/2021

2) 통계별 특성에 맞는 관리체계 미흡

○ 승인통계는 **작성방식에 따라** 조사통계, 가공통계, 보고통계로 구분되나 **통계관리방식은 통계별 특성과 상관없이 동일**

○ **작성기관으로 지정되면 모든 통계는 통계법의 엄격한 승인관리 대상이 되어 공공기관과 민간 협회, 단체에 과도한 부담으로 작용**

○ 통계의 중요도 및 활용성에 따른 차별화된 관리 부재로 **신뢰성 확보가 중요한 통계에 대한 집중적인 지원이나 관리 곤란**

⇒ 통계에 대한 맞춤형 관리, 통계작성기관 부담 경감 및 통계관리 효율화 필요

8/12/2021

Ⅲ. 추진 방향

8

<기본방향 : 통계 특성에 맞는 효율적 관리 추진>

1. 단기 추진방향 : 법령 개정 없이 제도개선을 추진
   ⇒ **빅데이터 활용 통계 등 새로운 방식의 통계 생산과 활용 활성화**

2. 중장기 추진방향 : 통계법 개정을 통한 관리체계 효율화
   ⇒ **국가 통계 관리체계 차등화 등을 통해 통계 작성·활용 확대 및 통계 품질 제고**

8/12/2021

Ⅳ. 추진 내용

9

1. 단기추진방향: 빅데이터 활용 통계관리방안 마련 : 실험통계 제도 도입

□ "실험통계(가칭)제도 도입" 추진

○ 개념 : 빅데이터 활용 등 기존 통계와는 다른 데이터나 새로운 방법론을 적용하여 작성한 통계로 작성 이후 적합성 및 타당성의 확인·점검이 필요한 통계

* 실험통계: 전통적인 조사방법 기준으로 보면 자료의 대표성과 신뢰성을 판단하기 어려운 통계
* 영국은 실험통계(Experimental Statistics) 용어 사용

○ 근거 : 통계법 시행령 제2조 2호에 따라 법 적용 대상에서 제외되는 시험적으로 작성하는 수량적 정보로 간주

8/12/2021

10

○ 관리방안

① 실험통계의 시험적 작성에 맞는 최소화된 기준을 적용

② 작성 기관이 확인을 요청하면 **실험통계(가칭)로 확인**

○ 판단 기준 (예시)

① 새로운 데이터나 방법론을 적용하여 작성되는지

② 추정 및 가공 방법은 검증된 통계적 기법을 사용하는지

③ 기존의 통계를 보완하여 상세한 통계 작성이 가능한지 등

8/12/2021

11

○ 모니터링 및 컨설팅 방안

① 시험 작성의 기간을 설정 (3년)

② 시험 작성의 적합성에 대한 모니터링 실시

③ 품질 제고를 위한 컨설팅 지원

○ 작성 완료

시험 작성의 적합성이 상실되거나, 작성 목적이 달성된 경우 작성이 완료된 것으로 판단

8/12/2021

12

2. 국가통계 종류 세분화 및 관리체계 효율화

1) 국가통계 종류 세분화

○ (기존) 승인·지정통계 → (개선) 신고·승인·지정통계

○ 신고제도를 도입하여 등록과 승인으로 관리 수준을 차등화

① 신고통계 : 기존통계와 중복성이 없고, 공공성, 대외 공표의 필요성이 있는 통계

* 주로, 정책부처에서 보유한 행정자료 등이 신고통계의 대상임
* 참고로, 기존 승인 통계들 중 '보고통계'도 신고통계로 전환

8/12/2021

13

② 승인통계 : 신뢰성과 타당성을 갖추어 통계청의 승인을 받은 통계로,
작성에 전문성이 요구되는 조사·가공통계는 원칙적으로 승인대상

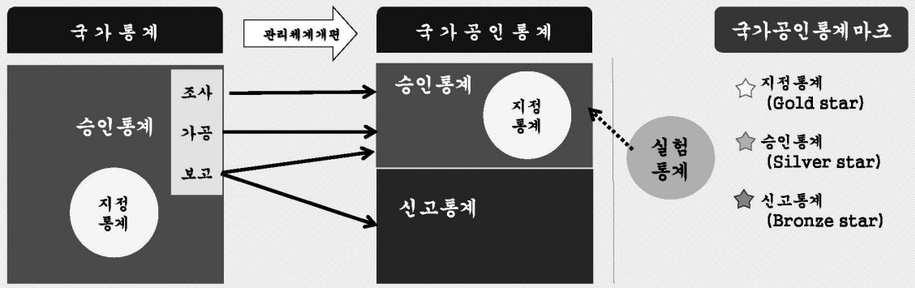
③ 지정통계 : 승인통계 중 통계법 제 17조에 따라 정책의 수립·평가 또는
다른 통계의 작성 등에 널리 활용되는 통계는 지정통계로 지정
* '21.3. 기준 92종 (인구총조사, 주택총조사, 경제총조사 등)

※ 실험통계는 제도 운영을 위한 근거조항을 신설하는 것으로 법령 개정 추진

8/12/2021

14

2) 국가통계 관리체계 효율화



8/12/2021

15

3. 국가 통계 품질관리 방안

1) 승인·지정통계 : 정기 + 자체 + 수시품질진단 실시

2) 신고통계 : X + 자체 + 수시품질진단 + 컨설팅 + 품질모니터링

3) 실험통계(가칭) : X + X + X + 컨설팅 + 품질모니터링

* 자체품질진단 권고

8/12/2021

16

4. 작성기관 범위 확대 검토

○ 실질적으로 행정기능을 수행하면서 통계를 작성하고 있는
법원행정처, 국회사무처 및 공공기관까지 통계작성기관 범위 적정화

* 삼권분립을 저해하지 않고 통계의 신뢰성 확보 목적 내에서 확대 검토하되,
입법부와 사법부의 소속기관은 논란이 있는 경우 작성기관 범위에서 제외

(예시) 예산정책처 통계작성·공표시 신뢰성 확보 필요성 제기

8/12/2021

Ⅴ. 향후 계획 및 기대 효과

17

□ **향후 계획**

① **통계승인업무 처리지침 개정 추진**(’ 21년 8월)

② 실험통계(가칭)제도 운영 및 결과 분석, 통계법령 개정 추진(’ 21년 하반기)

□ **기대 효과**

① 통계 관리범위 적정화로 **다양한 통계 작성 및 활용 활성화**

② 통계별 특성에 맞는 차등 관리로 **관리체계 효율성 제고**

8/12/2021

MEMO

MEMO

제11회 국가통계발전포럼
The 11th National Statistics Development Forum

세션 4

데이터 프라이버시 국제동향과 과제

김기태 UPS주식회사 대표

세션 4

데이터 프라이버시 국제동향과 과제

김기태
UPS주식회사 대표

학력

경남대학교 생물학과 학사
한라대학교 컴퓨터 공학과 석사과정

경력

(현) 한국인터넷진흥원 가명처리 검증전문가 과정 강의 (2017~)
(현) 금융보안원 가명처리과정 강사 (2017~)
(현) 비식별 컨설턴트 (2015~)
개인정보보호위원회 가명처리 전문가
한국인터넷진흥원 비식별 분야 자문위원
국세청, 국민연금, 금융보안원, 신용정보원, 한국인터넷진흥원,
금융결제원 반출심사위원

관련연구

한국인터넷진흥원 프레임워크 고도화연구
한국인터넷진흥원 비식별 프레임워크 연구
한국지능정보사회진흥원 비식별 기술 가이드라인 연구

1. 가명정보, 익명정보 개요

가명정보의 정의

개인정보보호법 제2조1호

"개인정보"란 살아있는 개인에 관한 정보로서 다음 각 목의 어느 하나에 해당하는 정보를 말한다.

가. 성명, 주민등록 번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보

나. 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보, 이 경우 쉽게 결합할 수 있는지 여부는 다른 정보의 입수 가능성 등 개인을 알아보는데 소요되는 시간, 비용, 기술 등을 합리적으로 고려하여야 한다.

다. 가목 또는 나목을 제 1호의 2에 다라 가명처리함으로써 원래의 상태로 복원하기 위한 추가정보의 사용·결합 없이는 특정 개인을 알아볼 수 없는 정보(이하 "가명정보"라 한다)

익명정보의 정의

(개인정보보호법 제 58조 2)

시간, 비용, 기술 등을 합리적으로 고려할 때 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 정보

1. 가명정보, 익명정보 개요

가명, 익명 정보에 대한 기준

- **가명정보의 기술적 기준 (ENISA)**
  - 식별 가능성 : 추가 정보를 사용하지 않은 상태에서의 가명정보의 분석을 통한 개인에 대한 식별 가능성
  - 복원 가능성 : 추가 정보를 사용하지 않은 상태에서 다시 원래의 정보로의 복원 가능성

- **익명정보의 기술적 기준 (ISO/IEC 20889:2018)**
  - 특정 가능성 : 데이터 주체를 고유 식별하기 위해 데이터 셋의 특성 집합을 관찰하여 개인에 속한 레코드를 격리(Isolation)해 낼 가능성이 없어야 함
  - 연결 가능성 : 동일한 데이터 주체 혹은 데이터 주체 그룹과 관련된 레코드를 별도의 데이터 셋에 연결하여 개인을 식별할 가능성이 없어야 함
  - 추론 가능성 : 무시할 수 없는 확률로 다른 속성 집합의 값에서 속성의 값을 추론하여 개인을 식별할 가능성이 없어야 함

  - → 이러한 기준들은 단순히 데이터의 처리 정도만으로 판단할 수 없음

1. 가명정보, 익명정보 개요

데이터 활용 환경

데이터 활용 환경

다양한 국가의 개인정보 비식별 모델에서는 데이터 활용 환경을 다음과 같이 분류

구분	UKAN의 Access Model	IPCO의 Release Model	NIST의 Data Sharing Models
완전 공개모델	Open Access	Publicly	The Release and Forget Model
제약적인 공개모델	Delivered access	semi-publicly	The Data Use Agreement(DUA) Model
한정된 장소 사용 모델	On-site safe settings	non-publicly	The Enclave Model
기타 모델	Virtual Access	없음	The Synthetic Data with Verification Model

- **데이터 활용 환경에 따라 필요한 수준의 가명 또는 익명처리가 필요**

→ 데이터 활용 환경은 어떤 요소들로 구성되어 있을까?

1. 가명정보, 익명정보 개요

데이터 활용 환경의 구성요소(ADF)

데이터 활용 환경의 구성요소(UKAN ADF 구성요소 3)

– **다른 데이터** : 활용을 하기 원하는 데이터에 연결되어 재식별을 가능하게 만들 수 있는 모든 정보, 다음과 같은 4가지의 세부 범주로 구분

- 개인 지식, 공개적으로 사용 가능한 데이터소스, 제한된 엑세스의 데이터소스, 기타 유사한 데이터의 공개여부

– **에이전트** : 데이터 흐름의 모든 시점에서 데이터에 대해 행위를 하고 상호작용할 수 있는 사람 및 단체

– **거버넌스 프로세스** : 에이전트와 데이터의 관계가 관리되는 방식, 데이터 엑세스 제어, 라이선스 계약 및 정책, 규범과 관행을 통한 비공식적인 행동 등

– **인프라** : 데이터가 흐르고 데이터 환경을 형성하도록 허용하는 구조와 설비를 포함하며 이러한 보안 인프라보다 광범위한 사회적이고 경제적인 구조가 포함됨

1. 가명정보, 익명정보 개요

•적정한 가명, 익명처리의 판단을 위한 조건•

- 적절한 수준의 가명, 익명처리를 위해서는 데이터에 대한 이해와 데이터를 이용하는 환경에 대한 이해가 모두 필요하고 이를 고려해야 함
- UKAN의 ADMF에서는 데이터와 데이터를 이용하는 환경을 모두 고려하는 것을 데이터 상황(Data Situation)이라고 하며 이는 NIST의 Context와 Content를 합한 것과 동일한 표현임
- 데이터 이용환경과 데이터에 대해 정확하게 판단하는 것은 매우 어려우며 이를 판단하기 위해 대부분의 국가들에서는 Risk Based Approach(위험기반 접근법)를 사용
- 위험기반의 접근법을 통한 위험에 대한 파악과 이에 대한 처리, 그리고 사후관리를 모두 포함한 Framework를 각국에서 만들거나 만들고 있으며 국제 표준에서도 이를 제정 중 (영국, 미국, 호주, 대한민국, ISO 등)

1. 가명정보, 익명정보 개요

또 하나의 문제

- 개인의 식별성을 낮추기 위해 다양한 기술을 사용
- 연소득 78,654,325원에 노이즈를 추가하여 연소득 78,723,158원으로 변경
  - ENISA의 가명처리 기술문서 : 블러링 기법을 사용
  - EU의 29번 작업반 비식별 기술문서 : Pertrubation(섭동) 기법을 사용
  - IHE IT Infrastructure Handbook De-Identification : Fuzzing기법을 사용

→ 동일한 기법에 대해 서로 다른 용어를 사용함으로써 혼돈을 가지고 오게 됨

→ 이러한 혼돈을 해결하기 위한 표준의 제정이 필요

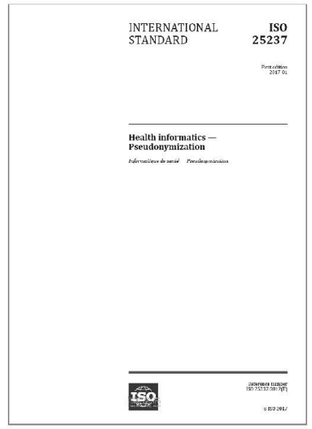
2. 해외의 프라이버시 동향

비식별 관련 국제표준의 동향

2017년

ISO 25237:2017
Health informatics — Pseudonymization

- 건강정보의 가명화에 대한 국제 표준
- 보건의료 정보의 종류, 보건의료 데이터의 비식별 프로세스 등을 정의
- 가명처리 기술에 대한 정의는 IHE의 IHE_ITI_Handbook_De-Identification-Mapping 를 인용


INTERNATIONAL STANDARD ISO 25237

Health informatics — Pseudonymization

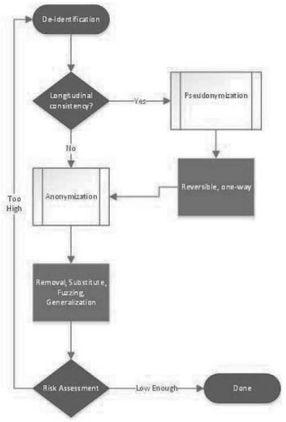
2. 해외의 프라이버시 동향

비식별 관련 국제 표준의 동향

2017년

ISO 25237:2017
Health informatics — Pseudonymization

- 가명처리 프로세스
- 가명정보의 재식별 위험을 판단하기 위한 3가지 레벨의 위험을 정의
  - Level1: 개인의 데이터 요소를 식별하는 것과 관련된 위험
  - Level2: 데이터 변수 집계와 관련된 위험
  - Level3: 데이터의 특이치와 관련된 위험
- 이러한 위험을 해결하기 위한 3가지의 보증 수준을 정의
  - 보증수준 1 : 데이터를 명확하게 식별하거나 간접적으로 쉽게 식별 가능한 데이터 제거, 재식별 가능성이 낮은 환경 고려시 충분한 보호 제공
  - 보증수준 2 : 외부 데이터를 사용하는 공격자를 고려한 수준, 다른 정보와의 결합가능성도 고려하는 수준
  - 보증수준 3 : 특이치에 대한 식별가능성까지 고려하는 수준



2. 해외의 프라이버시 동향

비식별 관련 국제 표준의 동향

2017년 2018년

ISO 25237:2017

ISO/IEC 20889 : 2018
Privacy enhancing data deidentification terminology and classification of techni

- 비식별 처리 기술의 표준
- 다양한 나라의 비식별 표준에서 기술에 대한 명칭의 불일치 등을 해결하기 위한 표준
- 정형데이터의 비식별 기술에 국한
- 개인식별자에 대해 가명과 익명에 맞춰 다음과 같이 소개

데이터 운영 환경을 고려한 기준	직접 식별자 (Direct identifier)	특정 분명 환경에서 데이터 주체를 고유하게 식별할 수 있도록 해주는 속성
	간접 식별자 (Indirect identifier)	데이터셋에 포함되어 있거나 외부에 속한 속성과 함께 특정 운영 환경에서 데이터 주체의 고유 식별을 가능하게 하는 속성
하나의 데이터셋 기준	고유 식별자 (Unique identifier)	데이터셋에서 데이터 주체를 독립적으로 선정(singles out)해 내는 데이터셋에서의 속성
	준식별자 (Quasi-identifier)	데이터셋에서 데이터셋에 포함된 다른 속성과 함께 고려할 때, 데이터 주체를 선정(singles out)해 내는 속성


INTERNATIONAL STANDARD ISO/IEC 20889

Privacy enhancing data de-identification terminology and classification of techniques

2. 해외의 프라이버시 동향

비식별 관련 국제 표준의 동향

2017년 2018년

ISO 25237:2017

ISO/IEC 20889 : 2018
Privacy enhancing data deidentification terminology and classification of techniques

- 비식별 기술을 8개 범주의 21개 기술로 소개하고 각 기술별 재식별 공격에 대한 효과를 부록에 정리

기술	세부기술	기술	세부기술
통계도구 (Statistical Tools)	샘플링(Sampling)	가명화 기술(Pseudonymization techniques)	
	총계처리(Aggregation)	해부화(Anatomization)	
암호화 도구 (cryptographic tools)	결정성 암호화 (Deterministic encryption)	일반화 기술 (Generalization techniques)	라운딩(Rounding) 제어라운딩(Controlled rounding)
	순서보존 암호화 (Order-Preserving encryption)		상하단 코딩 (Top and Bottom coding)
	형태보존 암호화 (Format-Preserving encryption)		특성집합을 단일특성으로 결합 (Combining a set of attributes into a single attribute)
	동형암호화 (Homomorphic encryption)		로컬 일반화 (Local genetrlization)
	동형비밀보존 (Homomorphic secret sharing)	무작위화 기술 (Randomization techniques)	잡음 추가(Noise addition)
삭제기술 (suppression Techniques)	마스킹(Masking)		순열(Permutation)
	로컬 삭제 (Local suppression)		부분총계(Microaggregation)
	레코드 삭제 (Record suppression)	재현데이터 (Synthetic data)	

2. 해외의 프라이버시 동향

비식별 관련 EU의 동향

2017년 2018년 2018년

ISO 25237:2017 ISO/IEC 20889 : 2018

ENISA
Recommendations on shaping technology according to GDPR provisions – An overview on data pseudonymisation

- 가명화 기술 및 가명화의 개념에 대한 안내서
- 2018년 11월 발간
- 가명화 기술
  - Hashing without key
  - Hashing with key or salt
  - Encryption as a pseudonymisation technique
  - Other cryptography-based techniques
  - Tokenisation
  - Other approaches

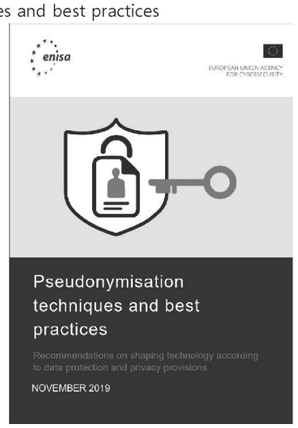
2. 해외의 프라이버시 동향

비식별 관련 EU의 동향

2017년 ISO 25237:2017 ISO/IEC 20889 : 2018 | 2018년 ENISA | 2018년 | 2019년

ENISA
Pseudonymisation techniques and best practices

- 가명화 기술과 가명처리의 모범적인 사례를 소개
- 데이터 사용 방법에 따른 6가지의 시나리오를 소개하고 다양한 형식의 데이터에 대한 가명처리의 사례를 소개
- 가명화 기술
  - Counter
  - Random number generator
  - Cryptographic hash function
  - Message authentication code
  - Encryption



2. 해외의 프라이버시 동향

비식별 관련 EU의 동향

2017년 ISO 25237:2017 ISO/IEC 20889 : 2018 | 2018년 ENISA | 2018년 | 2019년 ENISA | 2021년

ENISA
Data Pseudonymisation: Advanced Techniques and Use Cases

- 암호화의 고급 기술을 이용한 가명화 기술과 헬스케어, 사이버시큐리티등에서의 실제 사용 사례를 소개
- 가명화 기술
  - ASYMMETRIC ENCRYPTION
  - RING SIGNATURES AND GROUP PSEUDONYMS
  - CHAINING MODE
  - PSEUDONYMS BASED ON MULTIPLE IDENTIFIERS OR ATTRIBUTES
  - PSEUDONYMS WITH PROOF OF OWNERSHIP
  - SECURE MULTIPARTY COMPUTATION
  - SECRET SHARING SCHEMES

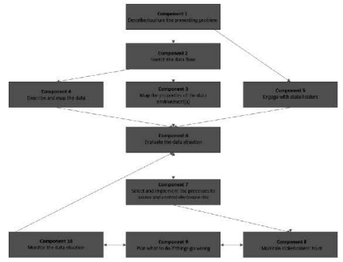
2. 해외의 프라이버시 동향

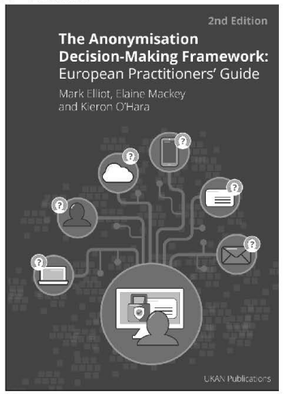
비식별 관련 영국의 동향

2017년 ISO 25237:2017 ISO/IEC 20889 : 2018 | 2018년 ENISA | 2018년 | 2019년 ENISA | 2021년 ENISA | 2021년

UKAN
The Anonymisation Decision-Making Framework : 2nd Edition

- 익명화에 대한 의사결정을 위한 프레임워크
- 2016년 초판에 이어 2021년 개정판을 발간
- 데이터 상황감사, 공개 리스크 평가 및 제어, 영향 관리에 대해 총 10단계의 단계별 가이드를 제공





2. 해외의 프라이버시 동향

비식별 관련 국제 표준의 동향

2017년 ISO 25237:2017 ISO/IEC 20889 : 2018 | 2018년 ENISA | 2018년 | 2019년 ENISA | 2021년 ENISA | 2021년 UKAN ADF II | 2022년

ISO/IEC 27559
Privacy-enhancing data de-identification framework

- 비식별 데이터의 수명주기동안 데이터를 적절하게 비식별화 하고 데이터 주체와 신뢰를 구축하고, 법적인 규정 준수 요구 사항을 충족할 수 있도록 하는 표준
- 컨텍스트 평가 : 데이터의 이용환경에 대한 평가
- 데이터 평가 : 데이터 및 데이터로 인한 잠재적인 공격 이해
- 식별 가능성 평가 : 부적절하게 익명화된 이용 가능한 데이터로부터 어떻게 식별이 발생하는지 이해
- 거버넌스 : 데이터의 활용과 연관된 사람들의 제어(적절한 역할과 책임)를 결정
- 2022년 표준 제정을 위해 진행 중

3. 과제

과제

- 미국 HI-Trust에서도 HIPAA 기반의 프레임워크를 가지고 있으며 헬스케어 분야에서 널리 활용 중
- 국내에서도 한국인터넷진흥원, 한라대학교 김순석교수, 유피에스데이터의 김기태 대표 등이 국내 법제에 적합한 비식별 프레임워크를 2019년 연구하였으며 개인정보보호법 개정 이후 이의 고도화를 수행
- 위험기반 접근법의 가명처리, 익명처리를 수행할 수 있도록 하고 있으며 2021년 다양한 산업군에서 이의 실증을 수행 중
- 가명처리, 익명처리의 안전성에 관해서는 프레임워크를 통해 많은 부분 충족 가능
- 이제 가명정보, 익명정보의 유용성에 대한 연구가 고민이 필요

감사합니다.

MEMO

MEMO

제11회
국가통계발전포럼

The 11th National Statistics Development Forum

데이터 시대와 국가통계의 역할