

K-통계시스템 구축방안[안]

2021. 2. 1.

통 계 청

보도자료	보도일시	2021. 02. 01.(월) 15:00	 통계청
	배포일시	2021. 02. 01.(월) 09:00	
	담당부서	통계데이터허브국 통계데이터기획과 통계정책국 통계정책과	
	담당자	과 장: 박상영(042-481-2313) 사 무 관: 백승항(042-481-2314) 과 장: 빈현준(042-481-2051) 사 무 관: 김병주(042-481-2426)	

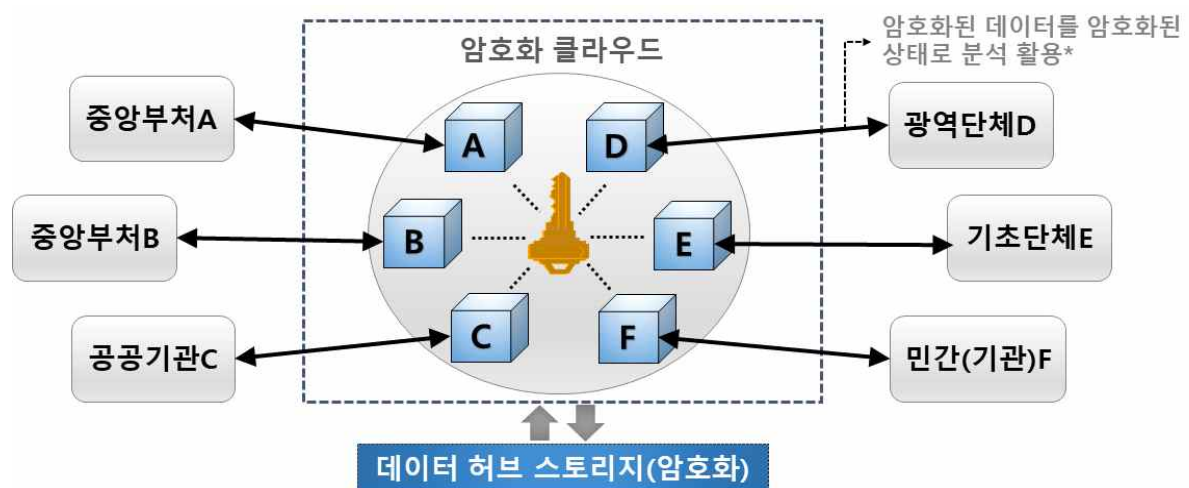
통계청, K-통계시스템 구축 추진

- 新 기술로 개인정보 보호하며 빅데이터 활용성 제고
- 데이터 결합 통한 부가가치 창출 및 4차 산업 활성화 지원 기대

통계청, K-통계시스템 구축해 나간다.

K-통계시스템이란 통계청만이 보유하고 있는 통계등록부 등 데이터베이스를 가교(架橋)로 삼아 각 정부기관과 공공기관에 산재되어 있는 데이터를 최신의 암호기술을 기반으로 연계한 클라우드 기반의 데이터 체계로 이해할 수 있다.

< K-통계시스템 개념도 >



* 암호화된 분석 결과를 활용을 위해 복호화할 때 필요한 복호화키는 별도 관리

데이터 3법 개정 · 제도개선에 대응하고 데이터 수요에 부응 필요

통계청이 새롭게 K-통계시스템 구축에 나서고자 하는 것은 데이터 이용 활성화를 위한 데이터 3법(개인정보보호법, 정보통신망법, 신용정보법) 개정 및 제도 개선에 따라 한편으로는 과학적 연구, 상업적 목적의 통계작성, 공익적 기록보존 등의 목적으로 정보주체의 동의 없이 가명정보의 활용이 허용되면서 대량의 데이터 저장 및 활용이 확산되는 가운데 다른 한편으로는 개인정보 유출 및 오남용에 대한 우려 등 불안감이 확대되고 있기 때문이다. 또한 데이터 간 연계 및 결합 활용에 대한 수요가 증가하고 있음에도 불구하고 유용한 데이터들이 정부 각 부처와 지방자치단체 등 공공기관에 산재한 채로 각각 분리 · 보관되고 있어 데이터의 잠재적 가치가 충분히 실현되지 않고 있기 때문이다.

최고의 보안수준을 유지하면서 데이터의 잠재적 가치를 극대화

통계청이 최근 연구가 활발하게 진행되고 있는 동형암호 등 최선의 암호기술을 이용하여 최고의 보안수준을 유지한 채 각종의 공공자료를 결합하여 활용할 수 있는 한국판 공공 빅데이터 체계를 구축하게 되면 데이터 집중화에 따른 개인정보 유출이나 빅 브라더 출현 등 우려를 불식하면서도 규모의 경제와 범위의 경제로 대표되는 데이터의 잠재적 가치를 극대화할 수 있을 것으로 기대하고 있다. 데이터는 일반적으로 각종 자료의 크기가 클수록, 또 보다 많은 특성이 결합할수록 그 가치가 배가되는 특징을 가지고 있기 때문이다.

경쟁과 혁신 생태계 구축 및 학문 발전, 산업계의 인센티브 제공

통계청은 K-통계시스템 구축을 통해 공공정책의 효율성과 형평성 제고는 물론 민간부문의 4차산업 활성화에도 크게 기여할 수 있을 것으로 기대한다. 나아가 공공 빅데이터 활용에 대한 접근성이 개방·확대되면서 신규 창업기업과 벤처기업을 적극적으로 지원할 수 있고, 기존 빅데이터 보유 플랫폼 기업과 신규 진입하는 기업 간 자료 격차를 축소하여 경제 전반의 경쟁과 혁신의 생태계를 촉진하는 데 큰 도움이 될 것으로 예상된다.

아울러 암호 상용화를 위한 학문적 발전과 산업계의 동형암호 가속기 등 개발에도 적극적인 인센티브를 제공하여 ICT 강국인 우리나라가 빅데이터 및 암호 상용화 분야에서도 세계에서 선도적 위치를 차지할 수 있는 큰 계기를 마련할 수 있을 것으로 기대하고 있다.

이용자 시각에서 통계·데이터 서비스 패러다임 획기적 전환

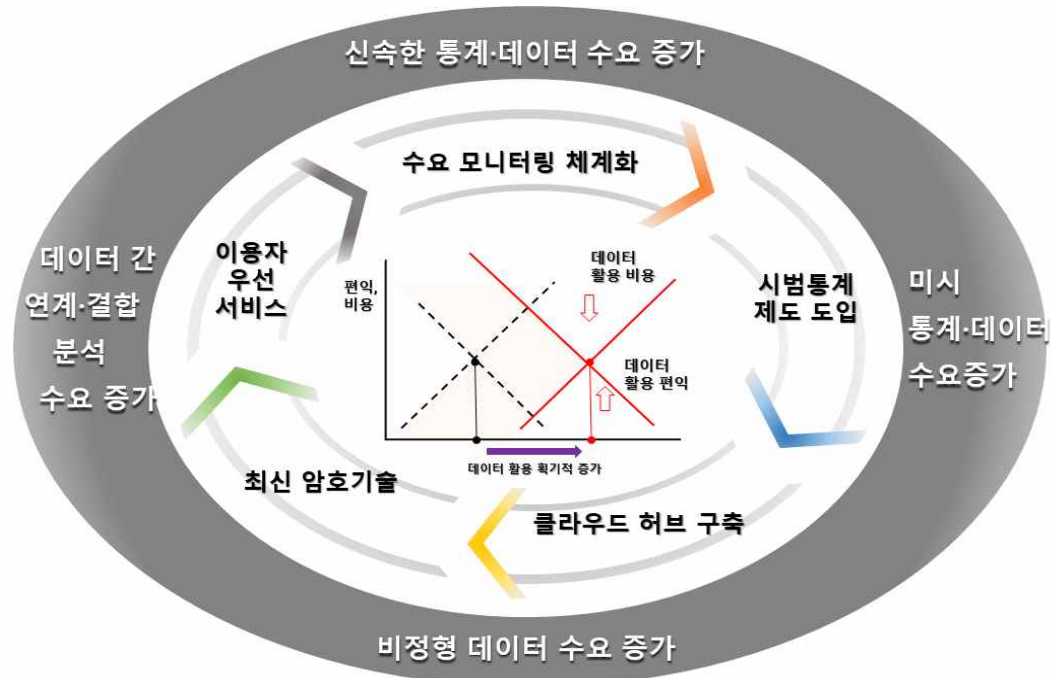
한편, 통계청은 데이터 간 연계·분석 및 맞춤형 정책 대응을 위한 보다 촘촘한 정보 생산 등 통계이용자의 수요 변화에 대응하여 국가통계관리도 탄력적으로 개선할 계획이다.

이를 위해 통계청은 우선 시범통계 제도를 마련하고 다양한 데이터 활용을 통한 통계작성을 촉진해 나가고자 한다. 시범통계란 빅데이터 활용 등 기존통계와 다른 데이터나 방법론을 적용하여 작성한 통계로 작성 이후 일정 기간 품질 제고를 위해 관찰이 필요한 준(準)통계를 의미한다. 이와 함께 통계청은 시범통계제도 도입에 맞추어 국가승인 통계와의 차이 및 운영방식 등을 구체화하여 통계이용자의 혼란을 방지할 계획이다.

이와 함께, 통계이용자의 자료 요청에 대하여 제공 가능 여부에 대한 판단이 모호한 경우 충분한 자료보안을 전제로 과학적 연구와 정부정책 근거 마련 등의 목적으로 일정 기간 자료에 대한 열람과 활용을 임시적으로 허용하기 위한 통계자료 샌드박스 제도도 도입할 계획이다.

이를 통해 통계·데이터의 양적인 면만 아니라 질적인 측면에서도 서비스 확대가 가능해지고 데이터의 접근성과 활용성이 확대되어 통계이용자가 체감하는 만족도도 높아질 것으로 기대하고 있다.

통계청은 국가데이터 활용 활성화를 통한 국가 혁신을 뒷받침하기 위한 K-통계시스템 구축사업이 성공적으로 추진될 수 있도록 관계부처와의 긴밀한 협력과 사회적 공감대 확산을 위한 노력도 병행해 나갈 계획이다.



【붙임 1】 K-통계시스템 구축방안(안)

K-통계시스템 구축방안[안]

2021. 2. 1.

통 계 청

순 서

I. 추진 배경	1
II. K-통계시스템 개요	2
III. 잠재적 응용 분야	3
IV. 통계·데이터 서비스 확대 추진	4
[참고1] 통계·데이터 환경 변화의 주요 특징	5
[참고2] 통합통계등록부 개요	6
[참고3] 동형암호기술 및 재현자료	7

I. 추진 배경

① 데이터 3법 개정 및 제도 개선 등 데이터 환경의 급변

- 과학적 연구, 상업적 목적의 통계작성, 공익적 기록보존 등의 목적으로 정보주체의 동의 없이 가명정보 활용 허용

② 법 개정에 따른 대량의 데이터 저장 및 활용 확산 예상

- 데이터는 크기가 클수록, 또 보다 많은 특성이 결합될수록 그 가치가 배가되는 특성을 보유
- 데이터 잠재가치를 극대화하기 위한 연계 및 결합 수요 증대

③ 개인정보 재식별 및 오남용 우려 등 불안감 확대

- 개인정보를 익명·가명화해도 다양한 원인에 의한 재식별화 가능성으로 인해 프라이버시 침해 및 법적 책임 문제 대두
- 강화된 보안체계에서도 외부 해킹, 관리 실수, 내부자에 의한 개인정보 고의 유출 위험 존재

* 일반적으로 데이터 결합시 정보보호를 위한 보호기법을 강하게 적용할수록 데이터 가치가 저하되는 데이터 보호와 활용상 상충관계 존재

④ 각 기관이 보유한 유용한 데이터가 분절화되어 운용

- 데이터 집중화에 따른 개인정보 유출 및 빅 브라더 출현 등 우려에 따라 데이터 공유의 어려움이 상존
- 유용한 데이터들이 정부부처와 지방자치단체 등 공공기관에 산재한 채 분리·보관되어 데이터의 잠재적 가치 실현 곤란
- 기관 데이터를 다른 기관이 활용하기 위한 신뢰체계 미확립

⇒ 중앙정부 및 지방자치단체가 보유한 공공 데이터를 안전하게 연계하여 누구나 활용할 수 있는 통계시스템 개발 필요

II. K-통계시스템 개요

◇ (정의) 기관별로 산재된 데이터를 최신의 암호기술을 기반으로 최고의 보안 수준 하에 결합 활용하는 한국판 공공빅데이터 체계

❶ 중앙부처, 지방정부, 공공기관이 암호화된 데이터를 실시간으로 업로드하고 분석 가능한 클라우드 기반의 시스템 개발 추진

- 연계의 기초로 활용될 수 있는 기관별 대용량 데이터를 암호화된 형태로 데이터 허브 클라우드 공간에 적재·활용

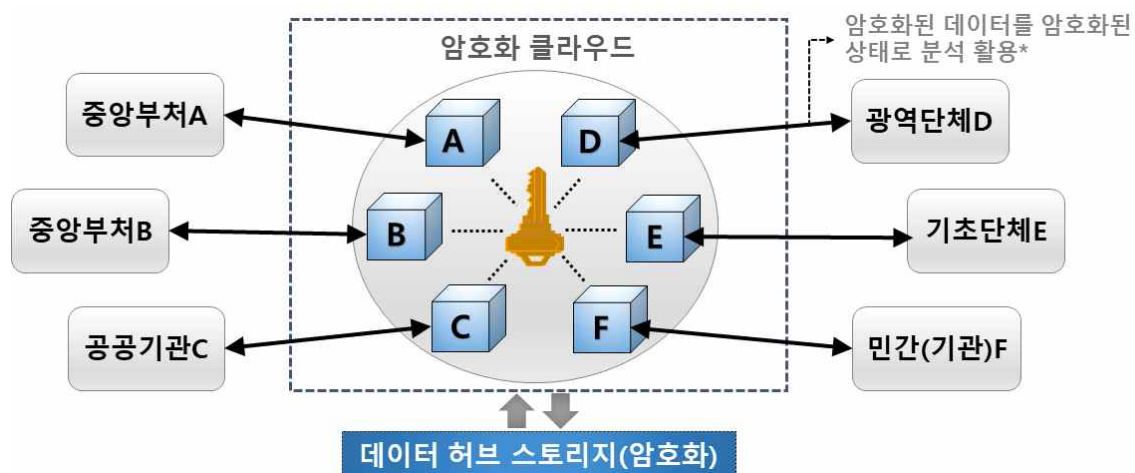
* 기관별 보유 평문데이터를 암호화된 형태로 변환하여 클라우드 공간에 각각 분리·저장하고 암호화된 상태에서 분리된 데이터 간 결합·연산이 가능

❷ 데이터 보유기관의 경우 K-통계시스템 참여 시 데이터 간 자유로운 연계·분석이 가능해 자발적인 참여 유인*을 제공

* 데이터는 규모의 경제와 범위의 경제가 가장 크게 실현되는 재화의 특성을 갖고 있어 K-통계시스템 참여를 통해 해당 기관의 데이터 가치 증대 효과

❸ 암호화된 상태의 데이터 적재·활용으로 개인 민감정보 유출 및 데이터 간 연계·결합에 따른 '빅 브라더' 우려 불식

<K-통계시스템 도식(예시)>



* 암호화된 분석 결과를 활용을 위해 복호화할 때 필요한 복호화키는 별도 관리

⇒ 최고 보안수준을 유지하면서 데이터의 잠재적 가치를 극대화

Ⅲ. 잠재적 응용 분야(예시)

① 공공정책의 효율성과 형평성 제고

- (효율성) 재정지출 심층분석 및 모델 개발
 - 클라우드 허브 구축 시 재정지출 자료(input)와 개인·가구 단위 재분배효과 자료(output)를 연계·결합함으로써 상세한 재정지출 효과 분석 및 맞춤형 정책 모델 개발 가능
- (형평성) 근거기반(evidence-based) 맞춤형 정책수립 제고
 - 형평성 있는 정책 대상자·지역 선정 등을 위한 기초자료 제공 가능

② 4차산업 활성화 기여

- 창업기업 및 벤처기업에 공공 빅데이터를 적극 지원하여 기존 빅데이터 보유 플랫폼 기업과 신규 진입 기업 간 격차 축소
 - 4차산업 활성화를 위한 경제 전반적인 경쟁과 혁신의 산업 생태계 구축에 기여

③ 빅데이터 및 암호 상용화 부문의 세계적 선도

- K-통계시스템은 암호 상용화를 위한 학문적 발전, 산업계의 동형암호 가속기 등 개발에 적극적 인센티브를 제공
- 한국이 ICT 강국을 넘어 빅데이터와 암호 상용화 분야에서도 세계에서 선도적 위치를 차지할 계기로 작용

⇒ 경쟁과 혁신생태계 구축 및 학문 발전, 산업계 인센티브 제공

IV. 통계 · 데이터 서비스 확대 추진

① (시범통계) 통계 다양성 확대를 위해 단기적으로 시범통계 제도 도입을 추진하여 데이터 활용 통계작성 활성화

- * 시범통계(Pilot Statistics)란 빅데이터 활용 등 기존통계와 다른 데이터나 방법론을 적용하여 작성한 통계로 작성 이후 품질 제고를 위한 관찰 기간이 필요한 통계
- 통계작성을 위한 가용정보 확대 및 수요 변화에 대응하여 탄력적으로 국가통계를 관리
- 시범통계제도 도입에 따른 시범통계와 국가승인통계와의 차이, 운용방식 등을 구체화하여 통계이용자 혼란을 방지
- 시범통계제도 도입에 맞춰 국가통계를 승인통계, 지정통계, 공공통계, 시범통계로 세분화하고 차등관리 방안 마련 검토
- * (예시)고도의 전문성이 요구되는 조사·가공통계 및 행정통계는 승인통계로 집중 관리하고, 공공·시범통계는 신고제도·자체관리를 통한 자율성, 책임성 확대

② (샌드박스) 특정 기관·이용자를 대상으로 통계 데이터 이용 범위를 제한적으로 넓혀주는 통계자료 샌드박스* 도입 검토

- * 제공 가능 여부 판단이 모호한 경우 충분한 자료보안을 전제로 일정기간 동안 과학적 연구, 정부정책 근거 마련 등 목적으로 자료 열람·활용 임시 허용
- (임시허가) 개인정보 등 식별정보를 제외한 자료 요청의 경우, 보안 환경에서 일정 기간 연구목적으로 자료 활용 가능
- (보안유지) 자료 재제공, 식별화 시도, 오남용 분석 등 부작용 발생에 대비한 접속 즉시 차단 등 보안 대비 태세 유지

⇒ 이용자 시각에서 통계 · 데이터 서비스 패러다임을 획기적으로 전환하여 이용자 체감 만족도 제고

① 시의성 높은 통계·데이터 수요 증대

① 정책환경 변화에 적기 대응을 위한 시의성 높은 데이터 요구 증가*

* 코로나19 실시간 상황판, 농가 및 차량이동 데이터 분석을 통한 조류독감 확산 예측 등 실시간 데이터의 품질과 유용성이 제고

② 국가승인통계의 87%(=1,075/1,238종)가 1년 이상 주기('20)이며 보다 빠른 주기(분기, 월, 일 등)의 정보 생산 확대 필요

② 미시 통계·데이터 수요 증가

① 맞춤형 정책 수립을 위한 소지역·소분류 기준 통계/분석 수요 확대*

* 지역별 정책에 활용하고자 공표 단위(전국·광역)보다 자세한(시·군·구) 고용통계 요구

② 미시자료의 유용성과 대표성 등 한계에 대한 균형 있는 접근 필요

③ 데이터 간 연계·결합, 분석 수요 증가

① 자료 간 연계·분석을 통한 데이터 활용도 제고에 대한 요구 지속

* 112 신고와 신용카드 데이터 결합으로 범죄위험도에 따른 경찰력 배치, 통화·택시 승하차 데이터 연계로 심야버스 노선 수립 사례 등

② 데이터의 가치는 선제적으로 알기 어렵고 연계·분석 등의 과정을 통해 유용성이 발견되고 새롭게 수요를 창출

④ 비정형 데이터의 증가

① '16~'20년 세계 데이터 총량 중 80%는 비정형 데이터라고 예측(IDC)

* 그림, 영상 등 형태가 구조화되어 있지 않고 직접 연산이 곤란한 데이터

1 통계등록부 개요

- ① 통계등록부는 인구, 사업체 등에 대해 조사자료(통계청 조사)와 각 부처 행정자료를 연계하여 작성한 통계단위별 모집단 자료

❖ 통계등록부 활용 현황

- (정책평가) 중기부는 기업등록부를 활용하여 중소기업기본통계 공표*('20.10.29.)
* 중소기업이 전체 기업의 99.9%, 종사자의 83.1% 비중 확인(9988 논란 종식)
- (행정효과) 등록부 기반 인구총조사* 및 경제총조사** 실시로 국가 예산 절감
* 인구주택총조사 : '15년 1,257억 → '20년 949억(△308억)
** 경제총조사 : '16년 456만 사업체, 610억 → '21년 620만 사업체, 571억

2 통합통계등록부 개요

- ① 통합통계등록부는 기구축한 통계등록부에 소득-교육 등 추가 자료를 입수하여 가명화된 방식으로 연계하여 구축
- 정부 정책 수립에 필요한 핵심 데이터 플랫폼 구축
 - '21년에는 법무부와 협업하여 외국인통계등록부 신규 작성

< 통합 통계등록부 체계도(안) >



- ② 인구, 가구, 기업 등의 핵심 속성정보를 종합적이고 상세하게 파악하여 정책수립 및 의사결정 고도화 지원
- 민간도 보유자료와 연계하여 심층분석 및 연구에 활용 가능

1 동형암호기술(Homomorphic Encryption)

- ① 동형암호는 평문과 암호문에서 같은 성질이 유지된다는 뜻으로 평문과 암호문의 연산결과가 값은 값을 가짐
 - 1970년대에 이론연구가 시작된 이후 2009년에 IBM 연구원인 Gentry에 의해 기술적 가능성이 증명됨



- ② 암호화상태에서 연산이 가능하여 정보유출 가능성이 낮으나, 아직까지 대용량화에 따른 낮은 연산속도 개선이 필요

2 재현자료(Synthetic Data)

- ① 전통적인 노출제한 방법의 정보 손실 문제를 극복하기 위해 구축된 원자료와 유사한 분포를 가지는 가상의 데이터
 - 실제 측정된 원자료와 다른 자료이지만 원자료를 적절하게 설명하는 가상의 모형에서 생성하여 모형의 특성이 유지됨

