

의안번호	제2022-03호
상 정	2022. 2. 21.
년 월 일	[통계데이터분과위원회]

**통계데이터센터 활성화를 위한
미래 발전전략 보고**

제 출 자	통 계 청 장
제출년월일	2022. 2. 21.

요 약

통계데이터센터 활성화를 위한 미래 발전전략 [요약]

◆ 4차 산업혁명 시대에 증거기반 정책환경과 데이터 생태계에 신속히 대응하기 위하여 **이용자 중심의 통계데이터센터 미래 발전전략** 수립

I 추진 배경

- 최근 데이터 3법 개정 및 4차 산업혁명 시대에 **데이터 개방 확대** 요구가 고조되는 등 **데이터 환경 변화**에 신속히 대응할 필요
- 동형암호와 같은 **신기술**을 이용하여 국민이 안전하게 데이터를 활용할 수 있는 K-통계 구축 등 새로운 **미래 발전과제** 발굴

II 추진 성과 및 평가

- 거점 지역센터 확충을 통해 **이용자 접근성**이 점차 개선되고 있고, 제한된 공간을 통해 **데이터 구축·개방 확대**에 부응하고 있으나,
 - 지역센터를 중심으로 유관기관 등과 협업을 통한 데이터 연계·활용 실적이 미흡하고, 적극적인 **연계·활용**을 위한 추가적인 노력이 필요
- 주문형 분석서비스, 재현자료 등 편의성이 크게 개선되고 있으나, **SDC/RDC 분리운영, 오프라인 위주의 서비스** 등 이용자 불편 지속

III 추진 전략

비전	데이터 혁신 기반 디지털 국가 경제 선도
목표	증거기반 정책지원과 데이터 생태계 변화 대응을 위한 이용자 중심 통계데이터센터 활성화
4대 추진전략	12개 추진과제
1 인프라 강화	<ul style="list-style-type: none"> 1. 관련 법·규정 등 제도 정비 2. 신기술을 통한 K-통계체계 적용 3. 유관기관과 대외협력 체계 구축
2 데이터 연계 활성화	<ul style="list-style-type: none"> 1. 핵심 데이터 서비스 확대 2. 데이터 연계·결합 전문기관 위상 제고 3. 데이터 연계·분석 서비스 강화
3 이용자 중심 서비스	<ul style="list-style-type: none"> 1. 데이터 서비스 일원화 및 확대 2. 지역 거점 클러스터 위상 강화 3. 포스트 코로나 대비 서비스 다변화
4 센터운영 내실화	<ul style="list-style-type: none"> 1. 통계데이터 미래조직 설계 2. 통계데이터센터 운영 효율화 3. 이용자 교육·홍보 및 인재양성 전문화

IV 주요 내용

1 인프라 강화

- 통계법 개정, 실험통계 및 이용자 가이드라인 등 법·규정을 정비하고, **최신기술**을 통한 K-통계체계 구축을 위하여 **test-bed 역할** 수행
- 다양한 **자료간 연계·결합**을 위하여 가명정보 결합전문기관, 플랫폼 발전협의체 등 데이터 유관기관과 **대외협력 체계**를 구축·운영

2 데이터 연계 활성화

- 통계등록부 및 민간데이터 확충 등 **핵심 데이터 서비스**를 확대하고, 금융데이터 연계·결합 활성화를 위한 **데이터전문기관** 지정 검토
- 통계등록부와 MD를 연계하여 대체번호를 부여하고, 이종 자료간 결합을 통한 맞춤형 DB를 구축하는 등 **연계·분석서비스 강화**

3 이용자 중심 서비스

- SDC와 RDC를 **단계적으로 통합**하여 서비스를 일원화하고, 통계등록부 및 마이크로데이터의 **제공범위 확대** 및 **이용체계 개선**
- 데이터 접근성 보장과 격차 해소를 위하여 **지역센터**를 신설하고, 지방행정 정책지원을 위하여 **클러스터 구축·운영**을 위한 기능개편

4 센터운영 활성화

- 국가데이터 활용강화를 위하여 통계데이터 **미래조직 설계**하고, 자료이용 전 단계에 대해 운영시스템 기능개선을 통해 **효율화**
- 데이터전문가를 **지역센터에 배치**하여 인재양성 기회를 제공하고, 기술노트 모음집 및 우수사례집 등을 발간하여 이용자에게 자료 공유

V 기대 효과

- 증거기반 정책지원을 위한 국가데이터 **연계 활성화** 기틀을 마련하고, 민간기업 지원 **활성화** 및 우리 경제의 **경쟁과 혁신** 촉진

VI 향후 일정

- 국가통계위원회 **안전 보고**(‘22.2월) 및 **진행사항 점검**(‘22.3월 이후)

목 차

목 차

I. 추진 배경	1
II. 주요 성과 및 평가	2
1. 그간의 추진성과	2
2. 성과평가	3
III. 추진 전략	4
IV. 주요 내용	5
1. 인프라 강화	5
2. 데이터 연계 활성화	7
3. 이용자 중심 서비스	9
4. 센터운영 내실화	11
V. 기대 효과	12
VI. 향후 일정	12
[참고 1] 통계데이터센터 운영 현황	13
[참고 2] 통계데이터센터 제공 자료	17

I. 추진 배경

□ 데이터 환경 변화에 신속히 대응할 필요

- 최근 데이터 3법 개정과 함께 자율주행, 인공지능, 빅데이터 기반 4차 산업혁명 시대에 데이터 개방 확대 요구 고조
- 데이터 활용을 통한 사회적 가치 창출을 위하여 데이터 활용 활성화와 개인정보보호 간의 균형에 대한 사회적 인식의 변화
 - 데이터 활용 촉진을 위한 가명정보 결합전문기관 제도 도입 운영중
- 코로나19 등으로 대면서비스 중단 상황에 대응하기 위한 비대면 통계서비스 지원방안이 보다 우선적으로 검토 필요

□ 데이터 패러다임 대전환을 위한 미래 발전전략 마련

- 통계데이터센터 출범('18년) 이후 그간의 추진성과에 대한 중간 점검과 함께 한 단계 도약을 위한 미래발전 과제발굴 필요
 - 포스트 코로나 및 이용자 편의제고를 위해 데이터센터 방문 없이 분석이 가능한 클라우드 기반의 비대면 서비스로 전환 필요
- K-통계체계 구축 등 발전전략 이행을 위한 새로운 도전과제 검토
 - 동형암호, 재현자료, 차등정보보호 등 데이터보안 신기술을 이용하여 국민 누구나 안전하게 데이터를 활용할 수 있는 플랫폼 제공

□ 외부 지적사항에 대한 해결방안 마련

- 이용자 접근이 쉽도록 부산, 광주, 대구 등 지역센터를 확충하고 있으나, 국회, 언론 등에서 이용 활성화 등을 지속적으로 요구
 - 직제개편 및 운영 효율화 등 통계데이터센터 활성화 방안 마련

⇒ 급변하는 대내외 환경 및 증거기반 정책에 신속히 대응하고 한 단계 도약하기 위하여 미래 발전전략 수립이 필요

II. 주요 성과 및 평가

1. 그간의 추진성과

□ 데이터 인프라 및 서비스

- 통계데이터센터는 '18년 정식 개소한 이후 지역센터 신설*, 맞춤형 신규 서비스 확대 등으로 이용실적이 양적으로 빠르게 성장

* 지역별 거점센터: 3곳('18년) → 3곳('19년) → 4곳('20년) → 6곳('21년)

(단위: 건, %)

구분	'18년	'19년	'20년(A)	'21년(B)	증감률 (B-A)/A
방문 이용 건수(건)	78	290	369	554	50.1
방문 이용 인원(명)	175	522	616	801	30.0
주문형 분석 건수	-	15	17	32	88.2

- 아동가구통계DB, 인구동태 코호트DB 등 통계등록부 구축*과 카드 매출정보 등 민간데이터 구매**로 데이터 콘텐츠 다양화 노력

* 통계등록부: 10종('18년) → 12종('19년) → 14종('20년) → 16종('21년)

** 민간데이터: 22종('18년) → 27종('19년) → 33종('20년) → 35종('21년)

- 통계데이터센터 이용자 대상 만족도 점수* 매년 지속적으로 상승

* 만족도: 3.65점('18년) → 3.86점('19년) → 4.09점('20년) → 4.09점('21년)

□ 데이터 저변 확산 및 이용 활성화 선도

- 통계데이터 분석·활용대회는 코로나19에도 불구하고 데이터 관심 증가, 제공자료 확충, 온라인서비스 등으로 많은 대학생, 일반인 등 참가(154팀)
 - 이용자들이 제출한 기술노트, 활용수기 등을 활용하여 우수사례 발굴

- 이용자 중심의 역량 강화를 위한 사전분석용 재현자료 생성 시스템 구축·운영과 함께 이용자 맞춤형 교육 확대 운영

- 데이터 프라이버시 신기술 선행연구 및 데이터센터 모범사례 제시
 - 동형암호, 재현자료, 차등정보보호 등 신기술 연구를 주도하고, 기재부, 국세청, 전북, 고용정보원 등 유관기관에 센터 설립·운영 노하우 전수

2. 성과평가

□ 데이터 인프라 및 서비스

- 부산, 대구, 광주 등 거점 지역센터 확충을 통해 데이터 사각지대 해소 및 이용자 접근성이 점차 개선되고 있으나,
 - 전체 실적 중 서울(52.5%), 대전(32.1%)과 그 외 지역센터간 실적 불균형
 - 지방청별 거점 지역센터를 중심으로 지자체, 대학, 유관기관 등과 협업을 통한 데이터 연계·활용 실적은 아직 미흡한 상황
- 각종 통계등록부와 함께 통신사, 신용카드 등 민간데이터가 제공되고 있으나, 보다 다양한 자료구축 및 항목 세분화 필요

□ 데이터간 연계·활용 지원

- 물리적으로 제한된 공간을 통해 개인정보 등 민감한 정보가 포함된 데이터 구축·개방 확대라는 당초 설립 취지에 부응하고 있으나,
 - 다양한 통계 수요에 대응한 데이터 개방을 넘어 좀 더 적극적인 데이터의 연계·활용을 위한 추가적인 지원 노력이 필요
- 다른 기관의 개인정보간 연계·결합을 활성화하기 위하여 개인정보 보호법에 따른 결합전문기관(‘20.11.)으로 지정·운영 중으로
 - 통계청은 국립암센터 등과 함께 최초 가명정보 결합 시범사례로 ‘폐암환자의 사망동향 및 사망예측 연구’ 결과를 발표(‘21.5.)
 - 반면, 카드, 부채 등 금융데이터 연계는 불가능하여 이를 활성화하기 위한 신용정보법에 따른 데이터전문기관 지정 검토 필요

□ 이용자 편의 제고

- 주문형 분석서비스, 온라인시각화, 재현자료 제공 등 다양한 통계 서비스 제공을 통해 이용자 편의성이 크게 개선되고 있으나,
 - 비식별화, 재현자료 등에 대한 추가연구, SDC/RDC 분리 운영에 따른 혼란 및 오프라인 위주의 서비스 등 이용자 불편은 지속

Ⅲ. 추진 전략

비전	데이터 혁신 기반 디지털 국가 경제 선도
목표	증거기반 정책지원과 데이터 생태계 변화 대응을 위한 이용자 중심 통계데이터센터 활성화
4대 추진전략	12개 추진과제
1 인프라 강화	<ol style="list-style-type: none"> 1 관련 법·규정 등 제도 정비 2 신기술을 통한 K-통계체계 적용 3 유관기관과 대외협력 체계 구축
2 데이터 연계 활성화	<ol style="list-style-type: none"> 1 핵심 데이터 서비스 확대 2 데이터 연계·결합 전문기관 위상 제고 3 데이터 연계·분석 서비스 강화
3 이용자 중심 서비스	<ol style="list-style-type: none"> 1 데이터 서비스 일원화 및 확대 2 지역 거점 클러스터 위상 강화 3 포스트 코로나 대비 서비스 다변화
4 센터운영 내실화	<ol style="list-style-type: none"> 1 통계데이터 미래조직 설계 2 통계데이터센터 운영 효율화 3 이용자 교육·홍보 및 인재양성 전문화

IV. 주요 내용

전략 1 인프라 강화

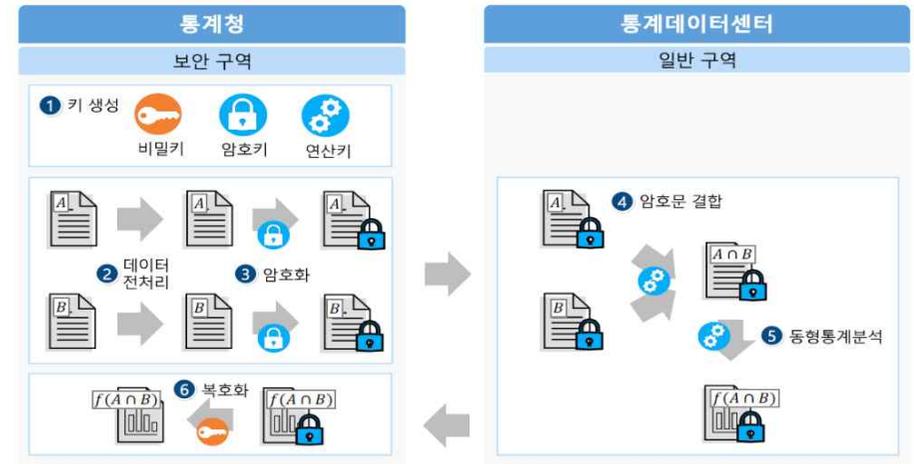
1 관련 법·규정 등 제도 정비

- (통계법 개정) 통계등록부를 통한 통계작성과 통계데이터센터를 통한 적극적인 데이터 공유·개방의 법적 근거 마련
 - 통계등록부 구축, 자료제공 및 이용, 통계데이터센터 구축 및 운영, 통계자료 제공 확대, 정보보호 강화 등
- (통계활성화) 통계의 다양성을 확대하고 빅데이터를 활용한 통계작성 등 새로운 방식의 통계 활성화를 위한 실험통계 제도 확대 운영
 - 또한, 개인정보 등 문제가 없으면 보다 많은 자료를 열람하고 이용할 수 있는 통계자료 '더' 이용제도 운영 내실화
- (데이터보안) 안전한 통계작성 및 데이터 활용을 위하여 '통계작성 및 제공을 위한 비식별화 가이드라인' 제정하고 작성기관에 배포
 - 국제표준(ISO/IEC 20889)에 따라 자료처리자가 비식별화 과정을 적절히 수행할 수 있도록 최신 기법을 활용한 비식별화 평가절차 제안

2 신기술을 통한 K-통계체계 적용

- (K-통계) 성공적인 K-통계시스템 구축을 위한 test-bed 역할 수행
 - 동형암호, 재현자료, 차등정보보호 등 데이터보호 최신기술의 R&D사업 연차별 연구결과를 통계데이터센터에 시범 적용
 - 각 기관이 보유한 데이터가 암호화된 상태로 분리·저장되며 암호화를 유지한 채 데이터센터에서 자료 분석·활용을 지원
 - 데이터센터 운영과 암·복호화 키 관리 업무 분리 등을 통해 국민들의 신뢰를 확보하기 위한 운영주체 검토
- ※ K-통계시스템 연차별 추진계획은 별도 검토하여 보고 예정(3월)

<동형암호 비밀키 관리 및 데이터 흐름도(예시)>



- (신기술 연구) 국가R&D사업 연구성과물(5개)을 과제 수행기관뿐만 아니라 분야별 전문가와 함께 매년 검증, K-통계 적용방안 검토
 - 과기부 국가R&D사업인 ① 동형암호 기술을 활용한 국가통계 분석시스템 연구개발(44억원, '21~'23년), ② 데이터 프라이버시 글로벌 선도기술 연구개발(155억원, '22~'24년)을 성공적으로 추진

3 유관기관과 대외협력 체계 구축

- (타임캡슐) 국가 데이터 정책 추진방향 11대 실천과제중 하나인 『코로나19 타임캡슐 프로젝트』 추진을 위한 주요부처 협의체* 참여
 - * 4차위, 보건복지부, 통계청, 기재부, 행안부 과기부, 질병청 등 주요부처
- (가명정보) 다양한 가명정보의 연계·결합을 통한 각종 시범사례를 성공적으로 추진하기 위하여 『결합전문기관 협의체』 참여
- (데이터플랫폼) 데이터 혁신 촉진에 필요한 정책, 품질, 표준 등을 논의하기 위한 과기부 주관 『민·관 데이터 플랫폼 발전협의체』 참여
 - * 정부(과기부, 통계청, 4차위, 기재부, 행안부 등), 공공/민간플랫폼, 학계, 산업계 등
- (이용자협의체) 타기관 데이터센터 운영자, 통계데이터센터 이용자 및 데이터 전문가 등으로 구성된 협의체를 구성·운영

전략 2 데이터 연계 활성화

1 핵심 데이터 서비스 확대

- (통계등록부) 맞춤형 통계개발을 위하여 각종 자료연계를 통한 다양한 통계등록부를 지속적으로 구축·서비스
 - 소지역 단위 자료 세분화뿐만 아니라 저출산·고령화, 가계부채, 외국인, 부동산, 보육·의료 등 수요가 많은 분야부터 우선 검토
- (공간통계) 승인통계 및 등록부 외에 격자 기반 공간통계서비스 추가
 - 인촌·경춘, 전국사업체 등의 자료를 100m, 500m, 1km 등 격자 크기별 통계값을 제공하고 '반출위원회' 승인후 결과 반출
- (민간데이터) 이용자 수요, 분석 연속성(시계열 업데이트), 대표성 및 세분화 정도 등을 고려하여 활용도 높은 민간자료 구매확충*
 - * 구매예산: 3.5억원('20년) → 5.7억원('21년) → 7.5억원('22년) → 13억원('23년예정)
- (데이터 품질제고) 전문인력을 활용하여 표준화 및 오류 관리 등 노력

2 데이터 연계·결합 전문기관 위상 제고

- (가명정보) 개인정보보호법에 따른 「가명정보 결합전문기관」 이용 활성화를 위해 다양한 사례를 발굴하여 데이터 가치 극대화
- (전문기관) 금융거래, 카드, 부채 등 금융데이터 연계·결합 활성화를 위하여 신용정보법에 따른 「데이터전문기관」 지정 검토 <요건충족>
 - * 정보결합, 가명·익명처리 조사·연구 및 적정성 평가, 표준화 등 업무 처리
 - 데이터 결합 대상 및 범위는 확대되나, 데이터전문기관 업무를 전담으로 하는 결합 담당 별도 인력 및 서버 구비 필요

<결합전문기관 및 데이터전문기관 지정 현황('21.12월)>

결합전문기관 (19개)	(개인정보위) 통계청, 삼성SDS, 지역정보원, 롯데정보통신, KCA, 한국정보인증, 신세계INC
	(과기정통부) NIA, SK C&C, 더존비즈온, 데이터산업진흥원, BC카드
	(보건복지부) 건강보험공단, 심사평가원, 보건산업진흥원, 국립암센터
	(국토부) 도로공사 (교육부) 학술정보원 (산자부) 한진KDN
데이터전문기관 (4개)	(금융위) 국세청, 금융결제원, 금융보안원, 신용정보원

3 데이터 연계·분석 서비스 강화

- (자료연계) 결정매칭*(Deterministic Maching) 및 확률매칭**(Probabilistic Maching) 서비스를 통해 보다 연계율이 높은 분석 서비스 제공
 - * 결정매칭 : 매칭대상변수가 일치할 경우 Match, 그렇지 않으면 Non-Match로 분류하는 방법으로 고유식별자(사업자등록번호 등) 및 준식별자(주소, 이름, 성, 생년월일 등)를 가지고 있는 경우에 활용
 - ** 확률매칭 : 식별자 등의 오류 등을 고려하여 매칭확률을 계산하는 기법
 - 특히, 통계등록부와 조사자료(MD)간 연계를 위하여 준식별자를 활용, MD에 대체식별번호 부여하고 청내 다른 시스템과 번호 통일
- (맞춤형DB) 이용자가 주로 이용하는 행정+조사자료, 행정+민간자료 등 이중 자료간 연계를 통한 맞춤형 DB를 구축·서비스
 - 예를 들면, 산업구조 연구(중사자-기업체DB+전사체), 수출입동향 연구(SBR+관세청 자료), 고용동향 연구(SBR+건강보험자료) 등
- (암데이터) 최근 이용수요가 많은 사망원인 자료의 활성화를 위해 보건복지부, 건보, 암센터 등과 암데이터 연계·결합 모델 정립
 - 보건복지부에서 추진중인 K-CURE 사업*과 연계하여 보안환경이 우수한 데이터센터간 공동 연구구역을 설정, 자료구축 및 관리
 - * Korea-Clinical data Utilization network for Research Excellence(암 임상데이터 네트워크)

<암 임상데이터 네트워크(예시)>

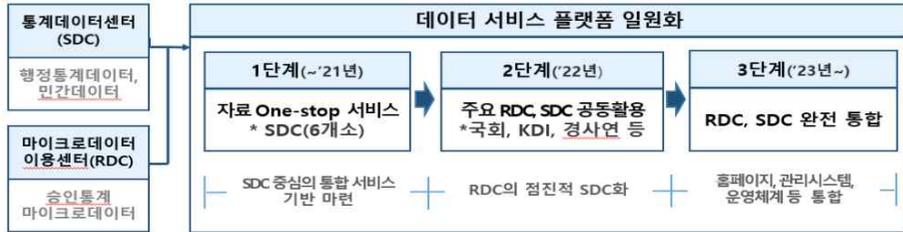


- (플랫폼간 연계) 빅데이터 플랫폼, 지자체 및 공공기관별 센터 등 타기관에서 운영하는 플랫폼과 연계를 통하여 이용자 접근성 제고
 - 우선 데이터센터 홈페이지 링크를 통한 연계를 추진하고, 장기적으로 메타데이터, 연계분석, 오픈API 개발 등 온라인 콘텐츠 추가 개발

전략 3 이용자 중심 서비스

1 데이터 서비스 일원화 및 확대

- (통합서비스) 이용자를 최우선으로 하는 통계서비스 혁신의 일환으로 SDC와 RDC를 단계적 통합하여 one-stop 서비스 추진
 - 단기적으로 센터간 자료호환이 가능하도록 시스템 및 작업공간 개선, 장기적으로 수수료체계 등 관련 규정을 정비하고 통합시스템 개발



- 보안환경, 운영방식(직영), 관련규정, 수수료체계, 이용체계 등에 대해 연구용역, 이용자 및 전문가 회의 등을 통해 **통합방안** 마련
- (서비스 확대·개선) 데이터 개방 요구에 부응하기 위하여 청내 통계등록부 및 MD자료의 **제공범위 확대** 및 **이용체계 개선**
 - **실사과**에서는 공표 즉시 MD자료를 **MD과**로 이관, **MD과**에서는 품질점검 및 관리 등 **정제된 DB**를 SDC로 이관
 - 제공자료는 물리적으로 독립되어 보안이 우수한 SDC에서만 분석하고, 개인정보가 비식별화 처리된 **집계표 형태**로만 반출 가능
 - 분석결과 반출시 “자료반출평가위원회”에 보안·표본, 실사과 등 다양한 전문가가 참여하여 비식별화, 집계결과의 타당성 등 **검토·승인**

<마이크로데이터 이용 절차도(예시)>



* 위원장(통계데이터기획과장), 위원(실사과, 데이터보안, 비밀보호, 표본업무 담당자 등 6명)

2 지역 거점 클러스터 위상 강화

- (지역센터 확충) 데이터 사각지대에 접근성 보장과 데이터 격차 해소를 위한 최소한의 지역별 거점센터 신설 추진
 - 센터가 미설치된 **경기권, 강원권, 제주권** 등을 중심으로 신규 설치하여 지역 삶의 질 제고를 위한 **맞춤형 지역정책 수립** 뒷받침
 - (지방청 기능조정) 지자체, 대학, 연구원 등과 협력하여 기획, 수요 발굴, 지역DB 구축 및 서비스 등 맞춤형 **지방행정 정책지원** 수행
 - **지역통계과 기능개편***을 통해 미래 먹거리 사업을 발굴, 적극적인 데이터 연계·활용을 통한 지역거점 데이터 클러스터 구축·운영
- * ① 지역단위 거버넌스(기획, 회의체, 컨설팅, 수요 등), ② 지역통계(입수, 연계, DB구축, 품질관리, 통계생산 등), ③ 지역센터 운영(채용, 분석지원, 교육·홍보 등)
- 단, 고도의 보안과 전문성이 요구되는 이용승인, 연계·반출 등은 본청 유지

3 포스트 코로나 대비 서비스 다변화

- (주문형 서비스) 이용자가 센터 방문없이 전문가에게 분석 요청, 결과표를 제공받는 주문형 분석서비스 업무를 **지방청으로 이관**
 - 개인, 국가, 공공기관 등 신청한 자료들의 성격상 통계생산에 준하는 것으로 **이용자 신청지역을** 기준으로 해당 지방청에서 작성·제공
- (비대면서비스) 재현자료 추가연구를 통해 **이용자 사전 분석용, 교육용** 등으로 제공하고, **온라인 시각화 및 지도개발** 등 다각화 시도
 - 베이스 및 다중대체 기법(CART), AI 기법(GAN/VAE) 등 재현자료 생성방법론 연구, 결과 검증, 재현자료 생성시스템 개발 등
- (서비스 차등화) 데이터의 **민감도와 재식별 가능성 및 보안 수준** 등을 반영한 차별적인 데이터서비스 제공체계 수립·운영

<서비스별 보안 수준과 제공 데이터 민감도>

데이터	유형	전수데이터	표본데이터	가공데이터
	민감도	높음 ←		→ 낮음
서비스	방식	오프라인		온라인
	채널	통계데이터센터	RDC	RAS 온라인 분석서비스
	보안수준	높음 ←		→ 낮음

전략 4 센터운영 내실화

1 통계데이터 미래조직 설계

- (국가통계 데이터) 통계데이터 개방·이용 활성화 및 국가 디지털 대전환 지원을 위한 예산·인력 등 인프라 확충 필요
 - 국가통계 데이터 활용강화를 위한 전 국민적 공감대 형성 필요

2 통계데이터센터 운영 효율화

- (이용단계별 관리) 자료입수, 품질관리, DB 구축·운영, 자료반출 등 자료이용 전 단계에 대해 운영시스템 기능개선을 통한 효율화
 - 민간데이터 DB 구축·관리, 이용실적 자동집계, 일괄결제 등
- (비식별화 시스템) 반출자료 자동화를 통해 수작업에 따른 오류를 보완하고 반출심의에 따른 업무 부담과 시간 단축 등 효율화
 - K-익명성 기반의 비식별화 로직 등을 재정리하여 자동화 시스템 개선

3 이용자 교육·홍보 및 인재양성 전문화

- (교육·홍보 강화) 전문기관을 통한 교육 콘텐츠 개발, 지역센터 중심으로 잠재적 이용자인 **중고생의 참여형 문제해결** 학습 지원
 - 보유 데이터, 자료간 연계·결합, 통계패키지, 분석지원 서비스 등 통계데이터센터 홍보와 함께 **흥미 위주의 현장학습** 병행
- (사례 발굴) 분석사례, 연구결과(기술노트) 등 분석방법 벤치마킹 기회를 제공하기 위하여 '**기술노트 모음집**'을 이용자에게 공유
 - 모범사례를 매년 선정, '**우수사례집**'을 발간하고 도서관 및 기관에 배포
- (인재양성) 데이터 연계·결합, 품질관리, 이용자 컨설팅 등 다양한 경험을 통해 **데이터 노하우 축적** 및 내부전문가 **인재양성** 기회 제공
 - 통계교육원 '**데이터 사이언티스트 양성과정**'과 연계, **데이터 전문가** 위주로 지역센터 우선 배치(상주)하고 컨설팅을 전담

V. 기대 효과

- 증거기반 정책지원을 위한 국가데이터 **연계 활성화** 기틀 마련
 - 다수 기관이 보유한 **행정자료**와 **통계등록부** 및 **MD 자료**, **민간자료** 등을 연계·결합하여 다양한 통계생산을 통한 정부 정책수립 지원
 - 포스트 코로나를 대비하여 통계데이터센터 방문 없이 분석이 가능한 **비대면 서비스**를 통해 이용자 접근성 및 편의제고
 - 데이터 활용 활성화와 개인정보보호 간 균형을 통해 **데이터 개방 확대** 요구에 대응하고 **사회적 가치 창출** 기대
- 민간기업 지원 **활성화** 및 우리 경제의 **경쟁과 혁신** 촉진
 - 데이터를 보유한 플랫폼 기업과 데이터 보유·접근이 어려운 신생 기업의 **데이터 격차**를 해소함으로써 경쟁과 혁신의 생태계 조성
 - ODA 사업과 연계하여 **신남방, 신북방** 국가 등을 대상으로 데이터 활용 **모범사례**를 전수하여 국위선양 및 국가 경쟁력 확보
 - K-통계와 같은 데이터 프라이버시가 확보된 플랫폼의 선도적인 구축은 국내 **디지털 뉴딜 사업 활성화**의 마중물 기대

VI. 향후 일정

- 국가통계위원회 안전 보고 : '22. 2월 중
- 진행사항 점검(허브국 주요업무 점검에 포함) : '22. 3월 이후

참고 1 통계데이터센터 운영 현황

- (센터개요) 수요자가 공공데이터를 편리하게 이용하고 민간자료 등 외부자료 융합이 가능하도록 구축된 데이터 플랫폼
 - (대상) 행정기관, MOU기관, 통계작성기관, 연구·분석을 목적으로 하는 연구기관 및 기업, 개인 등 일반이용자 등
 - (기능) 행정통계자료, 인가된 통계기초자료 및 민간자료 제공, 데이터간 연계지원, 분석전문가 상담 및 분석지원 서비스, 가명정보 결합* 등
 - * 중앙행정기관 제1호 결합전문기관 지정('20.11.25., 개보위)



외부와 차단된 공간에서 데이터를 자유롭게 결합·분석할 수 있는 데이터 안심존

- 추진경과
 - (법적·제도적) 통계데이터센터 운영규정(제정, '16.9.30.)
 - 통계빅데이터센터 운영 및 이용 등에 관한 규정(전부개정, '18.11.20.)
 - 통계데이터센터 운영 및 이용 등에 관한 규정(일부개정, '20.2.11.)
 - (물리적 공간) '16.10 시범운영, '18.11.부터 시스템 및 분석공간 구축
 - Zero-Client 단말기(자료저장불가)를 통해 가상화서버에 접속하여 자료분석, 가상화서버는 국가정보자원관리원에 설치
 - 분석공간 물리적 안전통제, 데이터 반·출입 통제, 데이터 접근 권한 최소화 부여 등 최상의 보안환경으로 운영



SDC (Statistics Data Center) 로고는 이성적 분석의 방향성과 통찰의 방향성을 함께 공유하여 더 나은 미래를 제시한다는 데이터센터의 비전을 표현

○ 센터운영

- 시범운영('16.10.~'18.10.) : 파주센터(시범운영종료), 대전센터
- 본격운영('18.11.20.) : 대전, 서울, 부산에 센터 설치
- 지역센터 확충 : 광주센터('20.3.), 대구센터 ('21.3.), 서울상암센터('21.9.)
- 지역센터 추가 설립예정 : 경기, 강원, 제주 등 검토중

구분	대전	서울(중구)	서울(상암)	부산	광주	대구
개소 일자	'18.11.20.	'18.11.20.	'21.9.27.	'18.7.3. ('20.5.확장)	'20.3.23.	'21.3.29.
위치	통계센터 13층	한국데이터산업진흥원	에스플렉스센터 15층	통합청사 1층	통합청사 4층	SW융합테크비즈센터 2층
센터 규모	540㎡ (163평)	56㎡ (16.9평)	14.4㎡ (4.4평)	73.57㎡ (22.3.평)	66㎡ (20.0평)	117.88㎡ (35.7평)
이용자 좌석	34	10	5	12	12	15
좌석당 면적	15.9	5.6	2.9	6.1	5.5	7.9
운영인력 (명)	13 (분석12,운영1)	2 (분석1,운영1)	1 (분석1)	2 (분석1,운영1)	2 (분석1,운영1)	2 (분석1,운영1)

○ 센터별 이용실적

(단위: 건, %)

구분	2018년	2019년	2020년	2021년	비중
					(100.0)
합계	78	290	369	554	(100.0)
대전	50	86	141	178	(32.1)
서울(중구)	20	176	169	276	(49.8)
서울(상암)	-	-	-	15	(2.7)
부산	8	13	26	11	(4.2)
광주	-	-	25	35	(7.4)
대구	-	-	-	19	(3.8)

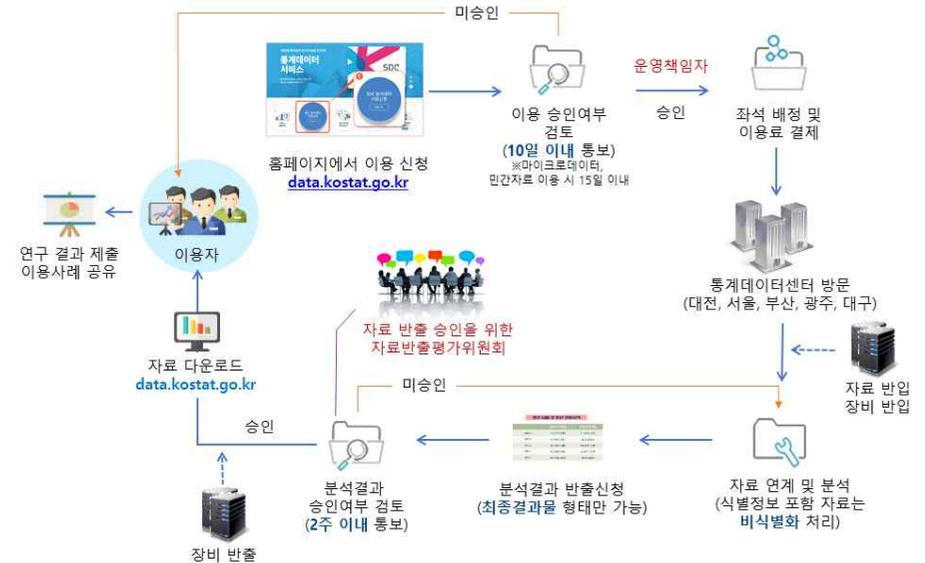
○ 센터별 역할

- (대전센터) 이용자분석지원, 데이터 구축(품질관리, 기초분석), 주문형 분석, 명부제공, 자료연계, 비식별처리 및 점검, 온라인시각화서비스, 반입자료관리, 시스템관리, 사용자 교육지원 등 총괄적인 서비스 지원업무
 - * 운영시스템 등은 가상화서버로 연계되어 방문이용자 지원 및 체험 교육을 제외하고는 대전센터에서 전담하여 지원
- (지역센터) 데이터 품질관리, 지역센터 이용자분석지원, 사용자 체험교육지원

□ 제공 서비스 현황

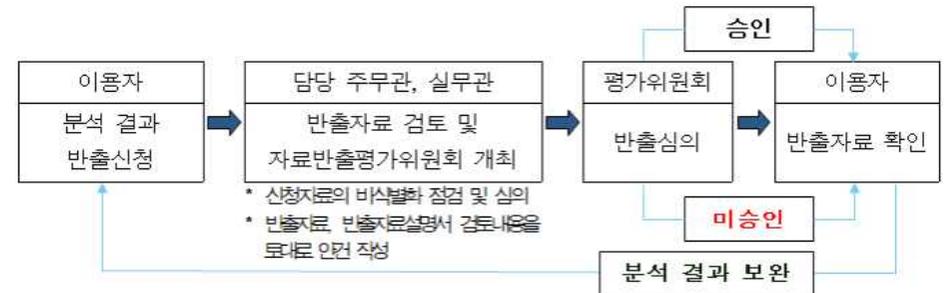
- **(데이터 구축)** 행정통계DB(16종), 민간자료(35종) 및 맞춤형 데이터셋 등 다양한 서비스 데이터 구축 및 제공
- **(전문가 분석 지원)** 자료 분석 경험이 부족한 이용자를 위한 자료이용 상담 및 연계·분석 자문 등 분석 컨설팅 서비스
 - * 분석전문가(17명) : DB구축, 자료연계, 비식별화 처리, 이용자 분석지원 수행
- **(이용자 주문형 분석서비스)** 센터 방문 또는 직접 자료 분석이 어려운 이용자를 위하여 온라인 상담을 통해 분석결과를 통계표 형태로 제공
- **(이용자 교육 지원)** 일반국민의 통계자료 이용 촉진* 및 데이터 리터러시** 실현을 위해 맞춤형 교육지원 및 분석사례 공유
 - * 온라인 교육시스템 운영(센터이용방법, 데이터체험, 데이터활용방법 분야)
 - * 맞춤형 교육지원 : 통계기초 및 데이터의 이해 교육, 대학생(대학원, 고등학생 포함)을 대상으로 연구활동, 논문작성 지원을 위해 이용료 감면
 - ** 데이터 리터러시(Data Literacy) : 데이터를 읽고 분석하여 활용할 수 있는 능력
- **(센터 이용 지원)** 승인받은 이용자가 센터에서 제공하는 데이터 및 이용자가 반입하는 자료 등을 직접 분석할 수 있도록 분석플랫폼 제공
 - 분석패키지(R, SAS, SPSS, STATA) 및 분석 플랫폼(SGIS-PRO, QGIS, Qlik-Sense, R-서버-PRO) 제공
 - 온라인(시각화 통계분석자료 및 사용자분석기능) 및 자료연계 (Felligi&Sunter자료 연계 모형 기반 시스템) 서비스
- **(결합전문기관 운영)** 통계청이 결합전문기관으로 지정*됨에 따라 결합 신청자(기관)의 다(多)기관·다(多)데이터 결합을 위한 전문적 역할 수행
 - 물리적 별도의 공간(결합지원실)에서 결합, 결합신청자의 접근성 용이한 지역의 센터 내 자료분석실에서 추가 가명·익명화 처리(별도 좌석 마련)
 - 이종자료간 데이터 연계를 통한 데이터 활용도 제고와 새로운 부가가치 창출을 위해 가명정보 및 결합관련 여건마련

□ 센터 이용절차



□ 반출관리

○ 반출절차



○ 반출평가위원회 운영

- 심의대상: 센터내에서 이용한 모든 반출요청건
- 반출위원: 행정자료, 자료보안, 비밀보호, 표본, 통계생산 담당자
- 의결요건: 재적위원 과반수 출석과 출석위원 2/3이상 찬성으로 반출

참고 2 통계데이터센터 제공자료(51종)

※ ○ 서비스시작, ● 예산구입(업데이트 포함)

1. 통계등록부(16종)

부분	자료명	범위	항목	시계열 (21.4.22.기준)	주기	갱신시점	'18	'19	'20	'21
							10	12	14	16
							10	12	14	16
경제·사회	기업등록부DB(사업자등록기준)	전국	61	2010~2019년	년	+1년 (3-4월중 이관)	○	○	○	○
	기업등록부DB(대표자기준)	전국	61	2010~2019년	년	+1년 (3-4월중 이관)	○	○	○	○
	기업등록부DB_분기(사업자등록기준)	전국	40	2018년 4분기~ 2020년 4분기	분기	+5개월		○	○	○
	기업등록부DB_분기(대표자기준)	전국	40	2018년 4분기~ 2020년 4분기	분기	+5개월		○	○	○
	사업장 기초DB(4대보험)	전국	20	2009~2019년	년	+1년 (3-4월중 이관)	○	○	○	○
	사업장 기초DB(법인)	전국	19	2008~2019년	년	+1년 (3-4월중 이관)	○	○	○	○
	종사자-기업체 연계DB	전국	18	2015~2019년	년	+1년 (3-4월중 이관)			○	○
	육아휴직 사용현황DB	전국	23	2015~2019년	년	+1년 (3-4월중 이관)			○	○
	등록센서스 인구DB	전국	12	2015~2019년	년	+1년 (3-4월중 이관)	○	○	○	○
	등록센서스 가구DB	전국	13	2015~2019년	년	+1년 (3-4월중 이관)	○	○	○	○
인구·가구	인구동태 코호트DB(예정)	전국								○
	아동가구 통계DB	전국								○
	등록센서스 주택DB	전국	10	2015~2019년	년	+1년 (3-4월중 이관)	○	○	○	○
	농업DB	전국	8	2015~2019년	년	+1년 (3-4월중 이관)	○	○	○	○
민요민애	임업DB	전국	8	2015~2019년	년	+1년 (3-4월중 이관)	○	○	○	○
	어업DB	전국	8	2015~2019년	년	+1년 (3-4월중 이관)	○	○	○	○

2. 민간데이터(35종)

부분	자료명	원천	범위	시계열 (20.10.1.기준)	주기	갱신 (구입)	유형	'18	'19	'20	'21
								22	26	33	35
인구·가구	성·연령별 유동인구	SKT	전국	2015.01.~2020.12.	월	'20.10. (12월 추가)	민간	○	○	●	●
	시간대별 유동인구	SKT	전국	2015.01.~2020.12.	월	'20.10. (12월 추가)	민간	○	○	●	●
	요일별 유동인구	SKT	전국	2015.01.~2020.12.	월	'20.10. (12월 추가)	민간	○	○	●	●
	성·연령별 유입인구	SKT	부산	2017.01.~2018.06.	월	'18.11.	민간	○	○	○	○
	성·연령별 주거인구	SKT 행안부	전국	2016~2019	년	'20.10.	공공·민간	○	○	●	●
	성·연령별 직장인구	SKT 국토부	전국	2016~2019	년	'20.10.	공공·민간	○	○	●	●
	가구소득	SKT 통계청	전국	2016~2017	년	'18.11.	공공·민간	○	○	○	○
	(코로나19)방문인구정보	SKT	전국	2019.01.~2020.12.	월	'20.10.	민간			●	●
	(코로나19)방문인구정보 (유입지null)	SKT	전국	2019.01.~2020.12.	월	'20.10.	민간			●	●
	경제·사회	성·연령별 카드매출1	현대카드	전국	2015.01.~2018.06.	월	'18.11.	민간	○	○	○
업종별 매출액		현대카드	전국	2015.01.~2018.06.	월	'18.11.	민간	○	○	○	○
성·연령별 카드매출2 (국내, 21년분리)		BC카드	전국	2018.10.~2020.12.	월	'20.10. (12월 추가)	민간		●	●	●
성·연령별 카드매출2 (해외, 21년 분리)		BC카드	전국	2018.10.~2020.12.	월	'20.10. (12월 추가)	민간				●
신용통계정보(대출)		KCB	전국	2018.10.~2020.12.	월	'20.10. (12월 추가)	민간		●	●	●
신용통계정보(소득)		KCB	전국	2018.10.~2020.12.	월	'20.10. (12월 추가)	민간		●	●	●
신용통계정보(카드소비)		KCB	전국	2018.10.~2020.12.	월	'20.10. (12월 추가)	민간		●	●	●
인구이동정보(전입인구)		KCB	전국	2019.4/4.~2020.4/4	분기	'20.10. (12월 추가)	민간			●	●
인구이동정보(전출인구)		KCB	전국	2019.4/4.~2020.4/4	분기	'20.10. (12월 추가)	민간			●	●
인구이동정보(통근인구)		KCB	전국	2019.4/4.~2020.4/4	분기	'20.10. (12월 추가)	민간			●	●
건물·지가	구매상품정보(오프라인)	롯데 멤버스	전국	2019.12.~2020.11.	월	'20.12.	민간			●	●
	1000대 상권 정보	SKT 부동산114	전국	2018	년	'18.11.	민간	○	○	○	○
	영업 중 업소	SKT 중기부	부산	2017.7.~2018.6	월	'18.11.	공공·민간	○	○	○	○
	표준공시지가(서울, 부산)	SKT 국토부	서울	2016~2017	년	'18.11.	공공·민간	○	○	○	○
	주택(단독/다세대/기타)	SKT 국토부	서울	2016~2017	년	'18.11.	공공·민간	○	○	○	○
	공동주택정보	SKT 부동산114	서울	2016~2017	년	'18.11.	공공·민간	○	○	○	○
	건축물정보	SKT 국토부	전국	2017.7.~2018.6	월	'18.11.	공공·민간	○	○	○	○
	인구집중유발시설	SKT	서울	2016.4.~2018.7 (3개월 간격)	분기	'18.11.	민간	○	○	○	○
	도로별 차량통행량	SKT 국토부	서울	2017	년	'18.11.	공공·민간	○	○	○	○
	시·군·구별 차량등록대수	SKT 국토부	전국	2017	년	'18.11.	공공·민간	○	○	○	○
기초정보	행정동코드 정보	SKT	전국	2018	년	'18.11.	공공·민간	○	○	○	○
	법정동코드 정보	SKT	전국	2018	년	'18.11.	공공·민간	○	○	○	○
	소지역정보(I, II)			~2018, 2019~							○
기타	SNS데이터	SKT	전국	2017	년	'18.11.	민간	○	○	○	○
	온라인가격 정보	온라인 수집	전국	2017.01.~2021.03	월	익월	민간	○	○	○	○
	모바일상품권 거래 정보	스마트콘	전국	2018	년	'20.04.	민간			○	○

< 의안 소관 부서명 >

통계청 통계데이터기획과	
연 락 처	(042) 481 - 1299