『데이터산업현황조사』

통계정보보고서

2024. 09.

본 이용자용 통계정보보고서는 정기통계품질진단 수행과정에서 통계작성기관이 작성한 보고서로 작성기준 시점에 따라 현재의 통계작성 정보와 다소 차이가 날 수 있습니다.

작성일자: 2025.07.31.



과학기술정보통신부

<차 례>

Ⅰ. 통계작성 기획 ····· 1

Ⅱ. 통계설계 ····· 9

Ⅲ. 자료수집 ····· 30

Ⅳ. 통계처리 및 분석 ····· 40

Ⅴ. 통계공표, 관리 및 이용자서비스 ····· 49

Ⅵ. 통계기반 및 개선 ····· 60

◆ 보고서 개요 ◆

이 보고서는 「데이터산업현황조사」 통계를 생산하기 위하여 과학기술정보통신부에서 수행하는 업무를 설명한 것이다. 보고서의 작성목적은 조사의 배경, 연혁, 이용자 및 용도와 통계에서 이용되는 개념과 방법론에 대하여 심층적으로 알고자 하는 통계작성 담당자(통계 전문이용자, 품질진단자 또는 승인담당자)에게 통계과정 전반에 대하여 포괄적이고 상세한 정보를 제공하는 것이다. 여기에는 통계작성 기획, 통계설계, 자료수집, 통계처리 및 분석, 통계공표, 관리 및 이용자서비스, 통계기반 및 개선 등에 대한 설명이 수록되어 있다.

Ⅰ. 통계작성 기획

□ 통계개요

○ 데이터산업현황조사(승인번호: 제127004호)

○ 통계작성기관/부서명

- 과학기술정보통신부/데이터진흥과

1. 조사방법

○ 전문조사원이 조사대상 기업을 대상으로 온라인, 이메일, 전화, 방문 등을 병행하여 실시

2. 조사 및 공표주기

○ 조사주기: 1년

○ 공표주기: 1년

3. 조사일정 및 일정별 수행 업무

○ 일반사항

- 본 조사는 조사기획 및 준비, 조사설계(조사표 설계 및 구성, 표본설계 등 포함), 조사원 교육, 실사준비, 조사실시, 자료처리 및 공표 순으로 진행
- 조사기획 단계부터 조사결과 공표까지 약 1년 정도의 기간이 소요

○ 일정별 수행 업무

단계	일 정	세부 추진내용
조사기획	6~7월	추진 방향 설정 및 일정 확인
	7~8월	조사업체 입찰 진행 및 계약
조사준비	7~8월	현황 검토 및 진행방향 설정
	9~10월	자문단 구성 및 회의 진행
조사설계	8~10월	조사 항목 선정 및 조사표 설계
	8~10월	모집단 정비 및 표본 설계
	11월	데이터산업 조사 변경승인
조사원교육	9~10월	조사지침서 및 가이드 작성
	11월	조사원 교육 수행
실사 준비	10~11월	조사 협조 요청 공문 작성
	10~11월	조사지침 가이드 및 조사표 작성 요령 확인
	11월	종이 조사표 디자인 및 인쇄
조사 실시	11~12월	조사실시 및 실사 관리
	실사기간	조사현황 보고 확인
	11~12월	조사내용 검토 및 확인(회수 에디팅, 조사품질 제고 등)
	12월	추가 조사실시(이상치, 무응답, 거절 등 대체표본조사 등)
	12월	조사현황 보고 실시
자료처리 및 공표	12월	데이터 클리닝(데이터 오류·이상치 식별, 수정판단 및 처리방법 결정 등)
	12월	가중치 적용
	익년 1월	조사결과 산출(결과표 작성 등)
	익년 5월	결과 공표(보도자료, 보고서 등)

4. 통계작성 문서화

4-1 통계작성 기본계획서

○ 본 조사는 과학기술정보통신부의 ‘데이터 기업역량 강화’ 사업의 일환으로 실시되고 있으며, 매년 동 사업의 수행계획, 세부 추진계획 등을 수립하여 단계적으로 업무 추진

4-2 업무편람(직무편람)

○ 본 조사의 개요, 수행절차(조사업체 선정, 모집단 정비, 표본 설계 등), 조사표 작성 요령 등으로 구성된 담당자 업무매뉴얼을 기준으로 관련 업무 수행 중

5. 통계연혁

5-1 작성통계의 최초개발 시기 및 배경

□ 최초개발 시기

○ 본 조사는 1995년에 “데이터베이스 산업 현황 및 이용실태조사”로 최초 작성되었으며, 2015년 데이터산업 현황조사로 조사명 변경 후 2016년 11월 16일 통계작성승인 획득

□ 개발 배경

○ 최초에는 데이터베이스 산업의 규모, 구조 및 이용 실태 등을 종합적으로 조사하여 국내 데이터베이스 산업 육성 및 활성화 정책 수립 등을 위한 기초자료를 제공하기 위해 개발

○ 4차 산업혁명 시대에 데이터가 인공지능, 사물인터넷 등의 주요 기술을 작동하는 핵심 자산으로 부각됨에 따라, 데이터베이스 산업을 포괄할 수 있는 데이터 산업으로 범위를 확장하여 데이터 관련 정책·사업화 전략 수립, 연구 등을 지원을 위한 통계자료 생산중

5-2 통계의 변경 또는 개편이력 관리(최근 진단년도 이후부터)

□ 주요 변경 또는 개편 이력

연도	변경 내용	통계작성기관
2020년	데이터산업 중분류 조정	과학기술정보통신부
2021년	문항 변경	
2022년	일반산업 분야 조사 신규 승인	
2023년	일반산업 분야 일부 업종 모집단 정의 변경	
2024년	문항 변경	

○ 2020년 통계 변경 세부 내역

- (명부 최신화) 통계 시의성 확보를 위해 최신 모집단 명부 반영
- (산업분류 조정) 대분류 4개, 중분류 13개 분류체계에서 대분류 4개, 중분류 12개 분류체계로 개편

< 2020년 통계작성 변경승인 내역 >

구분	변경전(2019년)	변경후(2020년)
통계작성 대상	데이터기업 모집단 8,300여개	데이터기업 모집단 8,600여개
조사시점	당해연도 하반기	
표본 추출틀	데이터기업: 2019년 ICT통합모집단	데이터기업: 2020년 ICT통합모집단
표본규모	데이터기업 표본 1,500개	데이터기업 표본 1,500개
자료수집 방법	웹, 이메일, 전화, 방문 등 병행조사	

< 2020년 데이터산업 분류체계 변동 현황(일부업종) >

변경전(2019년)		변경후(2020년)	
대분류	**중분류**	**대분류**	**중분류**
데이터 판매 및 제공 서비스업	데이터 판매 및 중개 서비스업	데이터 판매 및 제공 서비스업	데이터 판매 및 중개 서비스업
	데이터 분석 제공 서비스업		
	정보제공 서비스업		정보제공 서비스업
데이터 인프라 서비스업	온프레미스	데이터 인프라 서비스업	데이터 저장장치 및 시설 서비스업
	클라우드		데이터 네트워크 서비스업

○ 2021년 통계 변경 세부 내역

- (명부 최신화) 통계 시의성 확보를 위해 최신 모집단 명부 반영
- (문항 변경) 빅데이터 관련 데이터직무 인력 현황 문항 삭제, 비대면화 관련 사업 현황 문항 신설

< 2021년 통계작성 변경승인 내역 >

구분	변경전(2020년)	변경후(2022년)
통계작성 대상	데이터기업 모집단 8,600여개	데이터기업 모집단 9,300여개
조사시점	당해연도 하반기	
표본 추출틀	데이터기업: 2020년 ICT통합모집단	데이터기업: 2021년 ICT통합모집단
표본규모	데이터기업 표본 1,500개	데이터기업 표본 1,500개
자료수집 방법	웹, 이메일, 전화, 방문 등 병행조사	

○ 2022년 통계 변경 세부 내역

- (명부 최신화) 통계 시의성 확보를 위해 최신 모집단 명부 반영
- (시범조사 변경승인) 기존 시범조사로 진행한 일반산업 13개* 분야 종사자 수 100인 이상사업체의 조사내용 변경승인(일부업종 상이)

* 공공, 금융, 제조, 유통, 서비스, 의료, 통신·미디어, 물류, 교육, 유틸리티, 농축산·광업, 건설업, 숙박·음식점업

< 2022년 통계작성 변경승인 내역 >

구분	변경전(2021년)	변경후(2022년)
통계작성 대상	데이터기업 모집단 9,300여개	데이터기업 모집단 9,500여개 일반기업 모집단 13,000여개
조사시점	당해연도 하반기	
표본 추출틀	데이터기업: 2021년 ICT통합모집단	데이터기업: 2022년 ICT통합모집단 일반기업: 2020년 경제총조사
표본규모	데이터기업 표본 1,500개	데이터기업 표본 1,500개 일반기업 표본 1,200개
자료수집 방법	웹, 이메일, 전화, 방문 등 병행조사	

○ 2023년 통계 변경 세부 내역

- (명부 최신화) 통계 시의성 확보를 위해 최신 모집단 명부 반영
- (문항 변경) 비대면화 관련 사업 현황 문항 삭제
- (모집단 정의 조정) 일반산업 일부 업종(공공, 교육, 의료) 모집단 정의 조정
- (직무설명 현행화) 데이터 처리 절차(8단계) '기획>설계>운영/유지보수>수집>가공>분석>활용>거래/관리' 기준에 맞춰 데이터직무 순서 조정 및 설명 현행화

< 2023년 통계작성 변경승인 내역 >

구분	변경전(2022년)	변경후(2023년)
통계작성 대상	데이터기업 모집단 9,500여개 일반기업 모집단 13,000여개	데이터기업 모집단 10,300여개 일반기업 모집단 18,000여개
조사시점	당해연도 하반기	
표본 추출틀	데이터기업: 2022년 ICT통합모집단 일반기업: 2020년 경제총조사	데이터기업: 2023년 ICT통합모집단 일반기업: 2021년 전국사업체조사
표본규모	데이터기업 표본 1,500개 일반기업 표본 1,200개	
자료수집 방법	웹, 이메일, 전화, 방문 등 병행조사	

< 2023년 일반산업 모집단 정의 변경 내역 >

구분	변경전(2022년)	변경후(2023년)
공공	중앙정부기관, 지자체, 기획재정부의 공공기관 경영정보 시스템 (ALIO)에 등록된 준정부기관, 공기업, 기타공공 기관, 부설기관과 지방자치단체 포함	공공 행정, 국방 및 사회보장 행정에 해당되는 사업체 중 100명 이상 사업체
의료	의료기관 중 50병상 이상 병원	표준산업분류 중 보건업 및 사회복지 서비스업으로 등록된 사업체 중 100명 이상 사업체
교육	교육 서비스업 지정 교육 기관 중 대학교육협의회에 등록된 대학교	표준산업분류 중 교육 서비스업으로 등록된 사업체 중 100명 이상 사업체

○ 2024년 통계 변경 세부 내역

- (명부 최신화) 통계 시의성 확보를 위해 최신 모집단 명부 반영
- (문항 변경) 現 데이터산업에서 보편적으로 인지·사용 중인 용어로 자구수정 및 유사문항 병합

< 2024년 통계작성 변경승인 내역 >

구분	변경전(2023년)	변경후(2024년)
통계작성 대상	데이터기업 모집단 10,300여개 일반기업 모집단 18,000여개	데이터기업 모집단 10,600여개 일반기업 모집단 19,000여개
조사시점	당해연도 하반기	
표본 추출틀	데이터기업: 2023년 ICT통합모집단 일반기업: 2021년 전국사업체조사	데이터기업: 2024년 ICT통합모집단 일반기업: 2022년 전국사업체조사
표본규모	데이터기업 표본 1,500개 일반기업 표본 1,200개	
자료수집 방법	웹, 이메일, 전화, 방문 등 병행조사	

6. 통계의 작성목적

6-1 통계 작성 목적

○ 데이터 거래 및 활용기반 산업을 촉진하고 관련법에 따른 시책과 계획을 효율적으로 수립·추진하기 위해 매년 데이터산업 기초 통계조사 실시

- 데이터 정책·사업화 전략 수립을 위한 기초자료 확보를 위해 데이터산업의 시장규모, 데이터직무 인력 현황 등 데이터산업 전반에 대한 실태 파악을 통한 통계자료 생산

6-2 주된 통계활용 분야

○ (정책) 정부기관의 국내 데이터산업 육성 및 활성화를 위한 정책·전략 수립

- 데이터산업 관련 기술·정책 환경 변화에 대응하기 위한 데이터 생산·유통·거래·활용 등 데이터산업 전반에 걸친 정책·전략 수립의 핵심 기초자료로 활용

○ (학술) 대학·연구기관의 데이터산업 발전방향 모색 등을 위한 다각적인 측면의 연구 수행

- 데이터산업의 미래 방향성을 제시하거나 비즈니스 모델 발굴 및 정책 제언 등을 위해 국내 데이터 시장, 인력, 정책 등 데이터산업 현황을 분석하기 위한 연구자료로 활용

○ (언론) 데이터산업 관련 최신 동향과 정책 정보 등을 제공하기 위한 보도자료 작성

- 데이터산업 전반에 대한 대중의 이해도 제고와 사회적 공감대 형성을 위한 보도자료 작성 등에 필요한 객관적이고 신뢰성 있는 근거자료로 활용

6-3 국내외 관련 통계 또는 유사 사례에 대한 사전 검토 사항

□ 관련 통계 간 사전 검토 내용

○ 국내에서는 본 조사가 국내 데이터산업 전반의 현황과 수요 등을 파악할 수 있는 유일한 조사이므로, 인공지능·클라우드 등 유관 ICT 분야 산업 실태조사 등의 조사 항목 등을 참고

< '24년 데이터산업 유관 ICT 분야 조사 비교 >

구 분	데이터산업 현황조사	인공지능산업 실태조사	클라우드산업 실태조사
조사목적	데이터산업 육성 지원 정책 수립을 위한 기초통계 생산	인공지능 산업 생태계 조성 및 활성화 정책 수립을 위한 기초자료 확보	클라우드산업 활성화 정책수립 지원을 위한 기초 조사 자료를 확보
작성기관	과학기술정보통신부		
위탁기관	한국데이터산업진흥원	소프트웨어정책연구소	한국클라우드산업협회
승인연도	2016년 11월	2019년 11월	2018년 7월
조사주기	1년	1년	1년
조사대상	데이터산업 영위 기업, 일반산업 13개 분야 기업	인공지능산업 영위 기업	클라우드 서비스 공급 기업
조사항목	데이터 관련 매출 현황, 데이터직무 인력 현황, 데이터 보유·유통 현황, IT투자 현황	인공지능 매출현황, 인공지능 인력현황, 인공지능 투자현황, 비즈니스 애로사항 등	클라우드기업의 일반 현황 및 매출 현황, 수출 및 해외 진출 현황, 산업 및 인력 현황, 정책 건의 등

7. 유형별 주요 이용자 관리

□ 주요 이용자

○ (정책) 과학기술정통신부(통계작성기관)를 포함한 중앙정부·지자체 등 정부기관

○ (연구) 정보통신정책연구원(KISDI), 한국고용정보원(KEIS)) 등 연구기관

○ (기타) IT 미디어, 데이터 관련 학회, 민간기업 및 연구소 등

8. 이용자 의견수렴

8-1 이용자 의견수렴 실시 내용과 주요 결과

□ 이용자 의견수렴 실시 내용

○ 2024년 제1차 통계기획자문위원회 개최

- 일시/장소: 2024.9.11.(수), 14:00~18:00/한국데이터산업진흥원 회의실
- 참석자: 통계·데이터 등 분야별 전문가, 조사전문기관 등 12명
- 회의주제: 2024년 데이터산업 현황조사 내 산업·직무 분류체계 점검 및 자문
- 주요내용: 데이터산업, 데이터직무 등의 정의, 설명, 업계 현황 등의 적정성

검토 및 현행화, 산업 트렌드에 따른 신규 산업 및 직무 분류 도출 등에 관한 의견 수렴 등

○ 2024년 제2차 통계기획자문위원회 개최

- 일시/장소: 2024.9.25.(수), 14:00~17:00/한국데이터산업진흥원 회의실
- 참석자: 통계·데이터 등 분야별 전문가, 조사전문기관 등 12명
- 회의주제: 2024년 데이터산업 현황조사 조사문항 고도화
- 주요내용: 데이터산업 매출 현황, 데이터직무 인력 현황 등 기존 조사문항 검토 및 정비, 층별 표본설계 방식의 적절성 검토, 기타 조사문항 고도화 관련 개선 의견수렴

Ⅱ. 통계설계

1. 조사항목

1-1 조사항목의 체계 및 조사목적

□ 조사항목의 체계

○ 데이터산업: 데이터 매출 현황, 데이터직무 인력 현황 등 7개 조사영역 46개 문항

○ 일반산업: 데이터 도입 현황, 데이터직무 인력 현황 등 4개 조사영역 27개 문항

< '24년 데이터산업 현황조사 조사항목 체계 >

구 분		조 사 내 용
공통	기업 일반 현황	• 주사업 및 부사업(한국표준산업분류업종) • 전체 종사자 수 • 전체 매출액 • 설립연도
	의견수렴	• 데이터 사업 수행 시 애로사항(데이터산업) • 데이터 사업 수행 시 필요정보 • 데이터 산업 활성화 정책 수요 • 데이터 전문인력 양성 정책 수요
	데이터직무 인력 현황	• 데이터직무별 인력 수(전체/성별/기술등급별) • 현재/향후5년간 데이터직무 필요인력 수(전체/기술등급별) • 데이터직무 채용인력 수(최근2년, 익년) • 채용 우대 기술, 우대 자격증 • 채용 애로사항
데이터산업	데이터 매출 현황	• 데이터 매출액 • 데이터산업 분류별 매출 비중 • 매출 유형별 비중, 수요처 업종별 매출 비중
	데이터 보유·유통 현황	• 데이터 보유 현황(보유여부, 수집방법, 데이터형태) • 데이터 유통 현황(거래여부, 거래목적) • 데이터 거래 애로사항
	IT 투자 현황	• IT 예산 및 부문별 규모 • 전체 매출액 대비 IT 예산 비중
	해외 진출 현황	• 해외진출 경험 유무, 활동 유형, 진출 방식 • 데이터 관련 해외매출 발생여부 및 규모(국가별) • 해외진출 애로사항
일반산업	데이터 도입 현황	• 국내 기업의 데이터 관련 구축·활용 여부 • 데이터 활용 유형/분야별 비중 • 도입에 대한 관심 수준, 미도입 이유, 향후 도입 시기

□ 주요항목의 조사목적

○ 데이터 시장·인력·기업 등 데이터산업 변화와 정책 수요 등을 종합적으로 파악할 수 있는 통계자료를 생산하여 데이터 관련 정책·사업화 전략 수립을 위한 기초자료로 활용

1-2 주요 용어 및 항목별 정의

○ 데이터산업 정의

- 경제적 부가가치를 창출하기 위하여 데이터의 생산·유통·거래·활용 등 일련의 과정과 관련된 행위와 이와 관련되는 서비스를 제공하는 산업
- 데이터의 생성, 저장, 분석, 가공, 활용 등 제반 활동(생명주기 또는 가치사슬)을 포함해 데이터로부터 가치가 창출되는 일련의 모든 과정, 그리고 이와 연관된 활동을 포함

○ 데이터산업 범위

- 사업 유형별로는 ①데이터 처리 및 관리 솔루션 개발·공급업(데이터 수집·연계 솔루션 개발·공급업, 데이터베이스 관리 시스템 솔루션 개발·공급업, 데이터 분석 솔루션 개발·공급업, 데이터 관리 솔루션 개발·공급업, 데이터 보안 솔루션 개발·공급업, 데이터 통합 플랫폼 솔루션 개발·공급업), ②데이터 구축 및 컨설팅 서비스업(데이터 구축·가공 서비스업, 데이터 관련 컨설팅 서비스업), ③데이터 판매 및 제공 서비스업(데이터 판매·중개 서비스업, 정보제공 서비스업), ④데이터 인프라 서비스업(데이터 저장 장치·시설 서비스업, 데이터 네트워크 인프라 서비스업)을 제공하는 사업체 포함

< 데이터산업 비즈니스 정의 >

구분	비즈니스 정의 및 매출 발생 구조
데이터 처리 및 관리 솔루션 개발·공급업	• DBMS, DBMS관리, 데이터 모델링, 분석·시각화, 검색엔진, 품질 등 관련 솔루션 제품을 판매하는 등의 비즈니스를 의미하며, 주로 라이선스, 개발·커스터마이징, 유지보수를 통해 매출이 발생
데이터 구축 및 컨설팅 서비스업	• DB설계, 데이터 이행 등을 포함한 DB시스템 구축, 문서·음성·영상 등의 데이터를 DB로 변환, 정비, 가공, DW, Data Lake 구축 등의 데이터·DB 구축·가공하거나 데이터 관련 컨설팅 비즈니스를 의미(데이터·DB 관련 SI·IT아웃소싱 포함). 데이터 구축 및 컨설팅은 구축·개발, 유지보수·운영관리, 컨설팅을 통해 매출이 발생(DB·데이터 구축 과정에서의 컨설팅 매출은 구축 부문에 포함, SI 매출 중 HW·SW 공급 비용을 제외한 용역 매출만 포함)
데이터 판매 및 제공 서비스업	• 데이터·DB를 원천데이터 형태나 분석 및 활용이 가능한 상태로 판매하거나 중개, 데이터신디케이션 및 큐레이션 등을 통한 수요 맞춤형 데이터·DB 판매, 데이터를 가공·활용·분석한 주제분야별 정보서비스, 분석 결과 정보 등을 온·오프라인(모바일, 앱 등 포함)으로 제공하는 비즈니스를 의미함 • 데이터 이용료·수수료 등의 직접매출과 광고료 등의 간접매출로 수익이 발생

데이터인프라 서비스업 (시범영역)	• 데이터 기반 솔루션 구축과 서비스를 위해 반드시 필요한 서버, 스토리지, 네트워크 장비 등을 바탕으로 데이터 인프라 서비스를 제공하는 비즈니스를 의미 • 데이터 클라우드, 서버, 네트워크 인프라 이용료 등의 매출이 발생

※ 데이터인프라는 용어설명을 위해 수록하였으며, 현재 시범조사 영역으로 별도 조사 형태로 진행하였음

○ 데이터직무: 데이터 컨설턴트, 데이터 기획자, 데이터 아키텍트, 데이터 개발자, 데이터 엔지니어, 데이터베이스관리자, 데이터 분석가, 데이터 기획자 총 8개 직무로 구분

< 데이터직무 설명 >

데이터 직무명	설명
데이터 컨설턴트 (Data Consultant)	• 기업 또는 조직의 데이터 전략 및 프로세스 개선을 위한 조언 및 지원 업무 • 고객사의 데이터 기반 비즈니스 문제를 해결 및 경영에 필요한 해결책(솔루션)을 제시하는 업무 • 데이터 분석을 토대로 기업이 앞으로 나아갈 방향, 해결책 등을 제시하는 업무
데이터 기획자 (Data Planner)	• 기업이나 조직이 데이터를 효율적으로 활용하여 비즈니스 목표를 달성 및 필요한 전략을 이행하는 업무 • 기업의 비즈니스 목표에 따른 데이터 관련 프로젝트를 정의하고 계획하는 업무 • 정제된 DB 및 비정제 영상・이미지・음성, 데이터 관련 제품·서비스 기획 및 판매 • 데이터 활용·분석 등을 위한 데이터 큐레이팅・코디네이팅 등 데이터 수집 관련 기획 및 이에 필요한 거버넌스(정책, 법률, 규제 등) 수립 업무
데이터 아키텍트 (DA, Data Architect)	• 전사적 관점의 데이터 기반 IT 정책, 표준화, 구조, 설계 및 이행 • 전사 데이터 품질 관리 정책 및 거버넌스 체계 마련 • 개념적, 논리적, 물리적 데이터 설계 및 관리 체계 수립 예시) 데이터 아키텍트(DA), 데이터 모델러, 데이터 웨어하우스 아키텍트
데이터 개발자 (Data Developer)	• 데이터(DB)를 활용하여 서비스에 필요한 데이터 수집/정제/저장을 위한 프로세스 개발 • 데이터 활용 및 서비스 제공을 위한 API 개발 • QA(Quality Assurance) 및 오픈 라이브러리 활용, 개발 • 데이터 처리를 통한 응용 솔루션 개발 • 데이터 연계프로그램(ETL, EAI 등)을 활용한 솔루션 개발
데이터 엔지니어 (Data Engineer)	• 데이터 성능, 서비스, 제품에 대한 기술지원, 유지보수 • 머신러닝 모델 학습 및 배포, 도메인 지식 기반 데이터 전처리·추출 업무 • 대용량 처리가 가능한 데이터 파이프라인 및 플랫폼 설계 및 구축 • 데이터를 저장·처리·관리하거나 프로그램 개발 • 로그, 크롤링 자동화 등 데이터 수집 환경 구축 • 기계 학습용 데이터 생성·활용을 위한 참조구현 구축 • 데이터 구분·선별, 데이터 결합 및 포맷 변형 예시) 데이터 엔지니어, 데이터 플랫폼 및 시스템 엔지니어, 머신러닝 엔지니어, 피처 엔지니어 직무 포함
데이터베이스관리자 (DBA, Database Administrator)	• 요구사항 기반으로 데이터 관리 체계를 검토·개선·관리하는 업무 • 데이터(DB) 구성, 변경, 용량, 성능, 가용성(백업, 복구), 보안, 장애, 문제관리 등 운영시스템 관리 업무 • 외부 기관과 데이터를 송수신하는 체계를 수립하고, 물리적 보안 및 정보 보호 기능을 운영하는 업무 • 비즈니스 요구사항 기반으로 데이터베이스를 설계 및 스키마 구축 업무
데이터 분석가 (Data Analyst)	• 정형・비정형 데이터를 식별・관리・조작・분석하여 기업 경영의 의사결정에 활용할 수 있도록 자료를 만들어내는 직무

	• 통계, 머신러닝, 인공지능 기반 데이터 분석, 분석결과 시각화 업무 • 다양한 분석기법을 활용하여 데이터가 보여주는 현상을 해석하고 그 원인을 분석해 정보를 얻는 업무 • 핵심 비즈니스를 파악하고 개선하기 위한 실험 설계, 데이터를 분석하여 결과를 도출하는 업무 • 마케팅조사, 사회여론조사 등 조사 데이터를 분석하여 정보를 얻고 인사이트를 도출하는 업무 • 서로 다른 이종데이터 매쉬업과 가명정보 결합을 통해 데이터를 분석하는 업무
데이터 과학자 (Data Scientist)	• 조직 내외부 데이터의 관리·활용·분석 체계를 새롭게 만들고, 프로세스 혁신 및 신제품 개발, 마케팅 전략 결정 등의 의사결정을 이끌어내는 직무 • 데이터 간 관계, 패턴, 규칙 등을 찾아내 모형화하고 이에 적용하는 알고리즘을 개발하는 업무 • AI·통계 모델 설계 및 실무 적용, AI·통계 최신 기술 연구와 구현 및 적용 업무 • 기업의 비효율적인 업무 및 문제해결을 위한 분석 과제 발굴 및 시나리오 설계를 통한 예측 모델 개발

○ 데이터직무 기술등급: 경력·학력·자격 등에 따라 초급, 중급, 고급 3개 등급으로 구분

< 데이터직무 기술등급 설명 >

기술등급	기준
초급	• 정보처리기사, 정보처리산업기사 등 동급 데이터 관련 자격 취득자
	• 전문학사 이상의 학위 보유자
	• 고등학교 졸업 후 3년 이상 경력자
중급	• 기사자격 또는 데이터 관련 자격 취득 후 3년 이상 경력자 (데이터 관련 자격 : 데이터아키텍처준전문가, SQL개발자, OCP, 데이터분석준전문가, MCDBA, CSP, MySQL, DB Administrator, DB2 for Linux and Windows)
	• 산업기사 자격 취득 후 7년 이상 경력자
	• 석사 학위 취득 후 기사자격 또는 데이터 관련 자격 취득 후 2년 이상 경력자
고급	• 정보관리 기술사, 컴퓨터시스템응용 기술사
	• 중급 이후 3년 이상 경력자
	• 데이터 고급 자격 취득자 * 데이터 고급 자격: 데이터아키텍처전문가, SQL전문가, OCM, 데이터 분석 전문가
	• 박사 학위를 가진 자로서 기사 및 데이터 관련 자격 취득자

○ 국내 데이터 전문인력 자격증: 데이터 분석, SQL 등 국내 데이터 관련 자격증 9종으로 구분

< 국내 데이터 관련 전문인력 자격증 설명 >

자격증	내용
데이터아키텍처 전문가(DAP)	• 효과적인 데이터아키텍처 구축을 위해 전사아키텍처와 데이터품질관리에 대한 지식을 바탕으로 데이터 요건분석, 데이터 표준화, 데이터 모델링, 데이터베이스 설계와 이용 등의 직무를 수행하는 전문가자격검정
데이터아키텍처 준전문가(DAsP)	• 효과적인 데이터아키텍처 구축을 위해 전사 아키텍처에 대한 지식을 바탕으로 데이터 요건 분석, 데이터 표준화, 데이터 모델링 등의 직무를 수행하는 실무자자격검정

SQL 전문가 (SQLP)	• 데이터를 조작하고 추출하는데 있어서 정확하고 최적의 성능을 발휘하는 SQL을 작성할 수 있고, 이를 토대로 SQL을 내포하는 데이터베이스 프로그램이나 응용 소프트웨어의 성능을 최적화하거나, 이러한 성능 최적화를 지원할 수 있는 데이터베이스 개체(뷰, 인덱스 등)의 설계와 구현 등의 직무를 수행하는 전문가 자격검정
SQL 개발자 (SQLD)	• 데이터베이스와 데이터 모델링에 대한 지식을 바탕으로 응용 소프트웨어를 개발하면서 데이터를 조작하고 추출하는데 있어서 정확하고 최적의 성능을 발휘하는 SQL을 작성할 수 있는 개발자자격검정
데이터 분석 전문가(ADP)	• 데이터 이해 및 처리 기술에 대한 기본지식을 바탕으로 데이터 분석 기획, 데이터 분석, 데이터 시각화 업무를 수행하고 이를 통해 프로세스 혁신 및 마케팅 전략 결정 등의 과학적 의사결정을 지원할 수 있는 전문가자격검정
데이터 분석 준전문가(ADsP)	• 데이터 이해에 대한 기본지식을 바탕으로 데이터 분석 기획 및 데이터 분석 등의 직무를 수행할 수 있는 실무자자격검정
정보처리기사	• 정보시스템의 생명주기 전반에 걸친 프로젝트 업무를 수행하는 직무로서 계획수립, 분석, 설계, 구현, 시험, 운영, 유지보수 등의 업무를 수행할 수 있는 전문가자격검정
사회조사분석사	• 다양한 사회정보의 수집·분석·활용을 담당하는 새로운 직종으로 기업, 정당, 지방자치단체, 중앙정부 등 각종 단체의 시장조사 및 여론조사 등에 대한 계획을 수립하고 조사를 수행하며 그 결과를 분석, 보고서를 작성하는 전문가 자격검정
빅데이터 분석기사	• 대용량의 데이터 집합으로부터 유용한 정보를 찾고 결과를 예측하기 위해 목적에 따라 분석기술과 방법론을 기반으로 정형/비정형 대용량 데이터를 구축, 탐색, 분석하고 시각화 직무를 수행하는 전문가 자격검정

○ IT 예산 항목: Server, Storage, Network, SW/Solution, Service 등 5개 항목으로 구분

< IT 예산 항목 설명 >

항목	정의
Server	• 기업 내 컴퓨팅 프로세스를 처리하기 위한 하드웨어 서버 시스템 [H/W]
Storage	• 기업 내 주요 전산 데이터 및 파일을 저장하는 하드웨어 시스템 [H/W]
Network	• 기업 내 외부의 네트워크 프로세스를 처리하기 위한 네트워크 장비 시스템[H/W]
SW/Solution	• 소프트웨어 패키지 또는 응용프로그램을 처리하는 소프트웨어 시스템 [S/W]
Service	• 외부 용역형 IT시스템 개발, 구축, 유지보수 및 클라우드 등 서비스형 제품을 모두 포함

○ 해외진출 방식: 수출에 의한 진출, 계약에 의한 진출 등 3개 대분류와 10개 중분류로 구분

< 해외진출 방식 설명 >

대분류	중분류	개념
수출에 의한 진출	간접수출	• 무역상, 수출대행업자, 수출조합을 통해 자사 제품을 해외에 판매하는 형태 • 상품의 선적, 보험, 세관통관, 해외 판매 등 해외시장 진출에 관련된 모든 기능과 책임은 수출 대행업자가 짐
	직접수출	• 판매 대리인, 유통업자, 판매 지사, 판매 법인 등 현지에 자체 판매 조직을 보유하거나 현지 유통 업체를 통해 직접 상품을 유통시키는 방식
계약에 의한	라이센스	• 목표 시장국 내의 기업에게 기술이나 무형자산, 인력자원 등을 이전해 주는 계약관계를 통해 시장에 진입하는 방법

진출	프랜차이즈	• 제공기업이 상호, 상표, 기술 등의 사용권을 특정 기업이나 개인에게 허용할 뿐 아니라 상대방에 대해 조직, 마케팅 및 일반관리 분야에서의 지원을 제공하는 등 양자가 지속적이며 포괄적인 관계를 유지
	기타 계약 방식	• 주로 현지기업과의 해당 계약에 의해 해외사업 운영 • 계약생산(contract manufacturing), 관리생산(management contracts), 턴키계약(turnkey contracts), BOT방식(build, operate, transfer) 등이 대표적
직접 투자	단독투자	• 모기업이 현지 투자대상 기업의 의결권주 95% 이상을 소유하는 형태로 해외에 진출하는 경우
	합작투자	• 2개 이상의 기업, 개인 또는 정부기관이 영구적인 기반 아래 특정 기업체의 운영에 공동으로 참여하는 경우
	자회사 신설	• 과거에 존재하지 않았던 기업을 새롭게 설립하여 해외시장에 진출하는 경우
	인수합병	• 투자대상국에서 가동되고 있는 기업의 주식이나 자산 등을 매입하여 경영권을 확보하는 것으로, 결합형태에 따라 신설합병 또는 흡수합병으로 분류
	전략적제휴	• 2개 이상의 기업이 공동의 전략적 목표(기술개발, 생산, 판매, 자본 조달 등)를 달성하기 위해 협력하는 것. M&A와 달리 양 법인의 독립성은 유지

1-3 주요 용어의 정의나 개념 등에 대한 국내 또는 국제기준 비교

○ (국내) 「데이터 산업진흥 및 이용촉진에 관한 기본법」상 데이터 산업의 정의

< 데이터산업 정의 비교 >

구분	데이터 산업진흥 및 이용촉진에 관한 기본법 제2조제5호	데이터산업 현황조사
데이터산업 정의	"데이터산업"이란 **경제적 부가가치를 창출하기 위하여 데이터의 생산・유통・거래・활용 등 일련의 과정과 관련된 행위와 이와 관련되는 서비스를 제공하는 산업**을 말한다.	데이터산업은 '**경제적 부가가치를 창출하기 위하여 데이터의 생산·유통·거래·활용 등 일련의 과정과 관련된 행위와 이와 관련되는 서비스를 제공하는 산업**'으로 정의한다

○ (해외) IDC「European Data Market Study」보고서의 데이터기업 및 직무 등 용어 정의

< 데이터기업, 데이터직무 등 용어 정의 비교 >

구분	데이터산업현황조사	European Data Market Study
데이터기업 정의	경제적 부가가치를 창출하기 위하여 데이터의 생산·유통·거래·활용 등 일련의 과정과 관련된 행위와 이와 관련되는 서비스를 제공하는 1인이상 사업체	디지털 제품, 서비스, 기술의 형태로 데이터의 생산, 유통, 활용 등과 직접적으로 관련있는 기업으로 데이터 공급자와 데이터 사업자로 분류
데이터직무 인력	데이터아키텍트, 데이터개발자,	데이터 수집, 저장, 관리 및 분석,

정의 및 범위	데이터엔지니어, 데이터분석가, 데이터베이스관리자, 데이터과학자, 데이터컨설턴트, 데이터기획자 업무를 주업무로 하는 인력	해석, 시각화 작업을 주요 활동 또는 활동의 일부로 수행하는 인력으로 데이터 기술 전문가, 데이터 비즈니스 전문가, 데이터 소비자로 구분

2. 적용 분류체계

2-1 통계에서 사용하고 있는 분류체계 개요 및 내용

□ 데이터산업 분류체계

○ 한국표준산업분류와 비교

- 데이터산업은 한국표준산업분류상 J58221, J58222, J62021, J62090, J63111, J63120, J63910, J63991, J63999 업종에 주로 포함됨

○ 데이터산업 분류 내용

- 대분류 4개, 중분류 12개 분류(인프라 시범조사 영역 대분류 1개, 중분류 2개 포함)

< '24년 데이터산업 분류체계 >

한국표준산업분류	2024년 데이터산업 분류		
	대분류	중분류	내용
(J58221)시스템 소프트웨어 개발 및 공급업 (J58222)응용소프트웨어 개발 및 공급업	1. 데이터 처리 및 관리 솔루션 개발·공급업 **[데이터 솔루션]**	11. 데이터 수집·연계 솔루션 개발·공급업 **[데이터 수집]**	데이터 검색 솔루션, 데이터 수집 솔루션, 데이터 통합·연계 솔루션, 데이터 교환·개방 솔루션
		12. 데이터베이스관리 시스템 솔루션 개발·공급업 **[DBMS]**	관계형 데이터베이스 관리 솔루션, 비관계형 데이터베이스 관리 솔루션, 인메모리 데이터베이스 관리 솔루션, 기타 데이터베이스 솔루션
		13. 데이터 분석 솔루션 개발·공급업 **[데이터 분석]**	정형·비정형 데이터 분석 솔루션, 실시간 데이터 분석 솔루션, 데이터 시각화 분석 솔루션, 데이터 처리 솔루션
		14. 데이터 관리 솔루션 개발·공급업 **[데이터 관리]**	데이터 모델 관리 솔루션, 데이터 표준 관리 솔루션, 데이터 품질 관리 솔루션, 데이터베이스 운영·성능 관리 솔루션
		15. 데이터 보안 솔루션 개발·공급업 **[데이터 보안]**	데이터 보안 솔루션, 개인정보보호 솔루션
		16. 데이터 통합 플랫폼 솔루션 개발·공급업 **[데이터 플랫폼]**	데이터 플랫폼 솔루션
(J62021)컴퓨터시스템 통합 자문 및 구축 서비스업 (J62090)기타 정보기술 및 컴퓨터운영 관련 서비스업 (J63111)자료 처리업	2. 데이터 구축 및 컨설팅 서비스업 **[데이터 구축/컨설팅]**	21. 데이터 구축·가공 서비스업 **[데이터구축]**	데이터베이스 설계·구축 서비스, 데이터 이행 서비스, 데이터 가공 서비스
		22. 데이터 관련 컨설팅 서비스업 **[데이터컨설팅]**	데이터 설계 컨설팅, 데이터 품질 컨설팅, 데이터베이스 성능개선 컨설팅 , 데이터 거버넌스 컨설팅, 데이터 분석·활용 컨설팅
(J63120)포털 및 기타 인터넷 정보 매개서비스업	3. 데이터 판매 및 제공 서비스업 **[데이터 서비스]**	31. 데이터 판매 및 중개 서비스업 **[데이터 거래]**	데이터 판매 서비스, 데이터 중개 서비스, 분석 데이터 제공 서비스

(J63910)뉴스 제공업 (J63991)데이터베이스 및 온라인정보제공업 (J63999)그 외 기타 정보서비스업		32. 정보제공 서비스업 **[정보제공]**	포털·정보 매개서비스, 기타 정보 제공 서비스
(J61210)유선통신업 (J61220)무선 및 위성통신업 (J63112)호스팅 및 관련 서비스업	4. 데이터 인프라 서비스업 **[데이터 인프라]**	41. 데이터 저장 장치·시설 서비스업 **[데이터 저장장치]**	데이터 저장 및 인프라 제공·관리 서비스
		42. 데이터 네트워크 인프라 서비스업 **[데이터 네트워크]**	데이터 유·무선 네트워크 등

2-2 국내 또는 국제기준의 표준분류체계 사용 여부 또는 미사용 사유

○「데이터 산업진흥 및 이용촉진에 관한 기본법」제2조제5호 등에서는 데이터 가치사슬*을 중심으로 데이터산업을 정의하고 있어, 한국표준산업분류만으로 구분하기에는 한계가 존재

* 경제적 부가가치를 창출하기 위한 데이터의 생산 · 유통 · 거래 · 활용 등 일련의 과정 등

○ 이에, 데이터산업을 정의하여 한국표준산업분류와 매핑하여 활용하고 있으며, 급변하는 데이터산업 환경 변화를 파악하기 위해 매년 전문가 자문회의 등을 통해 업데이트 진행

< 데이터산업분류 vs 한국표준산업분류 >

데이터산업 분류		**한국표준산업분류**	
대분류	**중분류**		
1. 데이터 처리 및 관리 솔루션 개발·공급업	11. 데이터 수집·연계 솔루션 개발·공급업	J58221 J58222	시스템 소프트웨어 개발 및 공급업 응용소프트웨어 개발 및 공급업
	12. 데이터베이스관리시스템 솔루션 개발·공급업		
	13. 데이터 분석 솔루션 개발·공급업		
	14. 데이터 관리 솔루션 개발·공급업		
	15. 데이터 보안 솔루션 개발·공급업		
	16. 데이터 통합 플랫폼 솔루션 개발·공급업		
2. 데이터구축 및 컨설팅 서비스업	21. 데이터 구축·가공 서비스업	J62021 J62090 J63111	컴퓨터시스템 통합 자문 및 구축 서비스업 기타 정보기술 및 컴퓨터운영 관련 서비스업 자료 처리업
	22. 데이터 관련 컨설팅 서비스업		
3. 데이터 판매 및 제공 서비스업	31. 데이터 판매 및 중개 서비스업	J63120 J63910 J63991 J63999	포털 및 기타 인터넷 정보 매개 서비스업 뉴스 제공업 데이터베이스 및 온라인정보제공업 그 외 기타 정보서비스업
	32. 정보제공서비스업		

3. 조사표 설계 및 구성

3-1 조사표 설계 및 변경 절차나 방법

□ 조사표 설계 및 변경절차 방법

○ 기초 조사표 설계

- 조사결과의 시계열 분석을 위해 과거년도 조사 항목을 검토
- 과거 조사항목 결과 자료를 확인하고 계속 조사여부를 검토

○ 조사항목 개선(안) 계획 수립

- 산업변화를 반영하기 위해 관련기관 및 전문가의 다양한 의견을 수렴하여 일부 항목과 보기의 수정, 보완, 삭제 등 실시

○ 조사항목 확정

- 내부 검토 의견 정리 및 관련부처 협의: 의견수렴, 전문가회의 등 절차를 거쳐 최종 항목을 선정한 후 관련부처와 최종 협의를 통해 조사표 확정

3-2 조사표 구성 관련 내·외부 전문가 회의

□ 조사표 구성 및 변경의 기본 방향

○ 통계·데이터 등 분야별 전문가 7인 내외로 구성한 통계기획자문위원회를 통해 데이터산업 업계의 현황과 트렌드를 반영하여 파트별 기존 조사문항 정비 및 신규문항 추가 검토

○ 국어문화원 자문을 통해 조사표 내 지시문, 용어, 구성 등을 일원화

3-3 조사표 구성 관련 내·외부 전문가 회의 결과 반영 여부

□ 문항 관련 주요 요구사항 및 반영 결과

○ 조사표 내 지시문, 용어, 구성 등 일원화, 데이터산업 트렌드에 맞게 용어 변경

< '24년 통계기획자문위원회 검토 의견을 반영한 조사문항 변경 사례(일부) >

23년 조사(AS-IS)	24년 조사(TO-BE)	수정사유
PART A 빅데이터 관련 도입·활용 현황	PART A 데이터 관련 도입·활용 현황	
문1. 빅데이터 관련 인프라/솔루션/서비스에 대한 구축·활용 여부	문1. 데이터 관련 인프라/솔루션/서비스에 대한 구축·활용 여부	**'빅데이터' 용어**는 현업계에서 **사용하지 않는 추세**를 반영
문2. (도입·추진중인 기업만) 귀사의 빅데이터 구축 및 활용 유형	문2. (도입·추진중인 기업만) 귀사의 데이터 구축 및 활용 유형	
문3. (미도입 기업만) 빅데이터 도입에 대한 관심 수준	문3. (미도입 기업만) 데이터 도입에 대한 관심 수준	* 통계기획자문위원회 의견
문4. (미도입 기업만) 향후 빅데이터 도입 및 활용 시기	문4. (미도입 기업만) 향후 데이터 도입 및 활용 시기	

3-4 첨부된 조사표에 수록된 사항

□ 조사표 수록 사항

○ 조사명, 조사목적, 국가승인통계로고, 작성승인번호, 응답자 협조사항, 조사협조 감사인사, 조사기관 명시, 응답자 비밀보호정책, 문의사항 연락처

3-5 조사방법이 다양한 경우 조사방법별 특성을 고려한 조사표 설계 여부 및 검토 요소

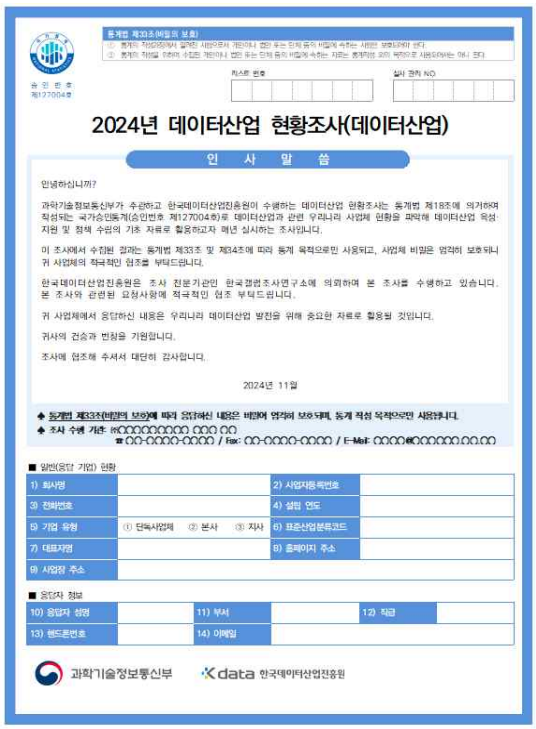
□ 전화, 웹(온라인), 이메일, 팩스, 방문 등 병행조사

○ 조사표는 면접조사를 종이 조사표로 했을 때를 기반으로 조사원과 응답자가 문항 이동 및 로직 파악이 손쉽게 이루어질 수 있도록 구성하였음

○ CAPI 시스템을 구축하였으며, CAPI 시스템은 웹(온라인) 조사, 방문 조사에서도 동일하게 활용할 수 있게 하였음

○ 사업체 조사는 조사항목에 따라 여러명의 응답자에게 응답을 받아야할 수도 있고, 파트별로 순서를 다르게 받게 될 수도 있음. CAPI 시스템에서도 파트별로 응답할 수 있고, 응답자가 다르다면 응답자를 다르게 기입할 수 있게 구현하였음

< 조사표 실물 >



2024년 데이터산업 현황조사(데이터산업)

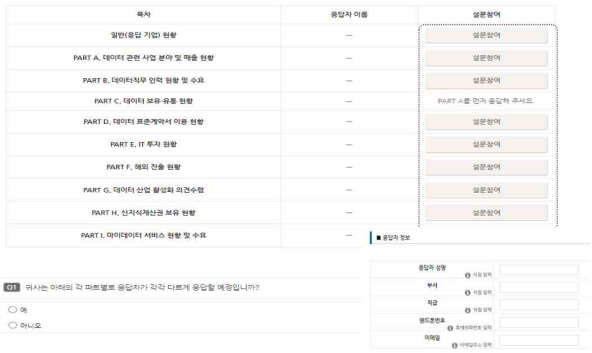
인 사 말 씀

인사말씀

과학기술정보통신부

한국데이터산업진흥원

< CAPI 시스템 >



3-6 응답유형별 응답 소요시간

□ 2024년 응답 소요시간

○ 조사 평균 응답시간: 약 20~30분

- 다만, 조사항목(매출액, 데이터사업 매출실적, 직무별 인력 수 등)에 따라 내부에서 자료를 취합해야 하기 때문에 조사 대상에 따라 소요시간 차이는 발생할 수 있음

4. 조사표 변경 이력 관리

4-1 & 4-2 & 4-3 조사표 변경 이력 및 변경이유, 변경승인일자 기록 · 관리

□ 조사항목 변경 관련 국가데이터처 변경승인 이력

○ 2016년 최초 승인

○ 2018년 변경승인(변경일: 2018년 11월 19일)

- 조사표 2종 → 1종 단일화: 산업분류별 일부문항을 제외하고 동일한 문항이므로 조사표 간소화
- 데이터산업분류 조정에 따른 시장구분 조정: (기존) 대분류 3개, 중분류 12개 → (변경) 대분류 4개, 중분류 13개
- 조사항목 추가(11개 항목, 국내 매출과 해외 매출 비중, 채용인력 현황, 해외진출 관련 문항 등): 데이터매출 명확화, 인력양성 정책 활용, 타 유관 산업통계와의 비교 등을 위한 항목 추가
- 조사항목 삭제(3개, 업종별 매출 비중, 자격증별 우대사항, 기술 관련 자산 및 데이터베이스의 매출 기여 비중): 응답률이 낮고 활용성이 떨어지는 항목 삭제
- 조사항목 수정(10개, 사업분야 선택단위, IT 예산 및 부문별 비중, 데이터 보유 현황, 그 외 일체 조사항목의 선택지 조정 등): 응답 간소화, 현재 시장 환경 반영

○ 2020년 변경승인(변경일: 2020년 11월 11일)

- 데이터산업분류 조정에 따른 시장구분 조정: (기존) 대분류 4개, 중분류 13개 → (변경) 대분류 4개, 중분류 12개
- 조사항목 추가(2개, 데이터 직무 필요기술, 일반업종별 매출 비중): 데이터 전문인력 통계 생산, 데이터 수요 현황 반영 등을 위한 항목 추가

○ 2021년 변경승인(변경일: 2021년 9월 16일)

- 조사항목 추가(7개, 비대면화 관련 사업 현황): 비대면 기술 공급 및 수요 파악을 위한 공통문항 신규 추가
- 조사항목 삭제(2개, 빅데이터 관련 인력, 빅데이터 관련 필요 인력): 데이터 직무 인력 정의에 빅데이터 직무 인력 정의 포함하여 조사 간소화 및 응답 정확도 제고
- 조사항목 수정(2개, 데이터 직무 채용 인력, 채용 시 자격증): 데이터 직무 정의에 빅데이터 직무 정의 포함, 데이터 관련 신설 자격증 선택지 반영

○ 2022년 변경승인(변경일: 2022년 10월 16일)

- 조사항목 추가(21개, 일반기업의 빅데이터 관련 도입-활용 현황, 데이터직무 인력 현황 및 수요, 데이터산업 활성화 의견 수렴 등): 기존 시범조사로 실시한 일반기업 대상 조사문항 승인 획득

- 조사항목 수정(9개, 분야별 매출 구성, 데이터 보유 여부, 데이터 수집 출처, 해외진출 애로사항 등): 세부 수요처 파악을 위해 분야 및 선택지 다양화

○ 2023년 변경승인(변경일: 2023년 10월 23일)
- 일반산업 '교육' 분야 모집단 정의 변경(대학교육협의회에 등록된 대학교 → 표준산업분류 교육서비스업으로 등록된 종사자 100인 이상 사업체): 분야별 일반산업 모집단 정의 일원화
- 조사항목 삭제(7개, 비대면화 현상으로 인한 변화): 코로나19 종식에 따른 관련 항목 삭제
- 조사항목 수정(1개, 데이터 거래 시 겪는 애로사항): 선택지 일부 조정

○ 2024년 변경승인(변경일: 2024년 11월 25일)
- 조사항목 추가(21개, 일반기업의 데이터 관련 도입-활용 현황, 데이터직무 인력 현황 및 수요, 데이터산업 활성화 의견 수렴 등): 기존 시범조사로 실시한 일반기업 대상 조사 문항 승인 획득
- 조사항목 삭제(1개, 데이터 관련 국내·해외 매출 비중): 해외 매출 범위 조정에 따라 발생한 중복항목 삭제
- 조사항목 수정(6개, 데이터산업 관련 해외매출, 데이터 관련 도입·활용 현황): 응답 범위 축소(해외 매출 범위(전체→데이터산업), 용어 정의(빅데이터→데이터)) 조정

5. 목표모집단과 조사모집단

5-1 목표모집단의 정의

○ **(데이터산업)** 경제적 부가가치를 창출하기 위하여 데이터의 생산·유통·거래·활용 등 일련의 과정과 관련된 행위와 이와 관련되는 서비스를 제공하는 1인 이상 사업체

○ **(일반산업)** 13개 업종*을 영위하는 종사자 수 100인 이상 사업체(단, 농축산·광업 50인 이상)

* 공공, 금융, 제조, 유통, 서비스, 의료, 통신·미디어, 교육, 물류, 유틸리티, 농축산·광업, 건설, 숙박·음식점업

5-2 조사모집단의 정의

○ **(데이터산업)** 2024년 ICT통합모집단을 기준으로 데이터산업을 영위하는 사업체 10,691개

○ **(일반산업)** 2022년 전국사업체조사 명부를 기준으로 13개 업종을 영위하는 종사자 수 100인 이상 사업체 19,816개

6. 조사모집단(전수조사) 또는 표본추출틀(표본조사)

6-1 조사모집단 또는 표본추출틀로 사용되는 자료의 출처(통계명, 작성기관, 작성연도)

□ **표본추출틀**

○ 통계명 : ICT통합모집단 명부

○ 작성기관 : 과학기술정보통신부

○ 작성연도 : 2024년 기준

○ 활용근거 : 과학기술정보통신부 정보통신·방송통계 관리규정

6-2 조사모집단 또는 표본추출틀로 선정한 이유

□ **표본추출틀 선정 이유**

○ 과학기술정보통신부 정보통신·방송통신 관리규정('18.1.1 시행) 제 5조에 의거 매년「ICT 통계조사 시행계획」수립에 따라 과학기술정보통신부 관련 ICT 통계는「ICT통합모집단」을 활용함

6-3 조사모집단 또는 표본추출틀의 구축(갱신) 주기, 방법 및 절차, 결과

□ **표본추출틀 구축 주기**

○ 매년 1년

- 본 조사모집단과 ICT통합모집단의 갱신주기는 1년으로 일치하나, 조사시점과 시간적 한계 존재

□ **표본추출틀 구축 방법**

○ 국가데이터처 전국사업체 조사 명부자료 + 과기정통부의 개별통계 명부자료 연계

- 전국 사업체조사 명부 내 데이터산업 특수분류코드(58221, 58222, 62021, 62090, 63111, 63120, 63910, 63991, 63999 등)와 ICT 통합모집단 명부를 바탕으로 표본추출틀을 구축하고 이를 매년 조사 시 업데이트하여 관리 및 사용

□ **표본추출틀 절차, 결과**

○ 전년도 조사대상 모집단 정비 및 산업의 최신 상황 반영하여 표본 추출틀 매년 개편

- (모집단 정비) 신용기관 DB 등을 통해 휴폐업 상태, 경영상태, 업종, 주소 등 확인하여 기존기업 생멸 확인 후 최종 정비
- (ICT통합분류체계 정비) 한국표준산업분류(KSIC) 연계, 누락·중복된 산업분류, 품목분류 추가·부분 통합 등 정비

6-4 분류별, 지역별 기타 하위모집단별 추출단위 분포, 관련 통계량, 상관관계 등 기록 및 관리

○ 결과보고서에 데이터산업 분류별·규모별 추출단위 기록 관리

< '24년 데이터산업 현황조사 모집단 분석내용(일부 발췌) >

[2024년 데이터산업 현황조사 모집단 구성]

구분		종사자 규모					
대분류	중분류	1-9인	10-49인	50-99인	100-299인	300인 이상	합계(개)
데이터 처리 및 관리 솔루션 개발·공급업	데이터 수집·연계 솔루션 개발·공급업	235	196	63	79	8	581
	데이터베이스 관리 시스템 솔루션 개발·공급업	87	96	12	3	4	202
	데이터 분석 솔루션 개발·공급업	207	225	39	21	3	495
	데이터 관리 솔루션 개발·공급업	625	535	97	70	15	1,342
	데이터 보안 솔루션 개발·공급업	214	153	46	19	7	439
	데이터 통합 플랫폼 솔루션 개발·공급업	169	139	17	8	1	334
소계		**1,537**	**1,344**	**274**	**200**	**38**	**3,393**
데이터 구축 및 컨설팅 서비스업	데이터 구축·가공 서비스업	1,259	841	250	102	63	2,515
	데이터 관련 컨설팅 서비스업	401	333	69	31	10	844
소계		**1,660**	**1,174**	**319**	**133**	**73**	**3,359**
데이터 판매 및 제공 서비스업	데이터 판매·중개 서비스업	269	320	74	67	34	764
	정보제공 서비스업	1,220	527	517	73	30	2,367
소계		**1,489**	**847**	**591**	**140**	**64**	**3,131**
데이터산업 합계		**4,686**	**3,365**	**1,184**	**473**	**175**	**9,883**
데이터 인프라 서비스업	데이터 저장 장치·시설 서비스업	214	86	37	11	9	357
	데이터 네트워크 인프라 서비스업	324	101	6	10	10	451
소계		**538**	**187**	**43**	**21**	**19**	**808**
총계(인프라 포함)		**5,224**	**3,552**	**1,227**	**494**	**194**	**10,691**

7. 표본설계 방법 및 결과(표본조사)

7-1 표본추출방법

□ 데이터산업

○ 표본추출방법은 계통 추출방법을 사용함

- 이는 동일 산업 분류, 동일 종사자 규모, 동일 지역을 유지하여 원표본과 매우 유사한 예비표본(대체표본)을 준비하기 위함임
- 이때 사용한 정렬변수는 종사자 수, 지역을 사용함

□ 일반산업

○ 표본추출방법은 계통 추출방법을 사용함

- 이는 동일 산업 분류, 동일 종사자 규모, 동일 지역을 유지하여 원표본과 매우 유사한 예비표본(대체표본)을 준비하기 위함임

- 이때 사용한 정렬변수는 종사자 수, 지역을 사용함

7-2 & 7-3 표본크기 결정, 표본배분 방법

□ 데이터산업

○ 1단계 : 산업 대분류별 표본 규모 배정

- 상대표준오차 기준 배정

· 상대표준오차는 추정량의 표준오차를 모수의 추정치로 나눈 값으로 다음과 같이 정의됨

$$\widehat{RSE}\left(\hat{\bar{Y}}\right) = \frac{\sqrt{\hat{V}\left(\hat{\bar{Y}}\right)}}{\hat{\bar{Y}}} \times 100(\%)$$

· 상대표준오차의 크기에 따라서 추정 결과의 공표 여부를 결정하게 됨. 예를 들어, 상대표준오차가 25% 이상이면 대외적인 공표보다는 업무상 내부 참고자료로 이용하는 것이 타당함. 구체적인 평균 추정량의 상대표준오차 공식은 다음과 같음

$$RSE^2 = \frac{S^2(1-f)/n}{\overline{Y}^2}$$

· 여기서 f는 흔히 추출률이라 부르며 $f = \frac{n}{N}$이고 n은 실제 조사자료 수, N은 최종 모집단 수 또는 가중치의 합임. 만약 층별로 구하게 되면 $f_h = \frac{n_h}{N_h}$가 되어 층별로 얻어진 조사 자료 수와 모집단 수 또는 층별 가중치 합을 사용함. 상대표준오차 공식인 $RSE^2 = \frac{S^2(1-f)/n}{\overline{Y}^2}$에서 과거 자료와 조사자료에서 얻어진 변이계수, $\frac{S}{\overline{Y}}$가 일정하다고 가정한다면 과거 자료의 상대표준오차는 $RSE_1^2 = \frac{S_1^2(1-f_1)/n_1}{\overline{Y_1}^2}$이고 조사자료의 상대표준오차는 $RSE_2^2 = \frac{S_2^2(1-f_2)/n_2}{\overline{Y_2}^2} = \frac{S_1^2(1-f_2)/n_2}{\overline{Y_1}^2}$이 되어 다음의 공식을 만족함.

$$\frac{n_1 RSE_1^2}{(1-f_1)} = \frac{n_2 RSE_2^2}{(1-f_2)}$$

· 여기서 $f_1 = \frac{n_1}{N_1}$, $f_2 = \frac{n_2}{N_2}$이고 RSE_1은 2023년 데이터산업현황조사 기준 상대표준오차이므로 n_1, N_1은 2023년 자료를 사용하며 목표 상대표준오차인 RSE_2를 정하면 n_2를 구할 수 있음.

○ 2 단계 : 산업 중분류별 표본 규모 배정

- 산업 중분류의 경우 상대표준오차 기준과 파워 배정을 이용하여 배정한 결과를 종합한 후 최종 중분류 표본 규모를 결정함.

1) 상대표준오차 기준 배정 : $\frac{n_1 RSE_1^2}{(1-f_1)} = \frac{n_2 RSE_2^2}{(1-f_2)}$

2) 파워(멱, 멱등) 배정(power allocation)

· 층별 모집단 규모에 차이가 큰 경우에는 파워 배정을 함으로써 모집단 규모 차이에 의한 표본 규모 차이를 극복할 수 있으며 또한 자료의 표준편차를 고려함으로써 추정의 정확도가 향상될 수 있음.

· 국내에서 흔히 사용되는 파워 배정은 다음과 같이 정의됨. 먼저 h대분류의 g번째 중분류의 모집단 크기와 표본 크기 그리고 표준편차를 각각 N_{hg}, n_{hg} 그리고 S_{hg}라고 할 때 h층에 포함된 g번째 중분류의 표본 크기는 다음과 같이 얻어짐. 흔히 $\lambda = 0.5$가 사용되기 때문에 본 표본설계에서도 0.5 값을 사용하였음

$$n_{hg} = n_h \times \frac{(N_{hg} S_{hg})^{\lambda}}{\sum (N_{hg} S_{hg})^{\lambda}}$$

○ 3단계 : 종사자 규모별 표본 규모 배정

- 종사자 규모의 경우 4개를 우선할당한 후 비례 배정, 제곱근 비례 배정을 이용하여 할당한 결과의 평균으로 최종배정값을 계산한 후에 모집단 분포 등을 고려하여 일부를 조정하였음.

1) 4개 우선할당 후 비례 배정

· 중분류별로 표본이 배정되면 사업체 규모, 즉 종사자 수를 종사자 규모로 층을 나누게 됨. h대분류, g중분류의 k번째 종사자 규모 모집단 크기를 N_{hgk}, h대분류, g중분류 층에 할당된 표본크기를 n_{hg+} 라고 하면 h대분류, g중분류의 k번째 종사자 규모층에 최소한 5개 표본을 할당하도록 4개를 우선할당한 후에 나머지는 비례배정으로 표본크기를 산출함

$$n_{hgk} = 4 + (n_{hg+} - 16) \times \frac{N_{hgk}}{\sum_{k=1} N_{hgk}}$$

2) 제곱근 비례 배정

· 모집단 규모에 차이가 큰 경우 비례 배정 또는 네이만 배정을 하게 되면 큰 규모 층에 배정된 표본이 매우 작을 수 있음. 또한, 각 세부 층의 분산 추정값이 적은 표본 수에 따라 정확하지 않을 수 있어 본 표본설계에서는 제곱근 배정을 함께 고려함. h대분류, g중분류의 k번째 종사자 규모 모집단

크기와 표본 크기를 각각 N_{hgk}, n_{hgk}라고 할 때 h대분류, g중분류 층인 hg층에 포함된 k 번째 종사자 규모의 표본 규모 공식은 다음과 같음.

$$n_{hgk} = n_{hg} \times \frac{\sqrt{N_{hgk}}}{\sum \sqrt{N_{hgk}}}$$

○ **4단계 : 절충배정**

- 우선할당 후 비례 배정과 제곱근 배정의 평균으로 최종 배정을 산출한 후에 각 층별 최소표본은 5개로 정하였으며 예상 상대표준오차가 매우 큰 경우에는 이를 종합적으로 고려하여 표본 규모를 조정함

□ 일반산업

○ **1단계 : 산업 분야별 표본 규모 배정**

- 상대표준오차 기준 배정

· 상대표준오차는 추정량의 표준오차를 모수의 추정치로 나눈 값으로 다음과 같이 정의됨.

$$\widehat{RSE}\left(\hat{\bar{Y}}\right) = \frac{\sqrt{\hat{V}\left(\hat{\bar{Y}}\right)}}{\hat{\bar{Y}}} \times 100(\%)$$

· 상대표준오차의 크기에 따라서 추정 결과의 공표 여부를 결정하게 됨. 예를 들어, 상대표준오차가 25% 이상이면 대외적인 공표보다는 업무상 내부 참고자료로 이용하는 것이 타당함. 구체적인 평균 추정량의 상대표준오차 공식은 다음과 같음.

$$RSE^2 = \frac{S^2(1-f)/n}{\overline{Y^2}}$$

· 여기서 f는 흔히 추출률이라 부르며 $f = \frac{n}{N}$이고 n은 실제 조사자료 수, N은 최종 모집단 수 또는 가중치의 합임. 만약 층별로 구하게 되면 $f_h = \frac{n_h}{N_h}$가 되어 층별로 얻어진 조사 자료 수와 모집단 수 또는 층별 가중치 합을 사용함. 상대표준오차 공식인 $RSE^2 = \frac{S^2(1-f)/n}{\overline{Y^2}}$에서 과거 자료와 조사자료에서 얻어진 변이계수, $\frac{S}{\overline{Y}}$가 일정하다고 가정한다면 과거 자료의 상대표준오차는 $RSE_1^2 = \frac{S_1^2(1-f_1)/n_1}{\overline{Y_1^2}}$이고 조사

자료의 상대표준오차는 $RSE_2^2 = \frac{S_2^2(1-f_2)/n_2}{\overline{Y_2^2}} = \frac{S_1^2(1-f_2)/n_2}{\overline{Y_1^2}}$ 이 되어 다음의 공식을 만족함.

$$\frac{n_1 RSE_1^2}{(1-f_1)} = \frac{n_2 RSE_2^2}{(1-f_2)}$$

· 여기서 $f_1 = \frac{n_1}{N_1}$, $f_2 = \frac{n_2}{N_2}$ 이고 RSE_1은 2023년 데이터산업현황조사 기준 상대표준오차이므로 n_1, N_1은 2023년 자료를 사용하며 목표 상대표준오차인 RSE_2를 정하면 n_2를 구할 수 있음.

- 파워(멱, 멱등) 배정(power allocation)

· 층별 모집단 규모에 차이가 큰 경우에는 파워 배정을 함으로써 모집단 규모 차이에 의한 표본 규모 차이를 극복할 수 있으며 또한 자료의 표준편차를 고려함으로써 추정의 정확도가 향상될 수 있음.

· 국내에서 흔히 사용되는 파워 배정은 다음과 같이 정의됨. 먼저 g산업 분야의 모집단 크기와 표본 크기 그리고 데이터산업 종사자의 표준편차를 각각 N_g, n_g 그리고 S_g라고 할 때 g층의 표본 크기는 다음과 같이 얻어짐. 흔히 $\lambda = 0.5$가 사용되기 때문에 본 표본설계에서도 0.5 값을 사용하였음

$$n_g = n \times \frac{(N_g S_g)^\lambda}{\sum (N_g S_g)^\lambda}$$

○ 2단계 : 종사자 규모별 표본 규모 배정

- 종사자 규모의 경우 제곱근 비례 배정과 우선할당 후 비례 배정법을 이용하여 배정한 결과를 종합한 후 최종 종사자 규모 층의 표본 규모를 결정함.

1) 3개 우선할당 후 비례 배정

· 산업 분야별로 표본이 배정되면 사업체 규모, 즉 종사자 수를 종사자 규모로 층을 나누게 됨. g산업 분야의 k번째 종사자 규모 모집단 크기와 표본크기를 각각 N_{gk}, n_{gk} 라고 할 때 g 산업분야에 포함된 k번째 종사자 규모의 표본크기는 아래 식으로 계산하였음.

$$n_{gk} = 3 + (n_g - 9) \times \frac{N_{gk}}{\sum N_{gk}}$$

2) 제곱근 비례 배정

· 모집단 규모에 차이가 큰 경우 비례 배정 또는 네이만 배정을 하게 되면 큰 규모 층에 배정된 표본이 매우 작을 수 있음. 또한, 각 세부 층의 분산 추정값이 적은 표본 수에 따라 정확하지 않을 수 있어 본 표본설계에서는 제곱근 배정을 함께 고려함. g산업분야의 k번째 종사자 규모 모집단 크기와 표본 크기를 각각 N_{gk}, n_{gk}라고 할 때 g산업분야 층에 포함된 k 번째 종사자 규모의 표본 규모 공식은 다음과 같음.

$$n_{gk} = n_g \times \frac{\sqrt{N_{gk}}}{\sum \sqrt{N_{gk}}}$$

○ 3단계 : 절충배정

- 최소표본은 5개로 정하였으며 예상 상대표준오차가 매우 큰 경우에는 이를 종합적으로 고려하여 최종 표본 규모는 2종의 배정 결과의 평균값으로 최종 표본 크기를 산출하였음.

7-4 표본추출 결과

□ 데이터산업

○ 2024년 데이터산업 업종별, 종사자 규모별 표본추출(표본배분) 결과는 다음과 같음

< '24년 데이터산업 업종별 표본배분 결과 >

구 분		종사자 규모					
대분류	중분류	1-9인	10-49인	50-99인	100-299인	300인 이상	합계
데이터 솔루션	데이터 수집	21	19	9	11	8	68
	DBMS	15	15	5	3	4	42
	데이터 분석	30	31	12	7	3	83
	데이터 관리	44	39	14	11	14	122
	데이터 보안	20	16	8	6	7	57
	데이터 플랫폼	28	24	7	6	1	66
	소계 1	**158**	**144**	**55**	**44**	**37**	**438**
데이터 구축/ 컨설팅	데이터 구축	142	105	45	27	61	380
	데이터 컨설팅	29	26	10	7	10	82
	소계 2	**171**	**131**	**55**	**34**	**71**	**462**
데이터 서비스	데이터 판매	44	49	20	16	35	164
	정보제공	114	61	64	19	28	286
	소계 3	**158**	**110**	**84**	**35**	**63**	**450**
합 계		**487**	**385**	**194**	**113**	**171**	**1,350**

데이터 인프라 (시범조사)	데이터 저장장치	21	13	7	5	9	55
	데이터 네트워크	50	24	5	6	10	95
	소계 4	**71**	**37**	**12**	**11**	**19**	**150**
합 계(인프라 포함)		**558**	**422**	**206**	**124**	**190**	**1,500**

□ 일반산업

○ 2024년 일반산업 업종별, 종사자 규모별 표본추출(표본배분) 결과는 다음과 같음

< '24년 일반산업 업종별 표본배분 결과 >

구 분	종사자 규모					
	50-99인	100-299인	300-999인	1000-2,499인	2,500인 이상	합계
공공	.	63	36	17	14	130
금융	.	48	19	8	7	82
제조	.	114	34	11	46	205
유통	.	40	13	5	1	59
서비스	.	104	44	14	36	198
의료	.	14	6	5	27	52
통신·미디어	.	50	18	5	5	78
교육	.	96	26	13	14	149
물류	.	70	17	6	5	98
유틸리티	.	20	7	5	1	33
농축산·광업	18	8	5	.	.	31
건설업	.	27	12	5	16	60
숙박·음식점업	.	15	8	2	.	25
합 계	**18**	**669**	**245**	**96**	**172**	**1,200**

* 농축산·광업은 동일한 종사자 수 기준으로 모집단을 구성하는데 한계가 있어 50인 이상으로 구성

7-5 표본설계보고서

○ 표본설계명세서에는 조사 개요 및 목적, 모집단 구축, 기존(2023년) 표본설계, 새로운 (2024년) 표본설계 등이 포함되어 있음

7-6 표본설계보고서에 모수 및 분산 추정방법

□ 추정식

○ 업종별 총계 추정치

$$\widehat{Y}_i = \sum y_{ic} + \sum_{h=1} (w_{ih}) \sum_{i=1} y_{ih}$$

여기에서 $\widehat{Y}_i$: i업종의 총계추정치

y_{ic} : 전수층에 속하는 사업체의 특성

w_{ih} : i업종, 종사자규모 h 표본층의 가중치

y_{hi} : i업종, 종사자규모 h 표본층 사업체의 특성

○ 분산추정치

$$\widehat{V}(\widehat{Y}) = \sum_{h=1} {}_iN_h^2 \left(\frac{{}_iN_h - {}_in_h}{{}_iN_h} \right) \frac{s_h^2}{{}_in_h}$$

여기에서 $s_h^2 = \frac{1}{{}_in_h - 1} \sum_{i=1} \left(y_{ih} - \overline{y_h} \right)^2$ 은 표본층의 분산추정치

여기에서 $\widehat{V}(\widehat{Y})$: 분산추정치

${}_iN_h$: i업종, 종사자규모 h 표본층의 크기

${}_in_h$: i업종, 종사자규모 h 표본층의 표본크기

y_{ih} : i업종, 종사자규모 h 표본층 사업체의 특성

$\overline{y}_h$: 종사자규모 h 표본층의 평균

○ 허용오차(95% 신뢰구간)

$$B = 1.96^* \sqrt{\widehat{V}(\widehat{Y})}$$

Ⅲ. 자료수집

1. 조사방법

1-1 조사방법 선택에 대한 검토

○ 조사방법

- 온라인, 이메일, 전화, 방문면접조사를 복합적으로 활용
- 온라인조사 : 웹설문지를 통해 회수
- 이메일: 사전 송부된 조사표를 이메일로 회수함
- 전화: 방문면접, 메일, 온라인 조사후 미흡한 부분에 대해 전화를 통해 응답받음
- 방문면접조사: 온라인 조사표를 사전에 응답자에게 송부 후, 방문면접일을 잡아 조사원이 직접 방문하여 조사 및 설문을 회수함

1-2 선택한 조사방법에 대한 조사과정

○ 1단계: 추출된 표본 대상으로 훈련받은 컨택원이 전화하여 기업명, 업종, 주소 등 사업체의 기본정보 확인한 후 조사응답에 적합한 대상을 찾아 컨택함

○ 2단계: 컨택된 기업의 조사의사 확인 및 조사방법을 선택. 편의에 따라 원하는 조사방법을 선택하게 하여 응답률 향상

○ 3단계: 각 조사방법에 따른 조사 진행

- (온라인 조사) 조사원이 사전 컨택 및 웹 설문페이지를 발송한 후, 주기적인 수신확인 및 독려. 웹페이지상 주요 문항의 상세설명이나 경고창 등 제공하여 응답편의성 제고
- (이메일, 전화조사) 응답자가 방문조사를 거부한 경우 다음의 조사를 활용하며, 응답자 이메일을 통해 조사표와 조사협조 공문을 발송한 후, 해당내용을 회수. 회수된 내용을 점검하고 부족한 내용이 있으면 전화를 통해 내용 보완
- (방문면접조사) 조사원교육을 참석하고 조사지침서를 숙지한 조사원이 조사표 및 조사지침서 등 조사에 필요한 용품을 지참하여 방문 면접조사 실시. 사전에 방문 일정을 조율한 후, 방문하여 조사 취지를 설명한 후 조사표 작성 및 의문사항 수시로 확인

< 조사방법별 프로세스 >

온라인조사	전화/이메일	방문면접조사
1) 조사메일 발송에 대한 안내문자 발송	1) 전화번호 확인 및 메일발송	1) 조사대상자와 방문약속 확정
2) 시스템 상에서 조사메일 수신 확인	2) 메일발송에 대한 전화 또는 문자로 안내	2) 방문 하루 전(또는 반일 전) 방문일시 재확인
3) 2일 이내 미수신인 경우 재전화 컨택 및 재발송 실시	3) 2일 이내 재전화 컨택으로 수신여부 확인	3) 방문일정 변경을 요청하는 경우, 그에 따라 일정 재조정
4) 시스템 상에서 조사참여 확인	4) 5일 이내 회신이 없을 경우 재전화 컨택 및 재발송 실시	4) 방문 전 조사표, 응답지침 등의 자료 확인과 복장 점검
5) 2일 이내 미참여자에게 재전화 컨택 및 재발송 실시	5) 조사완료 후 감사인사 및 검증연락에 대한 안내	5) 조사대상 방문하여 조사진행
6) 조사완료 후 감사인사 및 검증연락에 대한 안내		6) 조사완료 후 감사인사와 검증연락에 대한 안내

2. 조사원 채용 및 처우

○ 자격요건

< 조사원 자격요건 >

구분	주요내용
경력	면접조사 경력 5년 이상
	최근 3년간 면접조사 경험 50회 이상
	최근 3년간 사업체 관련 조사 경험 5회 이상
	전년도 조사 수행 우수 조사원 우선 투입
등급	당사 시스템 내 A등급 이상인 조사원
기타(전화컨택원)	상황대처 능력이 탁월하고 상대방과 커뮤니케이션 능력이 높은 조사원
	대규모 조사 유경험자(성실도, 목표 할당량 달성)
	과거 유사프로젝트 진행 성과평가가 높은 조사원
기타(방문 조사원)	용모 및 복장이 단정한 조사원
	민원 발생 경험이 없는 조사원

○ 지위 : 연단위 계약 개인사업자

○ 급여수준 : 조사완료 수당 16,000원

○ 지급방법 : 근로계약서에 기재된 계좌에 입금

○ 부가혜택

- 방문조사 등으로 출장 시, 출장비 지급
- 전수층 전담 면접원의 경우 성공보수 추가 제공

3. 조사원 교육

3-1 조사원 교육에 대한 일정

○ 수도권 조사원 교육은 집합교육으로 진행되었음(전화 사전컨택 면접원 및 본조사 면접원)

- 조사원 교육은 녹화하여 지방 교육과 교체된 조사원 교육 시 활용하였음

3-2 조사원 교육 내용 및 교육시간의 적절성

□ 조사원 교육 내용

○ 조사원 교육 내용은 크게 조사 및 기관 소개, 설문 내용 및 설문 조사 시 유의사항, 평가 및 질의응답으로 이루어졌음

- 추가적으로 컨택 프로세스 및 시스템 교육과 CAPI 시스템 교육을 진행하였음
- CAPI 시스템 교육에 CAPI 기능과 조사표 로직 교육이 포함됨

□ 교육시간의 적정성 검토

○ 교육시간은 다음과 같이 이루어짐

< '24년 데이터산업 현황조사 조사원 교육 일정 >

시간	내용	교육 담당자
10:00 ~ 10:10	• 조사 개요 소개 • 한국데이터산업진흥원 소개	OOO
10:10 ~ 11:30	• 설문 내용 및 설문 조사 시 유의사항	OOO
11:30 ~ 12:00	• 평가 및 질의 응답 • 최종 정리	OOO

3-3 교육 교재

○ 교육 교재는 2종으로 조사원 지침서(한글)과 조사원 교육자료로 구성됨

3-4 조사기간 중 교체된 조사원에 대한 교육 실시

□ 교체 조사원 교육

○ 교체된 조사원은 없음

3-5 조사원 대상 비밀보호 의무 교육 또는 서약서 작성

○ 비밀보호 의무 교육 실시

○ 조사원 대상 보안서약서 작성

3-6 조사원의 업무지식 숙지 정도에 대한 평가 및 평가 조치(재교육 실시 등)

○ 조사원 교육 실시 후 교육 내용 평가 진행

- 교육 점수 80점 이상만 조사에 투입하였으며, 80점 미만은 담당 SV와 1:1 재교육 실시 후 투입하였음

4. 조사업무 흐름도

4-1 조사실시에 대한 조사업무 흐름도 관리

□ 조사업무 흐름

< 조사업무 흐름도 >

주요 업무 내용	담당자
조사 실시 안내 및 안내자료 발송에 대해 사전 전화 컨택	컨택원
⇩	
조사 자료를 이메일로 조사대상 사업체 사전 발송	컨택원
⇩	
조사요령 등에 대한 조사원 교육 실시	연구원 조사원 담당SV
⇩	
조사원별 조사용품 수령(조사표, 지침서, 협조공문 등)	조사원 담당SV
⇩	
조사대상 사업체에 대한 정보 확인	조사원 담당SV
⇩	
해당 사업체 방문 후 개별면접조사, 해당 사업체와 직접 전화조사 등 실시	조사원
⇩	
조사 완료된 조사표를 최종 내용검토 후 제출(내검)	조사원

5. 조사준비 및 준비조사

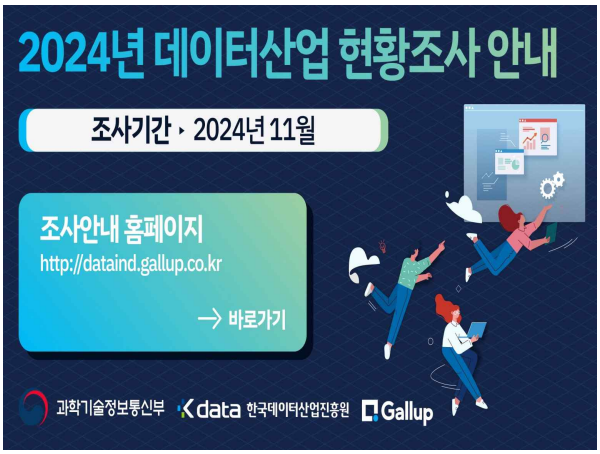
5-1 조사실시 홍보 내용과 방법

□ 조사의 홍보

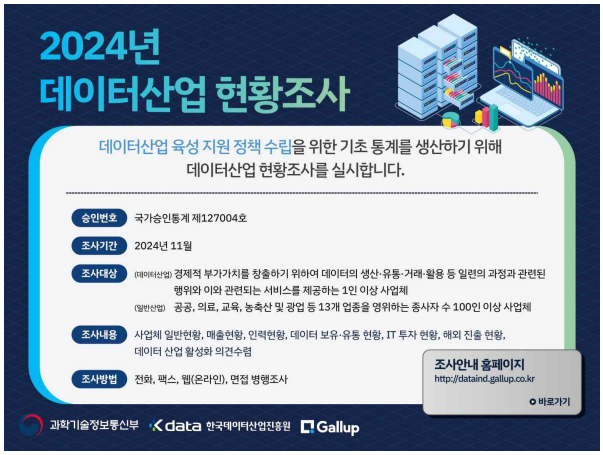
○ 웹배너 제작 및 통계작성위탁기관(한국데이터산업진흥원) 홈페이지 내 조사 안내 홍보물 게시

- 응답자의 조사에 대한 신뢰성 제고

< 웹배너 >



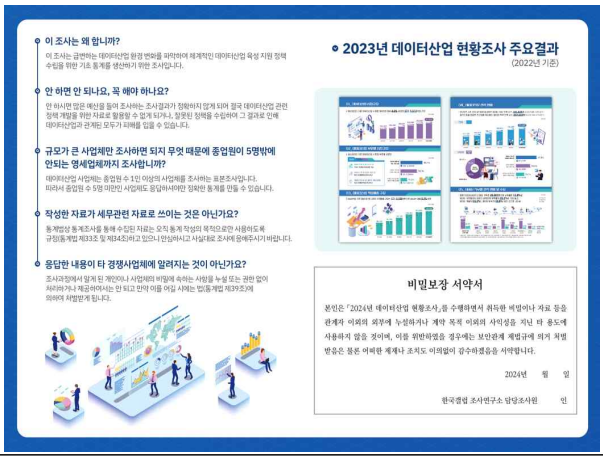
< 팝업 이미지 >



< 리플렛(앞) >



< 리플렛(뒤) >



5-2 응답자(조사대상) 사전 통지

□ 조사대상 사전 통지, 컨택

○ 조사대상 업체에 조사 이해도 향상과 참여의 필요성을 이해시키기 위하여 유선상 조사 실시 안내하고, 방문 사전 일정 협의

○ 조사목적, 법적근거, 비밀보호, 통계목적만으로 사용 등 조사 필요성과 안정성을 충분히 설명하고 공문 협조 요청

○ 조사 메일 내에 공문, 조사 안내 리플렛, 안내문을 포함하여 조사에 대한 설명 및 협조 요청

5-3 조사구 확인 또는 조사명부 보완

○ 조사대상이 개별 사업체이므로 해당 사업체에 해당하는 실적만 조사

○ 사업 종류(주사업) 내용이나 사업체가 변경된 경우 주사업 산업분류를 정확히 파악해 조사표 작성 안내

- 2023년 1년간 매출규모를 중심으로 매출액 비중이 가장 큰 산업 활동을 주사업으로 분류하고, 겸업의 경우 다음으로 매출액 비중이 가장 큰 산업을 부사업으로 기입

· 사업체의 종사자수나 매출액 규모와 관계없이 부사업 여부를 조사함

○ 사업체 정보가 부족하거나 변경된 경우, 현장 확인

- 사업체명 : 사업자등록상의 사업체명 기입 여부
- 사업자등록번호 : 세무서에서 부여한 고유번호로 사업자등록증에 표기된 번호
- 대표자명 : 지점, 영업소의 경우 본사 대표명을 기입하지 않도록 주의
- 소재지 : 도로명, 건물번호 및 빌딩, 시장, 상가 등 명칭과 동/층/호까지 구체적으로 기입
- 창설일 : 현행 산업활동을 처음 시작한 연월을 기입하고, 주요업종·조직형태·대표자가 변경된 경우 변경된 때를 창설연월로 기입(단, 법인대표변경, 일시적 휴·폐업 이후 동일 업종으로 재개한 경우, 장소만 이동한 경우 제외)

6. 조사항목별 조사 방법

6-1 주요 조사항목별 작성요령 및 유의사항

○ 전체항목 유의사항

- 수치자료는 아라비아숫자로 기입하며, 소수점 단위 미만 수치는 반올림하여 소수점 첫째자리까지 기입
- 사업실적(매출액), 투자금액 등 금액자료 조사는 조사표에 기재된 단위로 기입(억원, 백만원 등)
- 단일 선택형 질문은 제시된 보기 중에서 하나의 응답만을 선택하여 해당 번호에 표시하며, 중복응답의 경우 복수응답 가능
- 조사내용을 수정하는 경우 수정 전후 내용을 알아볼 수 있도록 수정 기입

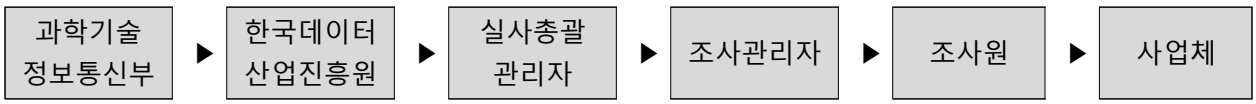
○ 문항별 유의사항

- 데이터사업 매출 현황: 데이터사업분야 소분류 코드 기입후 주 사업분야를 선택하고, 백만원 단위로 기입. 사업분야별 매출액과 매출비중의 합은 전체 값보다 같거나 작도록 작성. 국내 매출액과 해외 매출액의 합은 전체 값보다 같거나 작아야 함.
- 데이터직무 현황 : 1명이 여러 업무를 담당하는 경우 가장 주된 업무를 수행하는 분야로 작성. 초급/중급/고급 인력의 합이 전체 인력수와 일치하도록 작성. 각 영역별/등급별 인력 수의 합은 소계와 같아야 함.

- IT예산 및 해외진출 현황: 수출액 및 IT 예산은 백만원 단위로 기입. 해외진출 현황 상세 내역은 해외진출 경험자에 한해 기입
- 정책수요 등 그 외 문항: 복수응답으로 답변하되, 지정된 순위까지 빠짐없이 선택

7. 조사 관리

7-1 & 7-2 조사 관리 체계, 조사 관리 방법



○ 조사기간 중에는 조사업무에만 집중해야 함

○ 조사원은 지정된 일자 및 장소에서 조사지침에 대한 교육을 받은 후 조사업무를 수행하기 이전에 지침서의 내용을 충분히 숙지, 사업체 방문 시에도 조사지침서를 휴대하고 의문 발생 시 확인 후 조사

○ 조사기간 중 완료된 조사표는 수시로 조사관리자에게 제출하여 내용 검토를 받고, 실사총괄 관리자의 검토사항에 대해서 즉시 보완하도록 하며, 조사 진행사항 등도 조사관리자에게 수시로 보고하고 업무지시에 따라야함

○ 조사기간 중 조사원증을 지참하여 조사시 제시하여야 함

○ 특별한 사유(질병 등)로 조사업무를 수행하지 못하는 경우 지체 없이 조사관리자(또는 실사총괄관리자)에게 보고

○ 조사 과정에서 알게 된 개인이나 사업체의 조사 내용은 통계 업무 외의 목적에 사용해서는 안 됨[통계법 제 33조(비밀의 보호)]

○ 조사원증, 조사지침서 등 조사에 사용한 물품류는 조사가 끝나면 모두 반납

○ 안전사고에 유의, 영업행위나 종교 활동 금지, 안전사고에 따른 산재보험 혜택을 받을 수 있도록 조사원 각자는 미리 산재보험에 가입토록 함

7-3 조사 관리자 역할

○ 한국데이터산업진흥원 담당자

- 조사관련 업무 총괄 및 조사업무 제반 점검
- 조사 설계 및 조사 애로사항 청취 및 반영
- 통계 분석
- 조사원 대상 실사지도 실시
  · 조사원 교육: 데이터 산업분류 및 특성, 주요 확인사항 설명 등
  · 위탁기관 대상 실사 지도: 상시 진행하되, 주요 내용은 위와 같음

○ 실사총관리자

- 실사 총괄 및 일정 관리
- 조사관련 업무 지시사항 전달 및 보고

- 안내장 발송 관리

○ 조사관리자

- 조사원 관리, 민원 즉시 대응
- 조사진행 및 점검, 실사일정 관리
- 조사 문의사항 수렴 및 보고
- 내검 및 검증관리
- 답례품 관리

8. 조사 질의응답 체계

8-1 조사질의 및 응답체계 운영 방법

○ 응답자별 질의 방법

- 조사원에게 질의: 조사진행 시 즉시 응대가 가능한 방법이나 깊이있는 응답은 어려움
  · 주요 질의내용은 조사지침서에 사전 안내되어 응대가능하도록 함
- 조사관리자, 용역PM 질의: 조사진행시 즉시응대 어려우나 깊이있는 질의 가능
  · 공문, 조사표 등에 담당자 연락처를 기재하여 자유롭게 질의할 수 있도록 제공

○ 주요 질의응답 예시

< 주요 질의응답 예시 >

	예상 질의	답변
1	이 조사를 왜 합니까?	이 조사는 데이터산업 현황을 파악하여 데이터산업 정책 수립의 기초자료로 활용하기 위한 것입니다.
2	조사를 꼭 해야하나요?	사업체의 적극적인 협조가 없으면 정확한 통계가 작성되기 어렵고 정부에서는 부정확한 통계로 해당 산업에 대한 정책수립을 하게 되어 경제 전반에 커다란 손실이 올 수도 있습니다. 또한 국민의 세금으로 하는 조사인만큼 정확한 조사가 이루어지지 않을 경우 국민이 납부한 세금을 낭비하는 결과가 됩니다. 따라서 정부에서는 부실조사 방지를 위하여 통계법에 개인이나 법인 또는 단체 등으로 하여금 통계조사에 성실하게 응답(통계법 제32조)하도록 제도적으로 명시하고 있습니다. 바쁘시겠지만 데이터 산업 발전을 위해 귀중한 의견 부탁드립니다.
3	어디서 조사합니까?	과학기술정보통신부가 주관하고 한국데이터산업진흥원과 조사기관이 함께 조사를 수행하고 있습니다.

8-2 주요 질의 응답 · 오류사례 축적 및 관리

○ 유선 콜센터 및 메일 등을 통한 질의응답 채널 구축

○ 질의응답을 정리하여 조사원에게 전파 및 향후 조사원 교육 시 활용

○ 주요 질의 응답·오류사례

- 올해 신규 조사 대상 중, 2023년 매출액이 0인 경우가 있는데 조사해도 되나요? : 조사 진행은 가능하며, 매출액이 0인 이유에 대해 확인
- 주사업 분야에 회사 주력 사업을 넣는 것인지? : 매출 기준으로 주사업 분야를 선택
- 전년도 채용 인력이 현재는 퇴사했는데, 제외하고 응답해야하는지? : 현재 근무 여부와 관계없이 실제로 채용한 인력을 기준으로 응답
- 순위 문항은 꼭 3순위까지 선택해야하는지 : 가급적 3순위까지 모두 선택

9. 조사(또는 응답) 대상

9-1 적격 조사(또는 응답) 대상의 지정 이유

□ 설문 적격응답자

○ 기업의 전반적인 사항에 대한 질문이 많기 때문에 데이터 관련 사업부서 또는 해당 기업 의견을 대표할 수 있는 관리자급, 임원급 이상이 본 조사에 응답해야 함

- 응답자의 주요 업무를 확인후 설문내용과 유사한지 파악하여 적격 응답자여부 판단
- 만일, 관리자급 또는 임원급보다 실무진이 응답내용에 대해 더 많이 알고있는 경우에는 응답을 가장 잘할 수 있는 대상을 중심으로 조사를 진행할 수 있음

9-2 기억응답과 관련된 검토 결과

□ 장부 또는 경영 현황 자료에 의한 응답

○ 주요 직무인력 및 재무현황은 경영 장부 또는 결산 집계표 등을 참고하여 조사표를 작성하므로 기억응답은 거의 없음

- 임금지출대장, 매입·매출명세서, 현금 출납장부, 대차대조표, 매출 집계표 등의 객관적 자료를 1차로 활용
- 직무 인력의 경우, 주요 직무 업무별 보기 카드를 제시하여 응답

10. 무응답 대처

10-1 항목무응답 대처 방법

○ 데이터산업 현황조사에서 불가피하게 발생한 무응답 대체방법은 공시자료를 통해 조사기준시점과 동일한 자료가 존재하는 경우 외부 공시자료를 우선활용하고, 직전년도 응답값 활용을 통해 무응답을 보완하고자 함

10-2 단위무응답 대처 방법

○ 단위무응답 중 불응업체 경우 1주일 간격으로 3회까지 조사요청을 하며, 부재의 경우 다른 담당자를 확인후 추가 조사 협조를 요청하고, 이후에도 조사가 불가한 경우 응답 대체함

○ 전자공시시스템 등 확인할 수 있는 정보가 있는 경우 응답 대체값을 처리

○ 23년 조사에서 응답한 사업체 경우 23년 응답한 값에 24년 응답변화 양상을 적용 활용

11. 표본대체

11-1 표본대체 허용 기준

□ 표본 대체

○ 업종x종사자규모로 정렬 후, 계통추출을 통해 본표본과 예비표본을 뽑음.

○ 본표본 응답 거절, 불가 등일 경우 다음 예비표본 진행

○ 본표본 성공률을 높이기 위해, 최대 10회 비수신 시 응답 대체 진행

- 단순 거절일 경우에는 재컨택하여 응답 독려 진행
- 강력거절, 결번·번호오류, 휴폐업 등 조사가 불능일 경우에는 응답 대체 진행

○ 전수층의 경우 예비표본이 없기 때문에 응답 대체 불가능하며 업체의 편의를 고려하여 조사 진행

11-2 & 11-3 표본대체 기준, 절차 및 방법

○ 응답자의 조사거절, 불응, 조사협조가 불가능한 경우, 3회 이상 독려 전화 및 방문하여 조사 당위성을 강조하고 조사에 대한 협조 구함

○ 단위무응답이 전수층에서 발생한 경우에는 표본 대체 불가하였으며, 표본조사 업체 경우 동일층의 업체중 가장 근접한 업체를 선정한 후 조사

○ 단위무응답의 응답 대체 시, 원칙은 보강조사(인터뷰 등)를 통한 대상 사업체로부터 직접 응답을 확보하는 것이나 수차례 응답 거부한 사업체에 대해 응답 대체 값 처리

12. 사후조사

12-1 조사 실시 후 사후조사 실시

□ 데이터 리체크 및 검증 실시

○ 조사 완료 80%된 시점에 조사 리체크와 조사 검증 실시

- 응답 유무, 응답 내용 리체크 등 진행

12-2 사후조사 수행에 대한 결과 분석 및 사후조치 방안

□ 결과 및 활용

○ 응답 재확인 결과, 오류가 있는 응답은 수정 보완

Ⅳ. 통계처리 및 분석

1. 자료코딩

1-1 자료 코드체계 및 코딩(부호화) 방법

○ 회수된 모든 조사표는 불성실 응답, 논리오류, 허위작성 오류 등을 제거하고, 사업체 일반현황, 의견 등은 한글로 입력하며, 나머지 항목에 대해서는 코딩(수치화) 함

○ 항목 변수: 보기번호 입력

○ 연속변수(매출액, 인원 등) 변수 : 응답값 그대로 입력

- 재무정보 응답은 백만원 단위(최종보고서는 일부 억원 단위로 정리되나, 조사는 백만원 단위로 진행)
- 비율 응답은 소수점 한자리까지 표기

○ 그 외의 경우

- 응답거절의 경우 거절코드로 입력, 해당사항 없는 경우 결측

○ 코드체계

- 사업체 일반현황은 한글로 기입
- 유무변수: 예/아니오 입력변수 2개로 코드화
- 척도문항: 척도수준에 따라 코드화
- 우선순위 항목: 최대 3순위까지 코드화

2. 자료입력

2-1 조사결과 자료의 전산입력 방법

○ 사업체 대한 조사가 완료되면 조사 내용 입력

- 조사표의 모든 내용을 입력하며 입력오류를 최소화함

○ 입력오류 검출을 위한 적용

- 입력전에 일반 에러 체크 실시(별도 시스템)
- 내검 관련 규칙들이 시스템 내 내장되어 입력단계에서 자동으로 오류 위치 확인 가능
- 총계와 합산 등을 자동으로 계산하여 입력 오류 최소화
- 사업체 명부 정보가 변경되는 경우, 명부에 우선 정보 수정 후 조사표 입력

2-2 입력 시 오류 검출을 위한 방법

○ 본 조사는 CAPI시스템을 통해 응답하도록 되어있으며, 입력 시 항목 응답값을 벗어나는 경우 다음 문항으로 넘어가지 않거나, 오류를 알려주는 소프트 로직, 하드 로직이 걸려있음

○ 총계와 합산, 비율 등을 자동으로 계산하여 입력 오류 최소화함

2-3 입력매뉴얼(지침서)

○ 통계작성지침서의 조사표 작성요령에 수록되어 있으며, 조사원 교육자료에도 명시

2-4 자료입력 교육 실시 여부와 교육 일정 및 방법

○ 모든 조사는 CAPI 시스템을 통해 응답자 또는 조사원이 직접 입력하도록 되어있으며, 관련 교육 실시

3. 자료내검

3-1 조사현장 내검 내용 및 방법, 오류자료 처리방법

○ 조사현장 내검

- 조사대상 사업체별로 조사하였는지 여부 확인
- 조사 내용에 기존 사업체 정보와 다른 정보가 있는지, 변경하였는지 확인

4. 주요 항목무응답 실태

4-1 주요 항목에 대하여 항목무응답률 수치

○ 각 조사항목간 긴밀한 연관(논리오류 등)을 가지고 있어 항목 무응답을 인정하지 않음

- 예) 2023년 매출액을 받지 않으면 다음 문항 진행이 불가능함

5. 항목무응답 대체

5-1 항목무응답 대체방법

○ 항목 무응답대체

- 조사 응답표 점검 시 조사가 되어야 하는 항목 중 일부가 누락된 경우 그 이유를 반드시 확인하고, 실사 관리자는 조사표를 매일 점검하여 응답해야 할 질의 중 내용이 누락된 경우에는 해당 사업체에 직접 전화를 걸어 각 누락 사유별로 대응하여 최대한의 응답을 확보
- 실사 관리자가 2~3회 이상 반복 조사를 시행하여도 응답을 거부하거나 모름으로 일관한 경우에는 조사관리자에게 보고하여 각 문항의 응답을 확보하기 위해 다른 담당자를 컨택하여 관련 응답을 확보

5-2 항목무응답 대체 시 대체 비율, 대체값의 추정치 기여도, 대체값의 자료 표기 방법 등 분석

○ 반복적인 점검과 확인을 진행하고, 무응답인 경우 대체표본을 활용하여 항목 무응답률 최소화를 위해 노력하였음

○ 본 조사에서 항목무응답 존재하지 않음

6. 단위무응답 실태

6-1 단위무응답률 수치

○ 조사 중 단위무응답 발생 시 표본대체를 통해 단위무응답이 발생하지 않도록 하고 있으며, 전수층의 경우 표본대체가 불가능한 경우에만 단위무응답이 발생하였음

6-2 단위무응답률 산출식

○ 단위무응답률(%)

- $\text{단위무응답률}(\%) = \frac{\Sigma \text{단위무응답사업체수}}{\Sigma \text{최초표본사업체수}} \times 100$

6-3 주요 하위그룹 및 무응답 사유별 무응답률

○ 데이터산업

< '24년 데이터산업 종사자 규모별 무응답 현황 >

종사자 규모	최초 표본 (개)	무응답 표본(개)	무응답 률(%)	무응답 사유(개수)				
				비 수신	거절	결번·번호 오류	휴·폐업	규모 변동
1. 1-9인	558	0	-					
2. 10-49인	422	5	1.2	1	2	1	1	
3. 50-99인	206	12	5.8	2	1	3	6	
4. 100-299인	124	9	7.3	1	4	3	1	
5. 300인 이상	190	47	24.7	8	22	9		8
총합계	1,500	73	4.9	12	29	16	8	8

< '24년 데이터산업 업종별 무응답 현황 >

업종 중분류	최초 표본 (개)	무응답 표본(개)	무응답 률(%)	무응답 사유(개수)				
				비 수신	거절	결번·번호 오류	기타	규모 변동
1_1. 데이터 수집 ·연계 솔루션 개발·공급업	68	1	1.5		1			
1_2. 데이터베이스 관리 시스템 솔루션 개발·공급업	42	0	-					
1_3. 데이터 분석 솔루션 개발·공급업	83	2	2.4		1	1		
1_4. 데이터 관리 솔루션 개발·공급업	122	8	6.6	1	4	3	1	
1_5. 데이터 보안 솔루션 개발·공급업	57	2	3.5	1	1			
1_6. 데이터 통합 플랫폼 솔루션 개발·공급업	66	0	-					
2_1. 데이터 구축 및 가공 서비스업	380	16	4.2	5	5	2		4
2_2. 데이터 관련 컨설팅 서비스업	82	2	2.4		2	1		
3_1. 데이터 판매 및 중개 서비스업	164	15	9.1	1	6	5		4
3_2. 정보제공 서비스업	286	20	7.0	3	5	9	6	
4_1. 데이터 저장 장치/시설 서비스업	55	6	10.9	1	3	2	1	
4_2. 데이터 네트워크 장치/시설 서비스업	95	1	1.1		1			
총합계	1,500	73	4.9	12	29	23	8	8

○ 일반산업

< '24년 일반산업 종사자 규모별 무응답 현황 >

종사자 규모	최초 표본 (개)	무응답 표본(개)	무응답 률(%)	무응답 사유(개수)				
				비 수신	거절	결번·번호 오류	기타	규모 변동
1. 50-99인	18	0	-					
2. 100-299인	669	1	0.1					1
3. 300-999인	245	0	-					
4. 1,000-2,499인	96	0	-					
5. 2,500인 이상	172	57	33.1	5	24	6	4	23
총합계	1,200	58	4.8	5	24	6	4	24

< '24년 일반산업 업종별 무응답 현황 >

업종 분류	최초 표본(개)	무응답 표본(개)	무응답률(%)	무응답 사유(개수)				
				비 수신	거절	결번·번호 오류	기타	규모 변동
1. 공공	130	2	1.5			1	1	
2. 금융	82	0	-					
3. 제조	205	23	11.2	3	13	4	1	2
4. 유통	59	0	-					
5. 서비스	198	7	3.5	1	5	1		
6. 의료	52	10	19.2		1		2	7
7. 통신/미디어	78	3	3.8	1	2			
8. 교육	149	4	2.7					4
9. 물류	98	0	-					
10. 유틸리티	33	1	3.0					1
11. 농축산/광업	31	0	-					
12. 건설업	60	12	20.0		3			9
13. 숙박/음식점업	25	1	4.0					1
총합계	1200	63	5.3	5	24	6	4	24

6-4 단위무응답에 의한 편향 발생 및 분산 증가 가능성

○ 해당 사항 없음

○ 단위무응답 발생 시 무응답 대응 매뉴얼에 따라 표본대체를 실시하여 단위무응답 발생률이 낮아 단위무응답에 의한 편향 및 분산 발생 가능성은 미미함

6-5 항목 또는 단위무응답 발생 시, 응답자와 무응답자의 성향으로 인해 발생할 수 있는 편향을 줄이기 위한 조치

○ 항목무응답의 경우 조사 및 검토 과정에서 최대한 보완하여 무응답 발생 최소화

○ 단위무응답의 경우 사유에 따라 대처 방안 사용

○ 조사신뢰도 제고를 통한 응답 유도

- 조사안내문, 공문, 리플렛 발송
- 데이터산업현황조사 를 통한 조사 안내

○ 조사협조도 제고를 통합 응답 유도

- 조사완료 후 답례품 지급

7. 가중치 조정

7-1 설계가중치 산출과정 및 방법

○ 설계 가중치

- 전수 조사층 : 모든 사업체를 조사한다는 가정하에 조사된 표본사업체는 모집단 사업체 1개를 대변하므로 가중치는 1이 됨
- 표본 조사층 : 특정 i업종에서 종사자규모 h층의 조사모집단크기인 N_{ih}개와 표본조사된 사업체수 n_{ih}를 비교하여 표본조사된 사업체 1개는 모집단의 $\frac{N_{ih}}{n_{ih}}$를 대표하므로 이를 설계가중치라고 하며 다음과 같이 표현할 수 있음

$$w_{ih} = \frac{N_{ih}}{n_{ih}}$$

7-2 무응답 가중치 조정과정 및 방법

○ 무응답 조정 가중치

- 해당하는 i업종의 종사자 규모 h층 내에서 할당된 사업체와 조사된 사업체수가 다를 경우에는 무응답 사업체에 대한 무응답 조정을 위해서 가중치를 부여함

$$\text{무응답 조정 가중치} = \frac{n_{ih}}{r_{ih}}$$

- 여기서 n_h와 r_h는 각각 종사자규모 h층에 할당된 표본과 응답된 표본의 크기를 나타냄

7-3 사후가중치 조정과정 및 방법

○ 조사 전 모집단의 업종과 규모는 2024년 ICT 통합모집단에 기록된 내용으로 표본설계 진행. 2024년 ICT 통합모집단의 정보는 2022년 전국사업체조사의 DATA로 2024년 조사 당시와 2년의 차이가 발생

○ 본 조사에서 활용하는 데이터산업 분류는 23년 매출액을 기준으로 응답자가 데이터사업 5개를 응답 후, 그 중 업체의 주사업 하나를 선택함. 종사자규모는 문5에서 현재 기준 전체 종사자 수를 응답하고 그에 따라 쿼터 체크를 함

○ 업체의 사업 상황에 따라 사업과 주사업은 변경될 수 있고 종사자규모 또한 22년 전국사업체조사 당시와 변동이 있을 수 있으므로 실제 응답값에 따라 조사 할당을 관리함

○ 응답 업체의 변경된 업종과 종사자규모는 조사 완료 후 모집단에 반영하고 그에 따라 사후가중치를 조정함

8. 통계추정 산출식 및 내용

8-1 (표본조사)추정하고자 하는 주요 모수

○ 데이터산업 분류별 데이터 매출액, 데이터직무 인력 수 등

8-2 (표본조사)추정치를 계산하는 산출식

□ 추정식

○ 업종별 총계 추정치

$$\hat{Y}_i = \sum y_{ic} + \sum_{h=1} (w_{ih}) \sum_{i=1} y_{ih}$$

여기에서 $\hat{Y}_i$: i업종의 총계추정치

y_{ic} : 전수층에 속하는 사업체의 특성

w_{ih} : i업종, 종사자규모 h 표본층의 가중치

y_{hi} : i업종, 종사자규모 h 표본층 사업체의 특성

○ 분산추정치

$$\hat{V}(\hat{Y}) = \sum_{h=1} {}_iN_h^2 \left(\frac{{}_iN_h - {}_in_h}{{}_iN_h} \right) \frac{s_h^2}{{}_in_h}$$

여기에서 $s_h^2 = \frac{1}{{}_in_h - 1} \sum_{i=1} \left(y_{ih} - \overline{y_h} \right)^2$은 표본층의 분산추정치

여기에서 $\hat{V}(\hat{Y})$: 분산추정치

${}_iN_h$: i업종, 종사자규모 h 표본층의 크기

${}_in_h$: i업종, 종사자규모 h 표본층의 표본크기

y_{ih} : i업종, 종사자규모 h 표본층 사업체의 특성

$\overline{y}_h$: 종사자규모 h 표본층의 평균

○ 허용오차(95% 신뢰구간)

$$B = 1.96^* \sqrt{\hat{V}(\hat{Y})}$$

□ 단위무응답 사업체의 데이터 관련 매출액 추정 방안

○ 2023년 조사에 응답한 경우

- 과거 응답한 결과를 기준으로 증감률 개념을 적용. 즉, 해당 기업과 동일한 분류에 속하는 기업들의 2023년~2024년 평균 증감률을 도출하여 적용함으로써

무응답 기업의 2024년 데이터 관련 매출액을 추정

$$y(i)_{hk} = y^*(i)_{hk} \times \frac{\overline{y}(i)_h}{\overline{y^*}(i)_h}$$

$y(i)_{hk}$: i번째 항목에 대하여 h번째 세부 층의 k번째 사업체의 대체 값
$y^*(i)_{hk}$: i번째 항목에 대하여 h번째 세부 층의 k번째 사업체의 전년도 값
$\overline{y}(i)_h$: i번째 항목에 대하여 h번째 세부 층의 평균값
$\overline{y^*}(i)_h$: i번째 항목에 대하여 h번째 세부 층의 전년도 평균값

○ 2024년 조사에 응답한 경우

- 동일 중분류, 동일 규모에 해당하는 사업체들의 전체 매출액 대비 데이터 관련 매출액의 비율을 적용. 즉, 해당 사업체와 동일한 분류에 속하는 사업체들의 전체 매출액에서 데이터 관련 매출액이 차지하는 비율의 평균을 도출하여 적용함으로써 무응답 사업체의 2024년 데이터 관련 매출액을 추정

$$x(i)_{hk} = x^*(i)_{hk} \times \frac{\overline{x^*}(i)_h}{\overline{x}(i)_h}$$

$x(i)_{hk}$: i번째 항목에 대하여 h번째 세부 층의 k번째 사업체의 데이터 관련 매출
$x^*(i)_{hk}$: i번째 항목에 대하여 h번째 세부 층의 k번째 사업체의 전체 매출
$\overline{x}(i)_h$: i번째 항목에 대하여 h번째 세부 층의 전체 매출 합
$\overline{x^*}(i)_h$: i번째 항목에 대하여 h번째 세부 층의 데이터 관련 매출 합

단, 소속 분류의 응답률이 50% 미만인 경우, 상위분류의 비율을 활용하였다.
(데이터 관련 매출 합/전체 매출 합) 적용

□ 데이터직무 인력 수의 추정

○ 데이터직무 인력 수의 경우 매체 및 기업 관련 데이터를 통해 확인할 수 없는 정보이기 때문에 이전연도 응답 사업체가 아닌 경우 추정하는데 어려움 존재

○ 2023년 조사에 응답한 경우: 과거응답결과 활용

○ 2024년 조사에 응답한 경우: 동일 규모, 동일 중분류 응답 사업체들이 가지는 전체 종업원 수 대비 데이터직무 사업체의 인력 수 평균 비율을 적용. 즉, 무응답 사업체와 같은 중분류에 속하는 사업체 중 동일한 규모를 가진 사업체의 평균 데이터직무 인력 수 비율을 도출하여 적용

$$\text{평균 데이터직무 인력} = \frac{\text{데이터직무 인력수 (동일 중분류)}}{\text{전체 종사자수 (동일 중분류)}}$$

□ 단위무응답 사업체의 전체 매출액 및 종업원 수 추정

○ 2023년 자료, 전자공시시스템 등의 자료를 활용하여 대체. 전체 매출액 및 종업원 수 정보는 모집단 정비 과정에서 대부분 대체 가능하였음

9. 표본오차 추정 방법 및 결과(표본조사)

9-1 주요 항목에 대한 분산, 표준오차 등의 추정 방법

○ 업종별 데이터 매출, 종사자 인력 등의 총계 추정량, 분산 추정량 등은 다음과 같음

- 업종별 총계 추정량

총합 : $\widehat{Y}_g = \sum_{i=1}^{N_{gc}} y_{gci} + \sum_{k=1}^{4}\sum_{i=1}^{n_{gk}} w_{gk} y_{gki}$

평균: $\widehat{\overline{Y}}_g = \frac{1}{N_g}\widehat{Y}_g$

여기에서

$\widehat{Y}_g$	:	g업종 중분류의 총계추정치
y_{gci}	:	g업종 중분류의 전수층에 속하는 i 사업체의 특성
w_{gk}	:	g업종 중분류, 종사자 규모 k 표본층의 가중치
y_{gki}	:	g업종, 종사자 규모 k 표본층의 i 사업체의 특성
N_g	:	g업종에 속하는 모집단 사업체 수

- 업종별 평균 및 총합의 분산 추정량

총합 : $\hat{V}(\widehat{Y}_g) = \sum_{k=1}^{4} \hat{V}(\widehat{Y_{gk}})$

평균 : $\hat{V}(\widehat{\overline{Y}_g}) = \frac{1}{N_g^2}\hat{V}(\widehat{Y}_g) = \frac{1}{N_g^2}\sum_{k=1}^{4} \hat{V}(\widehat{Y_{gk}})$

- 상대표준오차(RSE)

$$RSE = \sqrt{\hat{V}(\widehat{Y}_g)} \,/\, \widehat{Y}_g \times 100$$

Ⅴ. 통계공표, 관리 및 이용자서비스

1. 공표통계 해석방법

1-1 주요 분류 수준별 세분화된 공표통계

□ 데이터산업 조사통계 분류수준

○ 공표되는 데이터산업 분류는 대분류(3개), 중분류(10개)임(시범조사 제외)

< '24년 데이터산업 조사 모집단 정의 >

대분류	중분류	내용
1. [데이터 솔루션] 데이터 처리 및 관리 솔루션 개발·공급업	**11. [데이터 수집]** 데이터 수집·연계 솔루션 개발·공급업	데이터 검색 솔루션, 데이터 수집 솔루션, 데이터 통합·연계 솔루션, 데이터 교환·개방 솔루션
	12. [DBMS] 데이터베이스관리시스템 솔루션 개발·공급업	관계형 데이터베이스 관리 솔루션, 비관계형 데이터베이스 관리 솔루션, 인메모리 데이터베이스 관리 솔루션, 기타 데이터베이스 솔루션
	13. [데이터 분석] 데이터 분석 솔루션 개발·공급업	정형·비정형 데이터 분석 솔루션, 실시간 데이터 분석 솔루션, 데이터 시각화 분석 솔루션, 데이터 처리 솔루션
	14. [데이터 관리] 데이터 관리 솔루션 개발·공급업	데이터 모델 관리 솔루션, 데이터 표준 관리 솔루션, 데이터 품질 관리 솔루션, 데이터베이스 운영·성능 관리 솔루션
	15. [데이터 보안] 데이터 보안 솔루션 개발·공급업	데이터 보안 솔루션, 개인정보보호 솔루션
	16. [데이터 플랫폼] 데이터 통합 플랫폼 솔루션 개발·공급업	데이터 플랫폼 솔루션
2. [데이터구축/컨설팅] 데이터 구축 및 컨설팅업	**21. [데이터 구축]** 데이터 구축·가공 서비스업	데이터베이스 설계·구축 서비스, 데이터 이행 서비스, 데이터 가공 서비스
	22. [데이터 컨설팅] 데이터 관련 컨설팅 서비스업	데이터 설계 컨설팅, 데이터 품질 컨설팅, 데이터베이스 성능개선 컨설팅, 데이터 거버넌스 컨설팅, 데이터 분석·활용 컨설팅
3. [데이터 서비스] 데이터 판매 및 제공 서비스업	**31. [데이터 거래]** 데이터 판매·중개 서비스업	데이터 판매 서비스, 데이터 중개 서비스, 데이터 분석 제공 서비스
	32. [정보제공] 정보제공 서비스업	포털·정보매개서비스, 기타 정보제공 서비스

* [ ] 안은 분류명 약어이며, 본 보고서에서는 데이터산업분류 약어로 사용

□ 일반산업 조사통계 분류수준

○ 공표되는 일반산업 분류는 공공, 의료, 교육, 농축산·광업 등(13개)임

< '24년 일반산업 조사 모집단 정의 >

구분	정의	관련 한국표준산업분류
공공	• 공공 행정, 국방 및 사회보장 행정에 해당되는 사업체 중 100명 이상 사업체	공공 행정, 국방 및 사회보장 행정(O84)
금융	• 금융업에 종사하는 사업체 중 100명 이상 사업체	금융업(K64), 보험 및 연금업(K65), 금융 및 보험관련 서비스업(K66)
제조	• 제조업에 해당되는 사업체 중 100명 이상 사업체	식료품 제조업(C10), 음료 제조업(C11), 담배 제조업(C12), 섬유제품 제조업; 의복 제외(C13), 의복, 의복 액세서리 및 모피제품 제조업(C14), 가죽, 가방 및 신발 제조업(C15), 목재 및 나무 제품 제조업; 가구 제외(C16), 펄프, 종이 및 종이제품 제조업(C17), 인쇄 및 기록매체 복제업(C18), 코크스, 연탄 및 석유정제품 제조업(C19), 화학 물질 및 화학제품 제조업; 의약품 제외(C20), 의료용 물질 및 의약품 제조업(C21), 고무 및 플라스틱제품 제조업(C22), 비금속 광물 제품 제조업(C23), 1차 금속 제조업(C24), 금속 가공제품 제조업; 기계 및 가구 제외(C25), 전자 부품, 컴퓨터, 영상, 음향 및 통신장비 제조업(C26), 의료, 정밀, 광학 기기 및 시계 제조업(C27), 전기장비 제조업(C28), 기타 기계 및 장비 제조업(C29), 자동차 및 트레일러 제조업(C30), 기타 운송장비 제조업(C31), 가구 제조업(C32), 기타 제품 제조업(C33), 산업용 기계 및 장비 수리업(C34)
유통	• 도매 및 소매업 중 100명 이상 사업체	자동차 및 부품 판매업(G45), 도매 및 상품 중개업(G46), 소매업; 자동차 제외(G47)
서비스	• 부동산업, 전문, 과학 및 기술 서비스, 사업시설 관리 및 지원, 예술, 스포츠 및 여가, 기타 서비스업 중 100명 이상 사업체	부동산업(L68), 연구개발업(M70), 전문 서비스업(M71), 건축 기술, 엔지니어링 및 기타 과학 기술 서비스업(M72), 기타 전문, 과학 및 기술 서비스업(M73), 사업시설 관리 및 조경 서비스업(N74), 사업 지원 서비스업(N75), 임대업; 부동산 제외(N76), 창작, 예술 및 여가관련 서비스업(R90), 스포츠 및 오락관련 서비스업(R91), 협회 및 단체(S94), 개인 및 소비용품 수리업(S95), 기타 개인 서비스업(S96)
의료	• 보건업 및 사회복지 서비스업으로 등록된 사업체 중 100명 이상 사업체	보건업(Q86), 사회복지 서비스업(Q87)
통신·미디어	• 정보서비스업으로 속하는 사업체 중 100명 이상 사업체 (데이터산업 모집단 제외)	출판업(J58), 영상・오디오 기록물 제작 및 배급업(J59), 방송업(J60), 우편 및 통신업(J61), 컴퓨터 프로그래밍, 시스템 통합 및 관리업(J62), 정보서비스업(J63)
물류	• 운수 및 창고업으로 등록된 사업체 중 100명 이상 사업체	육상 운송 및 파이프라인 운송업(H49), 수상 운송업(H50), 항공 운송업(H51), 창고 및 운송관련 서비스업(H52)
교육	• 교육 서비스업으로 등록된 사업체 중 100명 이상 사업체	교육 서비스업(P85)
유틸리티	• 전기, 가스, 증기, 수도, 하수, 폐기물 처리, 원료재생 및 환경복원업을 영위하는 사업체 중 100명 이상 사업체	전기, 가스, 증기 및 공기 조절 공급업(D35), 수도업(E36), 하수, 폐수 및 분뇨 처리업(E37), 폐기물 수집, 운반, 처리 및 원료 재생업(E38), 환경 정화 및 복원업(E39)
농축산·광업	• 농업, 임업, 축산업, 광업을 영위하고 있는 50명 이상 사업체	농업(A01), 임업(A02), 어업(A03), 석탄, 원유 및 천연가스 광업(B05), 금속 광업(B06), 비금속광물 광업; 연료용 제외(B07), 광업 지원 서비스업(B08)
건설업	• 건설업을 영위하고 있는 100명 이상 사업체	종합 건설업(F41), 전문직별 공사업(F42)
숙박·음식점업	• 숙박업 및 음식점업을 영위하는 100명 이상 사업체	숙박업(I55), 음식점 및 주점업(I56)

1-2 주요 통계표, 그래프

□ 주요 통계표

○ 국내 데이터산업 시장 규모는 '23년 29조 687억 원이며, '24년(E)은 30조 7,462억 원으로 추정됨

<데이터산업 시장 규모, 2020~2024년(E)>

(단위 : 억 원)

구분	2020년	2021년	2022년	2023년	2024년(E)	증감률 '23-'24(E)	CAGR '20-'24(E)
데이터 솔루션	25,133	29,843	41,318	49,108	55,343	12.7%	21.8%
데이터 구축/컨설팅	76,999	85,274	90,305	103,393	105,676	2.2%	8.2%
데이터 서비스	97,891	113,869	128,041	138,186	146,443	6.0%	10.6%
합계	**200,024**	**228,986**	**259,663**	**290,687**	**307,462**	**5.8%**	**11.3%**

○ 2024년 전 산업 데이터직무 인력은 22만 8,331명으로, 전년 대비 10.3% 증가함

<전 산업 데이터직무 인력 현황, 2020~2024년>

(단위: 명)

구분		2020년	2021년	2022년	2023년	2024년	증감률 '23-'24	CAGR '20-'24
데이터산업(A)	데이터 처리 및 관리 솔루션 개발·공급업	17,273	22,124	23,545	27,055	31,316	15.7%	16.0%
	데이터 구축 및 컨설팅 서비스업	48,644	58,733	64,248	67,072	71,306	6.3%	10.0%
	데이터 판매 및 제공 서비스업	36,050	41,574	46,517	46,309	49,683	7.3%	8.3%
	소계(명)	**101,967**	**122,431**	**134,310**	**140,435**	**152,305**	**8.5%**	**10.6%**
일반산업(B) (명)		48,210	58,442	63,492	66,592	76,027	14.2%	12.1%
전 산업(A+B) (명)		**150,177**	**180,873**	**197,802**	**207,027**	**228,331**	**10.3%**	**11.0%**

* 데이터직무 인력: 데이터 컨설턴트, 데이터 기획자, 데이터 아키텍트, 데이터 개발자, 데이터 엔지니어, 데이터베이스관리자, 데이터 분석가, 데이터 과학자 등 데이터 관련 업무를 수행하는 인력

1-3 공표되는 통계의 해석방법 및 이용 시 유의사항

○ 통계 이용 시 유의사항

○ 본 보고서는 「2024 데이터산업 현황조사」의 결과를 수록하였습니다.

○ 본 조사는 2024년 하반기에 실시되었으며, 매출액 조사기준 시점은 2023년 12월말(확정치)과 2024년 6월말(추정치), 인력 현황은 2024년 6월말 기준입니다.

○ 본 보고서에 수록된 2023년 시장규모는 확정치이며, 2024년 시장 규모는 추정치입니다.

○ 따라서 2024년 시장규모는 2025년 조사 실시 후에 보완될 수 있습니다.
(단, 데이터직무 인력 관련 통계 결과는 2024년까지 확정치입니다.)

○ 본 보고서에는 다음의 시범조사 결과를 함께 수록하고 있으며, 관련 통계 결과의 이용 시 주의가 필요합니다.

- 데이터인프라 서비스업 관련 비즈니스 영위기업 대상 통계 결과
- 일반산업 대상 통계 결과 일부(IT 및 데이터 관련 예산)

○ 통계결과는 반올림되어 부분의 합계가 전체와 일치하지 않을 수 있습니다.

○ 본 보고서의 통계표에 사용된 부호는 다음과 같습니다.

- 조사값 중 「-」는 해당 없음, 「0」은 해당 영역의 값이 0임을 의미

○ 본 보고서에 수록된 데이터산업 현황조사 결과 데이터는 국가통계포털(kosis.kr), ICT통계포털(www.itstat.go.kr)에서도 이용 가능합니다. 단, 시범조사 결과는 제외됩니다.

○ 본 보고서 내용과 관련한 문의사항은 한국데이터산업진흥원 산업기반본부 산업기획팀(02-3708-5361, 5365)으로 연락주시기 바랍니다.

○ 본 보고서의 내용을 인용할 때에는 반드시 과학기술정보통신부와 한국데이터산업진흥원 자료임을 밝혀야 합니다.

1-4 연도별(시계열) 통계결과 및 분석결과 관리

○ 2023년 일반산업 일부 부문 모집단 정의 변경

< 2023년 일반산업 모집단 정의 변경 내역 >

구분	변경전(2022년)	변경후(2023년)
공공	중앙정부기관, 지자체, 기획재정부의 공공기관 경영정보 시스템 (ALIO)에 등록된 준정부기관, 공기업 기타공공 기관, 부설기관과 지방자치단체 포함	공공 행정, 국방 및 사회보장 행정에 해당되는 사업체 중 100명 이상 사업체
의료	의료기관 중 50병상 이상 병원	표준산업분류 중 보건업 및 사회복지 서비스업으로 등록된 사업체 중 100명 이상 사업체
교육	교육 서비스업 지정 교육 기관 중 대학교육협의회에 등록된 대학교	표준산업분류 중 교육 서비스업으로 등록된 사업체 중 100명 이상 사업체

○ 데이터산업 부문은 승인통계 이후 시계열 단절 발생 없음('16년~'24년)

2. 조사대상 기간/조사 기준시점과 공표 시기

2-1 조사대상 기간/조사 기준시점과 통계공표 시점

○ 조사대상 기간: 2024년 11월~12월

○ 조사대상 시점: 매출액 2023년 12월말(확정치)/2024년 6월말(추정치), 인력현황

2024년 6월말(확정치)

○ 통계공표 시점

- 최종보고서 발표 : 2025년 5월

2-2 조사과정별 소요되는 기간

○ 조사수행 기간: 2024년 6월 ~ 12월

- 조사 설계 및 설문지 확정: 2024년 6월 ~ 10월
- 모집단 정비 및 표본설계 : 2024년 8월 ~ 10월
- 조사원 선발 교육 및 조사대상 컨택 : 2024년 11월
- 조사실시 : 2024년 11월 ~ 12월
- 자료검증 및 입력: 2024년 12월 ~ 2025년 1월
- 결과분석 및 보고서 작성: 2025년 1월 ~ 3월

2-3 조사 기준시점과 통계결과의 최초 공표일 간의 차이

○ 2024년 조사 기준시점: 2024년 6월말

○ 2024년 조사 공표시기: 2025년 5월말

2-4 통계작성 및 공표 기간 단축 가능성

○ 통계작성 기준시점부터 공표시점까지 차이 : 약 1년 5개월

○ 매년말~차년도 세부시행계획 수립 시 공표 일정 확정됨

○ 본 조사는 과기정통부 ICT통합모집단을 활용하도록 지정되어 있으나, 매년 상반기 모집단의 재정비로 인해 하반기 이후 활용 가능하므로 통계 시점 단축은 어려운 상황임

○ 이에, 매출액, 직무인력 등 조사수행 당해년도 추정치를 함께 조사/공개하여 이용자가 참고할 수 있도록 제시조사년도 추정치를 함께 발표하여 국가전략수립을 위한 정책통계 활용도와 시의성을 높이고자 노력함

3. 공표일정

3-1 사전에 공개된 통계공표 일정과 공개방법

○ 공표일 : 매년 5월

○ 공표방법

- 한국데이터산업진흥원 홈페이지(www.kdata.or.kr) / 보고서 공표('25년 5월)
- 국가통계포털 홈페이지(www.kosis.kr) / 주요 결과 공표('25.5~'25.8)
- 과학기술정보통신부 통계 홈페이지(www.itstat.go.kr) / 마이크로데이터 공표(~'25.12)

4. 통계 작성방법의 비교성

4-1 통계의 개념 동일 여부

○ 데이터산업 현황조사에 대한 조사통계의 개념은 매년 동일하게 적용됨

4-2 분류체계 동일 여부

○ 데이터산업 분류체계는 자체분류체계로 운영하고 있으며 매년 동일한 기준으로 조사실시 및 중분류 수준에서 결과 공표

- 중분류 체계는 2019년까지 일부 변동이 있었으나 포괄하고 있는 정의는 동일함

5. 국가간 비교성

5-1 동일한 목적을 갖는 외국 통계 명칭과 개요

○ 자료명: The European Data Market Study 보고서

○ 진행기관: IDC & The Lisbon Council

○ 주요내용: 데이터 시장의 범위를 데이터를 가공함으로써 제품 및 서비스로 재생산되는 '디지털 데이터 시장'으로 정의. 이미지 데이터처럼 디지털 기술을 통해 수집, 저장, 가공, 전송되는 멀티미디어 데이터 부문도 포함하여 본 통계와 차이가 있음

5-2 동일한 목적을 갖는 외국 통계와의 직접 비교 가능 여부, 가능하지 않은 경우 사유 및 이용 시 고려사항 등에 대한 검토

○ 데이터 시장에 대한 정의가 다르며, 시계열 문제로 수치적인 절대 비교는 불가

○ 연도별/국가별 증감 추이 수준에서 제한적인 비교 가능

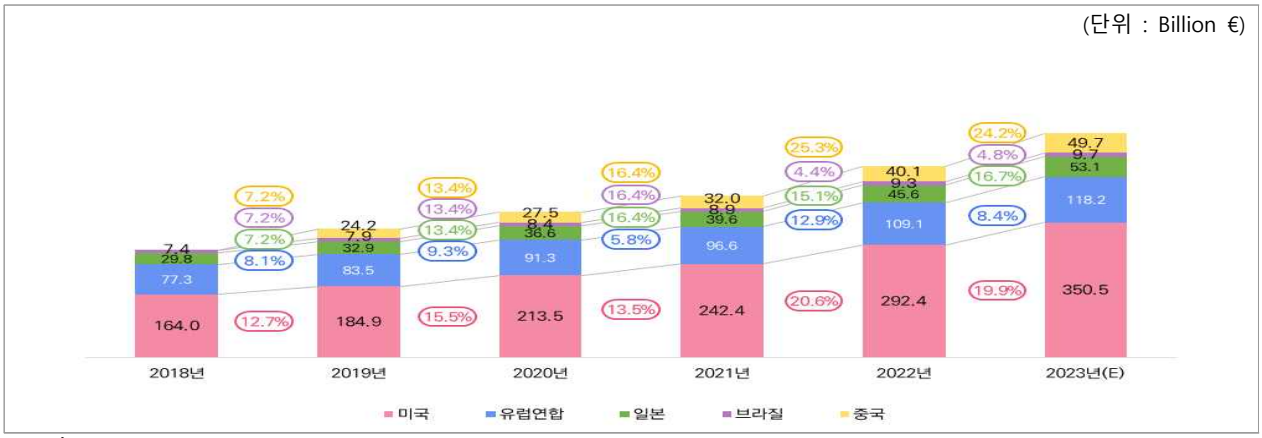
5-3 국가 간 주요 통계 내용을 비교한 통계표, 그래프

< 2018~2023(E) 주요국 데이터산업 시장규모 >

(단위 : Billion €, %)

구 분	2018년		2019년		2020년		2021년		2022년		2023년(E)	
	시장규모	증감률	시장규모	증감률	시장규모	증감률	시장규모	증감률	시장규모	증감률	시장규모	증감률
미국	164.0	11.6%	184.9	12.7%	213.5	15.5%	242.4	13.5%	292.4	20.6%	350.5	19.9%
유럽연합	77.3	12.3%	83.5	8.1%	91.3	9.3%	96.6	5.8%	109.1	12.9%	118.2	8.4%
일본	29.8	11.5%	32.9	10.5%	36.6	11.3%	39.6	8.2%	45.6	15.1%	53.1	16.7%
브라질	7.4	5.4%	7.9	7.2%	8.4	5.9%	8.9	6.0%	9.3	4.4%	9.7	4.8%
중국	-	-	24.2	-	27.5	13.4%	32.0	16.4%	40.1	25.3%	49.7	24.2%
전체	**278.5**	**11.6%**	**333.5**	**19.7%**	**377.3**	**13.1%**	**419.5**	**11.2%**	**496.5**	**18.4%**	**581.2**	**17.1%**

< 2018~2023(E) 주요국 데이터산업 시장규모 그림 >



* 자료: European DATA Market Study 2021–2023 D2.7 FINAL SECOND REPORT ON FACTS AND FIGURES (IDC & The Lisbon Council, 2024.2.)

6. 유사통계와 일관성

6-1 동일하거나 유사한 통계의 명칭과 개요

○ 해당없음

6-2 공표항목과 타통계 유사·동일 항목간 통계 수치 비교·검토

○ 해당없음

7. 통계의 이용자 서비스

7-1 통계공표 방법 다양화

○ 통계 제공 경로 및 내용

- 한국데이터산업진흥원 홈페이지(www.kdata.or.kr) : 보고서
- 과학기술정보통신부 홈페이지(www.msit.go.kr) : 보고서
- 국가통계포털 홈페이지(www.kosis.kr) : 주요결과 데이터
- 마이크로데이터 통합서비스(mdis.kostat.go.kr) : 마이크로데이터

7-2 국가통계포털(KOSIS) 자료 제공 여부

○ 국가통계포털(KOSIS) 수록: DB형식으로 제공

- 경로: 국내통계→주제별 통계→정보통신→데이터산업현황조사
- 경로: 국내통계→기관별 통계→과학기술정보통신부→데이터산업현황조사

8. 통계 설명자료 제공

8-1 통계설명자료에 대한 소재 정보

○ 통계설명자료(메타정보, 방법론 보고서, 품질보고서 등) 제공 경로

- 국가데이터처 통계설명자료서비스(www.k-stat.go.kr/metasvc)에 메타 정보 제공
- 국가데이터처 자료실(kostat.go.kr/board.es?mid=a10409060100&bid=67)에 데이터산업 현황조사에 대한 정기통계품질진단 보고서 제공
- 마이크로데이터 통합서비스(mdis.kostat.go.kr; MDIS)에서 마이크로데이터 신청 가능

9. 마이크로데이터 생성 · 관리

9-1 마이크로데이터 생성 방법

○ 원시자료(rawdata) 수집

○ 원시자료를 토대로 조사항목 간 논리관계, 작성된 통계표과 보고서와의 이상유무 확인

○ 응답사업체가 식별되지 않도록 민감데이터 마스킹 처리

○ 마이크로데이터 자료 생성

9-2 마이크로데이터 관리 방법

○ 조사기관 1차 생성·내검 후 제출 ⇄ 한국데이터산업진흥원(K-DATA) 2차 확인 및 보완, 제출 ⇄ 과학기술정보통신부 지정 정보통신·방송통계협의회(위탁기관:KISDI) 검토·관리

10. 마이크로데이터 서비스

10-1 마이크로데이터 제공여부

○ 국가데이터처 마이크로데이터 통합서비스(MDIS)에서 제공

10-2 마이크로데이터에 대한 설명자료

○ 마이크로데이터 통합서비스(mdis.kostat.go.kr) 통해 다운로드 제공

○ 제공일시 : 공표 및 품질점검 완료된 이후 해당사이트에서 다운로드 가능

○ 제공범위 : 승인통계 대상 자료 일체를 엑셀 형식으로 제공(원시자료, 코드설계서)

- 식별 가능성을 배제하기 위해 세부항목은 제공하지 않음

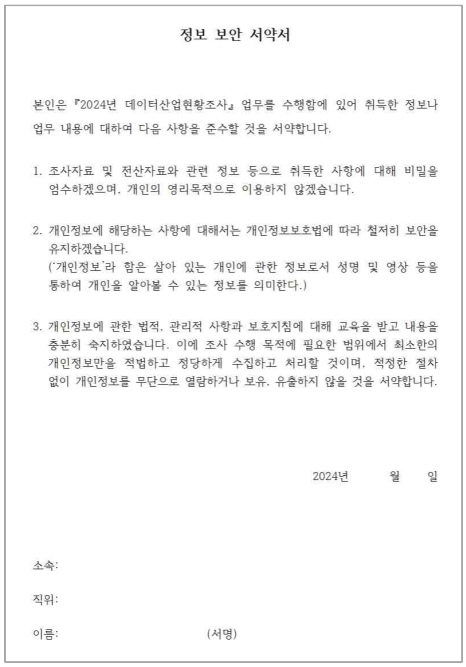
○ 제공대상 : 누구나 가능

○ 비용: 별도 구입 비용 없음

11. 자료 수집, 처리 및 보관 과정의 비밀보호

11-1 자료 수집과정에서 응답자 비밀보호를 위한 지침(법령, 규정)이나 조치

○ (조사원) 보안서약서 작성 의무화



정보 보안 서약서

본인은 『2024년 데이터산업현황조사』 업무를 수행함에 있어 취득한 정보나 업무 내용에 대하여 다음 사항을 준수할 것을 서약합니다.

1. 조사자료 및 전산자료와 관련 정보 등으로 취득한 사항에 대해 비밀을 엄수하겠으며, 개인의 영리목적으로 이용하지 않겠습니다.

2. 개인정보에 해당하는 사항에 대해서는 개인정보보호법에 따라 철저히 보안을 유지하겠습니다.
('개인정보'라 함은 살아 있는 개인에 관한 정보로서 성명 및 영상 등을 통하여 개인을 알아볼 수 있는 정보를 의미한다.)

3. 개인정보에 관한 법적, 관리적 사항과 보호지침에 대해 교육을 받고 내용을 충분히 숙지하였습니다. 이에 조사 수행 목적에 필요한 범위에서 최소한의 개인정보만을 적법하고 정당하게 수집하고 처리할 것이며, 적정한 절차 없이 개인정보를 무단으로 열람하거나 보유, 유출하지 않을 것을 서약합니다.

2024년 월 일

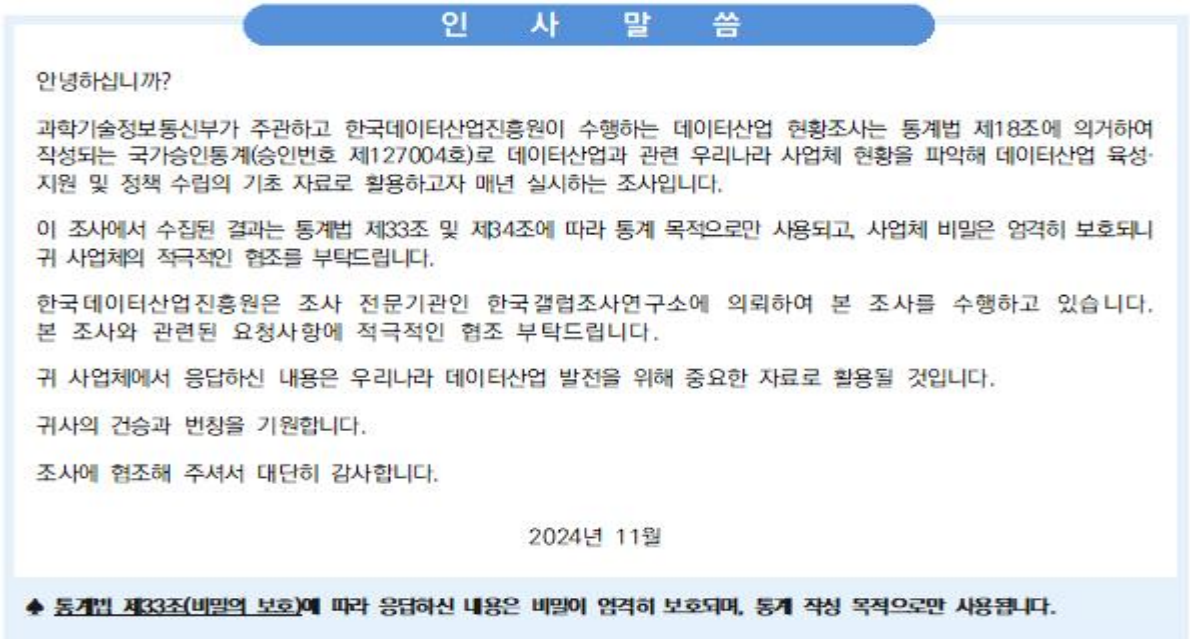
소속:

직위:

이름: (서명)

○ (응답자 보호) 주민등록번호 등 개인정보 미수집

○ (응답자 보호) 통계법 제33조에 의한 응답자 비밀보호



인 사 말 씀

안녕하십니까?

과학기술정보통신부가 주관하고 한국데이터산업진흥원이 수행하는 데이터산업 현황조사는 통계법 제18조에 의거하여 작성되는 국가승인통계(승인번호 제127004호)로 데이터산업과 관련 우리나라 사업체 현황을 파악해 데이터산업 육성·지원 및 정책 수립의 기초 자료로 활용하고자 매년 실시하는 조사입니다.

이 조사에서 수집된 결과는 통계법 제33조 및 제34조에 따라 통계 목적으로만 사용되고, 사업체 비밀은 엄격히 보호되니 귀 사업체의 적극적인 협조를 부탁드립니다.

한국데이터산업진흥원은 조사 전문기관인 한국갤럽조사연구소에 의뢰하여 본 조사를 수행하고 있습니다. 본 조사와 관련된 요청사항에 적극적인 협조 부탁드립니다.

귀 사업체에서 응답하신 내용은 우리나라 데이터산업 발전을 위해 중요한 자료로 활용될 것입니다.

귀사의 건승과 번창을 기원합니다.

조사에 협조해 주셔서 대단히 감사합니다.

2024년 11월

♠ 통계법 제33조(비밀의 보호)에 따라 응답하신 내용은 비밀이 엄격히 보호되며, 통계 작성 목적으로만 사용됩니다.

11-2 자료 처리과정에서 응답자 비밀보호를 위한 지침(법령, 규정)이나 조치

○ 본조사의 대다수가 온라인, TAPI 조사로 진행하므로 종이 조사표는 거의 존재하지 않음. 혹시나 응답 받은 종이 조사표는 조사 종료 후 1년 내 전수 폐기 원칙하고, 부득이한 경우 암호화하여 보관

○ 데이터는 사업체별 무작위 일련번호를 부여하고, 사업체 식별이 가능한 변수를 제외하여 공개용 데이터로 별도 관리, 보관용 데이터는 암호화하여 보관

○ (조사완료 후 자료 처리) 근거자료 등 보관이 필요한 자료 외에 조사 도중 작성된 자료 등은 조사 완료 후 전문업체에 의뢰하여 파쇄

○ (금지사항) 개인 USB 사용 금지, 개인 e-mail/카페 접속 금지, 퇴근 시 서류 일체를 캐비넷에 넣어 보관하며 개방된 곳에 보관 금지

11-3 자료 보관과정에서 응답자 비밀보호를 위한 지침(법령, 규정)이나 조치

○ 공표자료는 평균이나 분포값 등 통계값 제공을 위해 특정사업체와 응답값을 나타내지 않도록 함

○ 마이크로데이터는 사업체별 무작위 일련번호를 부여하고, 사업체 식별이 가능한 변수를 제외(사업체명, 주소, 전화번호 등)한 공개용 데이터로 제공

12. 공표자료의 비밀보호

12-1 공표자료에서 응답자 비밀보호를 위하여 취한 조치나 방법

○ 보안교육 수시 실시

○ 수집단계에서 주민등록번호 미수집

○ 컴퓨터 전원차단 및 암호 관리 준수

○ 식별 가능 조사항목(연락처, 응답자 명 등)은 내검 및 사후검증에서만 사용

○ 공표자료는 응답사업체 비식별화 조치 후 제공하므로 비밀이 보호됨

- 모든 통계자료는 산업 중분류 단위까지 한하여 제공

12-2 마이크로데이터 제공 과정에서 응답자 비밀 보호를 위하여 취한 조치나 방법

○ 응답 사업체 무작위 일련번호 부여

○ 사업체 식별이 가능한 데이터(기업명, 사업자등록번호, 주소, 전화번호 등 기업 관련 정보 일체)는 모두 삭제 처리

○ 통계 항목별 이상치를 점검하여 필요 시 마스킹 처리

○ DB 접근권한은 전담팀에게만 부여하고 해킹, 바이러스 대비 폐쇄망 서버에 조사 데이터 및 관련 자료 일체 보관

○ 결재과정 및 결재완료된 문서가 타 부서에 노출되지 않도록 전산권한을 조정

○ 랜섬웨어 등의 피해를 막기 위해 본건에 대한 조사 결과물 및 일체의 자료들은

실사기관의 폐쇄망 서버에 백업 관리되며, 일정 기간이 지나면 폐기함

13. 자료 보안 및 접근제한

13-1 자료 유실, 유출, 훼손 등을 예방하기 위한 자료보안 관련 지침(법령, 규정)이나 조치

○ 과업수행자의 보안서약서 제출 및 보안관리 책임자 지정

○ 참여 연구인력

- 본 조사를 수행하는 연구 인력과 관련하여 비밀 보호 서약서를 작성하여 자료의 외부 유출이 없도록 관리하고, 본 용역을 수행하는 인력 외에는 자료 접근 및 이용을 금지함

○ 조사원

- 본 조사를 수행하는 연구 인력과 관련하여 비밀보호 서약서를 작성하여 자료의 외부 유출이 없도록 관리하고, 본 용역을 수행하는 인력 외에는 자료 접근 및 이용을 금지함

○ 전문 보안 인력 및 프로그램 활용

- 정보 보호 책임자 및 정보 관리 처리 담당자 지정 운용
- 모든 PC는 보안 프로그램(오피스키퍼)를 적용하여 업무 자료에 대해 외부 저장매체나 메일 등을 통한 반출이 통제되며, 기록이 남음
- 개인 민감정보 등을 포함한 자료의 경우 암호화하여 관리하며, 보유 기간 이후에는 보안프로그램을 이용하여 복원이 불가능하도록 삭제함

○ 사업과정에서 생성된 자료는 발주기관 승인 없이 외부 공개 금지

○ 사업 수행과정에서 취득한 정보나 자료는 사업 수행 중은 물론 사업 완료 후에도 이를 외부에 유출할 수 없음(사업종료 시 폐기, 반납)

Ⅵ. 통계기반 및 개선

1. 기획 및 분석 인력, 사업예산

1-1 통계 기획 및 분석인력

○ 과학기술정보통신부 데이터진흥과 통계작성기관 : 한국데이터산업진흥원 위탁

- 위탁업무: 조사기획, 자료 처리, 통계의 공표 및 자료검수
- 민간위탁: 조사계획 수립, 자료 수집, 처리, 분석 등

2. 통계 품질관리 및 개선

2-1 통계품질제고 가능성에 대한 검토 결과나 개선계획 또는 추진 실적

○ 통계기획자문위원회 개최

- 데이터산업 및 데이터직무 분류체계, 조사 문항 점검 등 조사 고도화를 주제로 5~7인 내외 자문위원회를 구성하여 매년 2~3회 개최

2-2 전주기 통계품질진단 결과에 따른 개선과제 이행내역 및 이행률(중점관리과제)

○ 2020년 정기통계품질진단 개선과제 4건 중 4건 이행(이행률 100%)

< 전주기 정기통계품질진단 결과에 따른 개선과제 현황 >

개선과제명	세부 개선과제	이행여부	이행내용
표본관리 강화	- 각 층별 목표표본수 미달시 대처방안에 대한 기준 명시 - 작성기관과 위탁기관의 지속적인 관리·감독 필요	이행	- 응답 거부(및 비논리적 응답)의 경우 이차적으로는 응답 업체에 재연락을 취해 오기입 및 무응답에 대해 질의하여 관련 응답을 수정(확보)하였으며, 이에 재차 응답을 거부한 경우 대체 표본을 활용하여 보완 조사를 실시하였음. - 표본 대체가 불가능한 전수층(300인 이상 그룹)은 무응답 추정 방법(해당 기업의 전자공시 자료, 이전연도 응답, 기업 관련 정보 수집 등)에 따라 추정값으로 대체하였음 * 2021년 보고서 32~38페이지 참고 - '19년 대비 '21년의 종사자 규모에 따른 표본 회수율은 하기와 같이 개선되었음 · 50인~99인 76.8% => 104.6% · 100~299인 41.6% => 97.0% · 300인 이상 27.1% => 56.4%
해외 유사통계 비교	- 해외 유사 통계에 대한 비교표 작성	이행	- EU집행위원회의 '유럽 데이터 시장 연구 보고서(European Data Market Study)'의 데이터기업 포함 범위와 주요국 데이터산업 시장규모를 수록하여, 국내 데이터산업 현황과 비교 가능 *2021년 보고서 19, 65~66페이지 참고
무응답 개선	- 조사거절 사유 분석을 통한 각 기관별 접근방법 필요 - 작성기관과 위탁기관의 지속적인 관리·감독 필요	이행	- 항목 무응답의 경우 해당 표본(기업)에 직접 전화를 걸어 응답을 확보하거나, 해당 기업의 다른 담당자를 컨택하는 등 응답 현황 관리를 통해 최대한 관련 응답을 확보하였음. - 단위 무응답의 경우 응답 대체값 처리를 위해 외부 기업정보 DB 및 공시정보 등을 추가로 활용해 최신정보 수집 * 2021년 보고서 36~38페이지 참고 - '19년 대비 '21년 단위무응답 규모는 89건에서 54건으로 감소
마이크로데이터 제공방법 변경	- 마이크로데이터 통합서비스를 통한 공공용 다운로드서비스 제공	이행	- 기존 과기정통부에서 관리하던 ITSTAT을 포함하여 국가데이터처 마이크로데이터 통합서비스에도 마이크로데이터 제공