

# 문화·관광·콘텐츠 데이터 공동활용 기반 구축 연구

2021-07

정책연구

A Study on the Building of  
the Joint Utilization of Culture, Tourism,  
and Content Data

송철재  
김수경  
장선녀  
김정림



한국문화관광연구원  
Korea Culture & Tourism Institute



# 문화·관광·콘텐츠 데이터 공동활용 기반 구축 연구

A Study on the Building of the Joint Utilization of Culture,  
Tourism, and Content Data

송철재·김수경·장선녀·김정림



한국문화관광연구원  
Korea Culture & Tourism Institute



## 연구책임

---

송철재 한국문화관광연구원 책임전문위원

## 공동연구

---

김수경 한국문화관광연구원 책임전문위원

장선녀 한국문화관광연구원 차석전문위원

김정림 한국문화관광연구원 차석전문위원



문화·관광·콘텐츠 데이터 공동활용 기반 구축 연구



## 연구개요





# 1. 서론

## 가. 연구 배경 및 목적

### 1) 연구 배경

- 데이터기반행정 활성화를 위하여 데이터 공동활용에 대한 요구 증가
- 문체부 소속 및 산하기관별 폐쇄적·독점적 데이터 관리
- 문화·관광·콘텐츠 공공데이터 공동활용 기반의 확충 필요

### 2) 연구 목적

- 문화체육관광부 산하 소속·산하기관을 대상으로 공공데이터 현황조사, 보유 및 수요 데이터 전수조사, 데이터 업무 활용 실태조사를 실시
- 데이터 현황조사 결과를 수집·정리하여 기관별 데이터 보유 현황 및 분석업무에 필요한 타 기관 데이터 수요를 파악하여 데이터기반행정 활성화의 추진기반 마련
- 데이터기반행정 공동활용 활성화 방안 마련

## 나. 연구 범위 및 방법

### 1) 연구 범위

- 대상적 범위: 문체부 산하 소속·공공기관 고유의 행정업무 수행을 통해 생성, 취득하여 보유 또는 운영·관리하는 모든 데이터를 대상
- 시간적 범위: 연구진행을 위한 시간적 범위는 기관별로 보유하고 있는 데이터 유형, 데이터 최초 작성시점, 작성주기, 시계열 연속성 등이 상이함을 고려하여 시간적 범위를 특정시점으로 한정하지는 않음
- 내용적 범위: 데이터기반행정 추진여건 및 환경변화 분석, 공공데이터 공동활용 사례분석, 문체부 산하 소속·공공기관 보유 및 활용 데이터 현황조사, 조사결과 활용 방안 마련, 데이터기반행정 추진기반 및 역량 강화 방안 제시

## 2) 연구 방법

- 정부의 최근 데이터정책 기초, 데이터3법, 공공데이터법, 데이터기반행정 활성화에 관한 법률안 등 거시적 환경분석
- 국내외 연구보고서, 학술논문, 정책자료 등을 기초로 기관별로 산재되어 있는 공공데이터 활용 사례분석
- 공공데이터 현황조사, 보유 및 수요 데이터 조사, 데이터 업무 활용 실태조사
- 문체부 산하 소속·공공기관별 데이터 담당자에 대한 심층인터뷰 및 관련 전문가 자문

## 2. 데이터기반행정 추진여건 및 환경분석

### 가. 데이터기반행정의 의의

- 데이터기반행정법을 토대로 데이터 및 데이터기반행정 등 주요 용어를 정의함
  - (데이터의 정의) 기계 판독이 가능한 형태로 존재하는 정형 또는 비정형의 정보를 의미함(데이터기반행정법 제2조제1호)
  - (데이터기반행정의 정의) 공공기관이 생성하거나 다른 공공기관 등으로부터 취득하여 관리하고 있는 데이터를 수집·저장·가공·분석·표현하는 등의 방법으로 정책 수립 및 의사결정에 활용함으로써 객관적이고 과학적으로 수행하는 행정(데이터기반행정법 제2조제2호)
- 데이터기반행정 이전의 데이터 관련 정책과 제도 및 주요 특징을 비교함

구분	정의	특징	한계
데이터 기반 행정	데이터 기반의 객관적·과학적 행정체계를 구축하기 위한 제도적 기반	사회·환경변화를 대응해 행정의 책임성, 대응성, 신뢰성 향상	데이터기반행정의 범위가 포괄적이기 때문에 기능과 목적이 겹칠 수 있는 기존의 법 제도들과의 중복성이 있음
공공 데이터	정부, 지자체, 공공기관이 전자적으로 생성 또는 취득하여 관리하고 있는 모든 전자화된 파일(DB)	다양한 공공기관이 생성 및 관리해왔던 공공데이터의 통합적 활용을 통해 데이터기반행정의 주요 재원의 역할	부처별로 데이터 공유가 미진할 경우, 데이터가 구조화되지 않을 경우 활용도가 떨어짐
전자 정부 제도	행정기관 및 공공기관의 업무에 정보 기술을 활용, 전자화하여 행정업무를 효율적으로 수행하는 정부	행정의 생산성과 투명성, 민주성이 제고를 통해 대국민 행정서비스의 질적 향상 도모	하드웨어에 치중, 중복투자와 공동활용 미흡으로 인해 비효율성 및 통합서비스의 어려움
증거 기반 정책	과학적 증거에 기반을 두어 정책을 결정하려는 시도	신뢰성과 타당성 있는 정책 과정을 통해 생성된 정보와 지식이 축적되어 더 나은 정책 형성 가능	과학적 증거 부재로 인한 적용범위의 제한성, 담당자의 이해부족에 따른 기술적 해석의 한계

## 나. 데이터기반행정 정책환경 분석

- 국내외 데이터 환경 변화를 중심으로 데이터기반행정의 정책환경을 분석함
  - (데이터 제반 환경) 데이터 시장의 성장과 함께 데이터 경제의 가속화가 진행, 정부 주도의 데이터경제로의 전환을 위한 정책적 대응체계 마련
  - (데이터 공동활용 요구 증대) 2020년 코로나19 확산에 따른 위기상황 극복, 사회현안 해결에 대한 선제적 대응 등을 위해 정부의 맞춤형 서비스의 요구와 필요 증대
- 데이터기반행정 활성화를 위한 관련 제도 현황을 파악하여 시사점 도출
  - (법 제도) 데이터기반행정법과 밀접한 관련성이 있는 주요 법(공공데이터법, 전자정부법, 데이터 3법 등)의 특징을 파악하고, 데이터기반 활성화에 어떠한 영향이 있는지 분석
  - (관련 계획) 공공기관의 데이터 연계를 위한 정부의 주요 계획을 비교·분석함
  - (데이터 관련 조직) 데이터포털 사이트 운영 현황 및 조직도 상의 명칭 등을 통해 데이터 업무 수행 가능한 정부부처 및 공공기관 현황 파악하고, 데이터기반 활성화를 위한 조직 정비 등에 대해 제언함

## 다. 데이터 공동활용 관련 선행연구 분석

- 데이터기반행정법과 국내외 사례 연구를 통해 데이터 연계와 공동활용 개념 정의
  - (데이터 연계) 서로 다른 복수의 데이터 파일을 결합하여 보다 풍부한 정보를 제공해 줄 수 있는 통합 데이터를 만드는 방법을 뜻함(오미애, 2015)
  - (데이터 공동활용) 데이터 연계 가능한 공공데이터를 선별하여 기관 간 공유하고, 데이터 통합기반을 구축하여 기존의 데이터 가치 이상의 인사이트를 도출함
- 데이터기반행정법 상의 데이터 연계 관련 이슈 정리
  - (데이터 수집 및 관리) 원활한 데이터 연계를 위한 기관 간 데이터 제공 협조체계 구축 필요
  - (데이터 연계의 필요성) 데이터의 단순한 연계는 실제 정책결정에 활용 불가능하므로, 데이터 연계 시 목적성에 대한 지속적인 고민이 필요

### 3. 데이터 공동활용 사례 분석

#### 가. 사례 분석 방법

- 국내사례는 정부부처 및 지방자치단체가 실시한 주요 사업 중 데이터기반행정의 우수사례로 평가받은 사례를 중심으로 살펴봄
- 해외사례는 데이터기반행정의 모범적인 사례로 평가받고 있는 영국, 스웨덴, 네덜란드를 중심으로 정책환경 및 주요 사업 추진 사례를 살펴봄

#### 나. 국내 사례

- 행정안전부·경찰청·인천광역시: 주요 데이터를 공동활용한 무질서·범죄 위험도 예측모델 구축하여 ‘112 신고데이터 기반 범죄 위험도 분석’
- 교육부·보건복지부·국토교통부·창원시·청주시: 아파트 시공 전 초등학생 돌봄 수요를 예측해 이에 부합하는 ‘다함께돌봄센터’를 설치하는 ‘초등돌봄 수요예측 모델’ 사업을 실시
- 병무청·법무부·경찰청 등 빅데이터 분석을 통한 병역면탈 의심자 포착
- 서울특별시, 심야 전용 시내버스 ‘올빼미 버스’

#### 다. 해외 사례

- 영국, ECHILD 데이터 세트
  - 행정데이터 연구(ARD UK), 통계청(ONS), 유니버시티 칼리지 런던(UCL), 국가 의료서비스 디지털(NHS Digital), 교육부(ED)의 합동 프로젝트임
  - 데이터베이스는 1995년 9월 1일에서 2020년 8월 31일 사이에 태어난 영국의 모든 어린이 및 청소년(25세 이하)약 1,600만 명의 20년 이상의 교육과 사회복지, 그리고 건강정보를 포함
- 스웨덴의 LISA(노동시장 장기통합 데이터)
  - LISA는 사회보험청의 건강보험 데이터와 노동부의 노동시장 통합 데이터임

- 데이터는 등록 기반 노동 시장 통계(RAMS), 총인구 등록부(RTB), 소득세 등록부(IoT), 인구 및 주택 센서스 등 공동 활용함
- 네덜란드, VIVET Programme
  - 2019년 네덜란드 통계청(CBS: Statistics Netherlands), 네덜란드 기업청(RVO.nl) 및 네덜란드 환경평가청(PBL: Netherlands Assessment Agency), 토지등록부(Land Registry) 등의 합동프로젝트
  - 등록 행정데이터와 빅데이터, 서베이자료 등 필수적인 수집자료와 함께 기술적인 자료들이 포함

## 라. 시사점

- 국내사례의 경우 주요 이슈와 관련된 활용사업을 개발하는 과정에서 데이터 통합이나 연계 없이 추진하는 단계별로 필요한 통계 수치 및 데이터를 활용하여 사업 및 정책을 추진하였으며, 이후 데이터의 재사용 또는 개방이 소극적임
- 해외사례의 경우 데이터 개방성이 높은 나라로 각국의 복수의 정부부처가 협력하여 각자 보유한 행정 데이터를 통합 또는 연계한 데이터 세트를 만들었음
- 우리나라의 경우 데이터기반행정 활성화 및 데이터 공동활용을 적극 추진할 수 있는 동력이 부족한 상황임
- 향후 단일조직 또는 다수의 정부 부처 관계자로 구성된 공식 위원회를 구성하여 데이터 공동활용을 적극 추진할 수 있는 동력을 확보해야 함

## 4. 데이터 현황조사

### 가. 데이터 현황조사 수행 체계

- 문화체육관광부 소속 및 산하기관이 보유하고 있는 데이터 소재 및 현황을 파악하기 위한 전제로서, 데이터 대한 명확한 개념 정의가 사전에 이루어져야 하며, 조사 대상 데이터에 대한 정확한 조작적 정의가 제시되어야 함

- 데이터 현황조사는 공공데이터 현황조사, 보유 데이터 전수조사, 수요 데이터 전수조사, 데이터 업무 활용 실태조사 이상의 4가지 조사로 구성
- 데이터 현황조사를 수행하는 궁극적인 목적은 문화체육관광 분야 공공데이터의 정확한 소재 및 현황을 파악하고 나아가서는 공동활용 데이터 발굴·관리를 통한 문체부 소속 및 산하기관 간에 공동활용 데이터 이용 활성화 방안을 마련하고자 하는 것임

## 나. 공공데이터 현황조사

- 공공데이터 현황조사는 행정안전부 메타데이터관리시스템, 공공데이터포털과 한국문화정보원 문화데이터광장 등 문체부 소속 및 산하기관의 데이터를 통합적으로 관리·운영하고 있는 공공데이터 정보 파악을 목적으로 진행
- 개인정보가 마스킹 처리된 필수 메타데이터 정보만을 입수하여 공공데이터 현황조사 진행 및 결과분석
- 문화체육관광부 소속 및 산하기관이 공공데이터 포털에 개방한 데이터를 살펴보면, 본부 및 소속기관 15곳에서 총 161개의 공공데이터를 개방하고 있었는데 그 중 파일데이터는 112개, 오픈API는 49개를 차지하는 것으로 나타남

## 다. 보유 및 수요 데이터 전수조사

- 기관별 데이터 보유 현황 및 분석업무에 필요한 타 기관 데이터 수요 파악을 목적으로 보유 및 수요 데이터 전수조사 진행
- 문체부 소속 및 산하기관 21곳에서 총 452개의 데이터를 보유 또는 운영·관리하고 있었으며, 데이터 생성·취득 방법별로는 2차가공 208개(46.0%), 직접조사 147개(32.5%), 행정자료 97개(21.5%) 순으로 나타남
- 문체부 소속 및 산하기관 8곳에서 총 49개 데이터를 업무 수행 시 필요한 타 기관 데이터로 응답하였으며, 공동활용 정책분야별로는 “특화된 대책 및 맞춤형 서비스(32개)” 분야에 활용할 수 있는 데이터에 대한 수요가 가장 많은 것으로 나타남

## 라. 데이터 업무 활용 실태조사

- 데이터 공동활용의 제약조건 및 관계 분석을 위해 데이터 업무 활용 관련 경험 가능성이 높은 문체부 및 소속·산하기관 데이터 및 정책업무 담당자와 연구원을 대상으로 설문조사 실시
- 데이터 및 정책업무 담당자들은 담당 업무를 수행할 때 데이터를 많이 활용하고 있으며, 데이터기반행정 활성화를 위해서는 인력, 예산, 장비·시설의 확충이 중요하다고 생각하고 있었음
- 또한 데이터기반행정 강화를 위한 데이터 공동활용의 중요성은 높게 생각했으나 실현 정도는 잘 이뤄지지 않고 있는 것으로 조사됨
- 데이터 공동활용의 제약조건 및 관계 분석을 위해 데이터 업무 활용 관련 경험 가능성이 높은 정책업무 담당자와 연구원을 대상으로 설문조사 실시
- 데이터기반행정 실태를 조사한 결과, 주로 활용하는 데이터 유형은 ‘통계조사 자료’(78.1%), 데이터 형식은 ‘정형 데이터’(46.4%), 데이터 구성은 ‘단일 데이터’(63.5%)가 높게 나타남
- 데이터기반행정 인식을 조사한 결과, 담당 업무 수행 시 데이터 활용도는 ‘활용한다’(약간 활용한다 + 매우 많이 활용한다)(75.5%)[정책기획](81.8%), [정책집행](70.8%), [정책평가](75.5%) 업무에서도 역시 ‘활용한다’(약간 활용한다 + 매우 많이 활용한다)가 높게 나타남
- 데이터 활용 실태를 조사한 결과, 데이터 공동활용 정책 및 사업 참여 경험자는 12.0%로 주 업무는 ‘데이터 분석’(65.2%)이 가장 많았으며, 사업 기간은 평균 12.5개월이며, 예산은 평균 59,501.7만 원, 평균 투입 인력은 12.5명으로 나타남
- 데이터 공동활용 인식을 조사한 결과, 데이터기반행정 활성화를 위해 필요한 과제를 묻은 결과, ‘중요하다’(중요하다 + 매우 중요하다)는 응답이 [데이터 관련 예산 확보](89.1%), [데이터 관련 부서 간 협력](87.0%), [데이터 관련 전문인력 확보](86.5%) 등의 순으로 높게 나타남
- 전반적으로 문화체육관광부 소속 및 산하기관 데이터 및 정책업무 담당자들은 담

당 업무를 수행할 때 데이터를 많이 활용하고 있으며, 데이터기반행정 활성화를 위해서는 인력, 예산, 장비·시설의 확충이 중요하다고 생각하고 있음

## 마. 조사결과 활용

### 1) 데이터 소재 및 현황 파악

- 문화체육관광부 소속 및 산하기관이 보유 또는 생성·취득하여 운영·관리하고 있는 데이터 총괄 현황(공공데이터 현황조사+데이터 전수조사)을 분석함
- 데이터 총괄 현황을 크게 ‘정부기능분류 기준’과 ‘문화체육관광부 조직분류 기준’으로 구분하여 살펴볼 수 있으며, 데이터기반행정법에서 제시하고 있는 5가지의 정책영역을 각각의 분류기준별 교차하여 문체부 소속 및 산하기관이 보유하고 있는 데이터 소재 및 현황을 파악할 수 있음
- 문화체육관광 분야 데이터 1,364개 중 예술정책 데이터가 284개(20.8%)로 가장 많았고, 다음으로 문화정책 데이터가 274개(20.1%), 미디어정책 179개(13.1%), 등의 순으로 데이터를 보유 또는 운영·관리하고 있는 것으로 나타남

### 2) 공동활용 데이터 발굴

- 데이터 현황조사 결과를 활용하여 데이터기반행정법에 명시된 정책분야에 부합하는 공동활용 데이터 목록화 작업을 진행
- 행정안전부가 권고하고 있는 공동활용 정책분야에 포함되는 데이터로서 기관 간 공동 활용성이 높다고 판단되는 보유 데이터를 공동활용 목록화 작업 대상으로 설정
- 범정부 차원의 데이터기반행정 활성화를 위한 지정등록 데이터 대응 및 문체부 차원의 자체 분석수요에 맞는 기초자료로 활용할 수 있음

### 3) 정책분야별 분석활용(안) 도출

- 데이터현황조사 결과를 기반으로 수집·정리된 문체부 소속 및 산하기관에서 분산 축적·관리되고 있는 다양한 유형의 데이터 정보 및 공동활용 가능 데이터로 목록



화된 데이터를 기준으로 정책분야별 분석활용 과제(안)을 도출

- 문화체육관광 분야 기획과제 및 기관 간 협업 과제 등으로 발굴될 수 있는 활용(안)을 분석주제, 분석 배경 및 목적, 공동활용 대상 데이터, 분석 설명 및 방법, 기대 효과 등으로 구체화하여 12개 분석과제(안)를 주요 정책분야별로 제시

## 5. 데이터기반행정 공동활용 활성화 방안 마련

### 가. 제도적 측면

- 데이터기반행정법이 시행됨에 따라 데이터기반행정을 수행하기 위해서는 기관의 특성에 맞추어 데이터의 수집·활용 계획의 수립여부, 적절성 등을 사전에 확인하는 제도 도입 등 관련 업무 규정이나 기타 제도 등의 정비가 필요
- 데이터의 공동활용 및 데이터기반행정을 제한하는 장애요소와 법적 제약사항 등을 꾸준히 정비하고 개선해나가야 함
- 문화체육관광부 내부적으로 데이터 공동활용에 대한 강제성을 가지는 규정을 추가하거나 제도 및 업무절차를 개선해야할 필요가 있음

### 나. 기술적 측면

- 데이터 표준화는 데이터의 정보 요소의 명칭과 정의, 형식, 규칙 등에 대한 원칙을 수립하여 지속적으로 관리하기 위한 활동
- 데이터의 품질을 보장하기 위한 기본 활동인 데이터 표준화는 개별 기관이 가진 데이터의 구조와 기준 등은 기관의 특성에 따라 구성되어 있기 때문에 이를 통합 관리하기 위해서는 중앙으로부터의 표준화 방안 수립이 필수
- 부처 차원에서 정책수립 또는 의사결정자의 수요를 파악하여 해당 부처의 소속 및 소관 공공기관 등을 대상으로 데이터를 수집한다면 공동활용 데이터의 등록이 좀 더 명확하게 이루어질 것임

### 다. 데이터 활용역량 강화

- 데이터기반행정의 안착과 실제 해당 제도를 통해 운영되는 공공기관의 데이터 역량 강화는 매우 중요한 사안임
- 전사적인 차원에서 기관 내 종사자들의 데이터에 대한 이해 제고 차원의 데이터 인식제고 교육을 시행하는 방안을 우선 추진이 필요
- 데이터 인식강화를 위한 전사적인 교육은 ① 기관 내부 교육으로 진행하며, ② 교육 방식은 기관의 여건을 고려하여 온라인 교육과 오프라인 교육을 병행하고, ③ 교육 방법은 강의 중심으로 진행하되, 기관의 제반 현황에 따라 실습과 토론 교육을 실시함

## 6. 결론 및 제언

### 가. 결론

- 본 연구는 문체부 소속 및 산하기관에서 분산 축적·관리되고 있는 다양한 유형의 데이터 정보를 파악하고, 데이터 공동활용 기반을 마련하였다는 점에서 의의가 있음
- 기관 간, 부서 간 보유하고 있는 데이터를 공동활용할 수 있는 데이터기반의 행정 문화를 조성하고 확산해 나갈 필요가 있음
- 지속적인 데이터기반의 행정추진 가운데 각 기관에 데이터 역량과 데이터기반 문화가 조성됨으로써 객관적이며 과학적인 행정 및 의사결정이 가능해질 것으로 봄

### 나. 정책 제언

- 데이터 공동활용 활성화 방안의 구체적인 역할 및 권한을 제도화·지침화해야 함
- 문화체육관광부의 소속 및 산하기관 내에서는 개인정보에 대한 비밀유지 계약을 체결하여 데이터를 제공하는 등의 개인정보보호 문제를 해결해야할 필요가 있음

- 데이터 공동활용 측면에서 보다 활용도 높은 데이터 정보를 확인하기 위해서는 관련 법제도를 근거로 공동활용할 필요가 있는 데이터를 발굴·관리해 나가야 함
- 데이터 공동활용을 제한하는 여러 요소 중 하나인 개인정보와 관련하여 강력한 데이터 공동활용을 위하여 개인정보보호 문제 해결 필요
- 데이터 문제해결능력을 포함하여 데이터 전문성을 지닌 데이터 전문인력을 확보하고, 기존 인력 중 데이터분석 업무의 관심과 역량을 지닌 인력의 배치 전환을 통하여 전문인력 확보
- 기존 인력의 전문성을 제고하기 위하여 외부 교육기관을 통한 외부교육과 기관 내부적으로 이루어지는 내부 교육 등을 추진
- 데이터기반행정 문화의 확산을 위하여 단계적 로드맵 마련 및 데이터 역량을 보유한 리더의 배치, 부서 간 협력적 조직문화 형성 필요
- 기 구축 데이터에 대한 표준화를 위하여 중앙 연계 시 변환하여 수집하며, 장기적 관점에서 표준화를 추진하되 정부차원의 컨설팅 등 지원 필요
- 데이터를 생성·관리·활용 등의 모든 과정에 대하여 데이터의 품질관리 추진



---

# 목차

<b>제1장 서론 .....</b>	<b>1</b>
제1절 연구 배경 및 목적	3
1. 연구 배경	3
2. 연구 목적	4
제2절 연구 범위 및 방법	5
1. 연구 범위	5
2. 연구 방법	6
3. 연구 수행절차	8
제3절 연구 기대효과	9
 <b>제2장 데이터기반행정 추진여건 및 환경 분석 .....</b>	 <b>11</b>
제1절 데이터기반행정의 의의	13
1. 데이터의 정의	13
2. 데이터기반행정의 개념	16
3. 데이터기반행정과 기존의 데이터 관련 제도와의 관계	17
제2절 데이터기반행정 정책환경 분석	23
1. 데이터기반행정 정책환경	23
2. 데이터기반행정 관련 제도 현황	25
제3절 데이터 공동활용 관련 선행연구 분석	33
1. 데이터 연계의 개념 및 의의	33
2. 데이터 공동활용 개념	39
3. 데이터 공동활용 선행연구	41
제4절 소결	47

<b>제3장 데이터 공동활용 사례 분석 .....</b>	<b>51</b>
제1절 사례 분석 방법	53
제2절 국내 사례	54
1. 경찰청, 112 신고데이터 기반 범죄 위험도 분석	55
2. 교육부·보건복지부 등, 신규 공동주택 초등돌봄시설 수요예측	59
3. 병무청, 병역면탈 의심자 포착	62
4. 서울특별시, 심야 전용 시내버스 ‘올빼미 버스’	65
제3절 해외 사례	68
1. 영국, ECHILD 데이터 세트	69
2. 스웨덴, 노동시장 장기통합 데이터(LISA)	71
3. 네덜란드, VIVET Programme	74
제4절 소결	78
 <b>제4장 데이터 현황조사 .....</b>	 <b>81</b>
제1절 데이터 현황조사 수행 체계	83
1. 데이터 현황조사의 역할과 의미	83
2. 주요 원칙	84
3. 데이터 현황조사의 기본 방향	85
제2절 공공데이터 현황조사	88
1. 조사개요	88
2. 조사결과	90
제3절 보유 및 수요 데이터 전수조사	96
1. 조사개요	96
2. 조사결과	104
제4절 데이터 업무 활용 실태조사	115
1. 조사개요	115
2. 조사결과	119
제5절 조사결과 종합 및 활용	140
1. 조사결과 종합	140
2. 조사결과 활용	145
3. 조사결과 활용상의 유의점	177

<b>제5장 데이터 공동활용 활성화 방안 마련 .....</b>	<b>179</b>
제1절 데이터 관리 및 공동활용 기반 마련	181
1. 제도적 측면	181
2. 기술적 측면	188
제2절 데이터 공동활용 활성화	199
1. 데이터 활용역량 강화	199
2. 데이터기반행정 활성화 문화 조성	210
3. 공동활용 제약 요소 발굴 및 개선사항 마련	211
제3절 소결	213
 <b>제6장 결론 및 제언 .....</b>	 <b>215</b>
제1절 결론	217
제2절 정책 제언	222
 <b>참고문헌 /</b>	 <b>225</b>
 <b>ABSTRACT /</b>	 <b>229</b>
 <b>부록 데이터 업무 활용 실태조사 조사표 /</b>	 <b>231</b>

---

# 표 목차

〈표 1-1〉 대상적 범위 기준	5
〈표 2-1〉 「데이터기반행정법(제2조제1호)」 데이터 정의	13
〈표 2-2〉 「공공데이터법(제2조제1호)」 공공데이터 정의	14
〈표 2-3〉 「문화체육관광부 데이터 관리 규정」 빅데이터 정의	14
〈표 2-4〉 데이터와 공공데이터의 차이 비교	15
〈표 2-5〉 문화체육관광 분야 데이터 정의	15
〈표 2-6〉 데이터기반행정의 정의	16
〈표 2-7〉 데이터기반행정법과 공공데이터법의 차이 비교	18
〈표 2-8〉 데이터기반행정과 증거기반 비교	20
〈표 2-9〉 데이터기반행정 관련 개념의 정의와 특징	21
〈표 2-10〉 주요국 데이터 전략 현황	23
〈표 2-11〉 데이터 3법의 개정 주요 내용	28
〈표 2-12〉 공공데이터 연계 관련 정부 주요 계획	29
〈표 2-13〉 데이터기반행정 활성화 위원회에 참여 기관	32
〈표 2-14〉 데이터 연계 개념 정의 관련 국내외 선행연구	34
〈표 2-15〉 데이터 공동활용 촉진	40
〈표 2-16〉 공공기관의 공동활용 데이터 등록 활성화	40
〈표 2-17〉 개인정보 비식별 가이드라인과의 차이점	42
〈표 2-18〉 신혼부부 신용DB 구축 시 사용한 주요 데이터	43
〈표 2-19〉 신혼부부 횡단분석 및 종단분석에 따른 주요 결과	44
〈표 2-20〉 데이터기반행정 관련 개념의 정의와 특징	48
〈표 2-21〉 연구 분석틀 설정	50
〈표 3-1〉 사례분석 방법	53
〈표 3-2〉 국내 데이터공동활용 주요 사례	54
〈표 3-3〉 공동주택 구조·환경 및 인구 관련 데이터 목록(1)	60
〈표 3-4〉 공동주택 구조·환경 및 인구 관련 데이터 목록(2)	60
〈표 3-5〉 입력변수 및 출력 예시	60



〈표 3-6〉 가계동향조사 가구유형별 맞벌이 비율	60
〈표 3-7〉 병역면탈 의심자 포착 관련 데이터 목록	63
〈표 3-8〉 해외 데이터공동활용 주요 사례	68
〈표 4-1〉 데이터 현황조사 수행 체계	87
〈표 4-2〉 조사 제외 대상 예시	88
〈표 4-3〉 조사 수행 절차	89
〈표 4-4〉 문체부 소속·산하기관 데이터 시스템 구축·운영현황	90
〈표 4-5〉 문체부 소속·산하기관 공공데이터 포털 개방 현황	90
〈표 4-6〉 문체부 소속기관 공공데이터 포털 개방 현황	91
〈표 4-7〉 문체부 산하기관 공공데이터 포털 개방 현황	92
〈표 4-8〉 문화체육관광 분야 공공데이터 포털 개방 현황	93
〈표 4-9〉 문화정보통합시스템 데이터 수집·관리 현황(1)	94
〈표 4-10〉 문화정보통합시스템 데이터 수집·관리 현황(2)	95
〈표 4-11〉 조사 설계 개요	97
〈표 4-12〉 조사대상 데이터	97
〈표 4-13〉 주요 용어 정의	98
〈표 4-14〉 조사절차	99
〈표 4-15〉 보유 데이터 전수조사 양식	100
〈표 4-16〉 수요 데이터 전수조사 양식	101
〈표 4-17〉 조사표 작성 참고 자료(공동활용 정책분야)	102
〈표 4-18〉 조사표 작성 참고 자료(제공 예외조건)	102
〈표 4-19〉 조사표 작성 참고 자료(정부기능분류체계*, BRM)	102
〈표 4-20〉 정부기능분류체계 대기능 재분류	105
〈표 4-21〉 보유 및 수요 데이터 전수조사 응답(참여) 기관	106
〈표 4-22〉 문체부 소속·산하기관 생성·취득 방법별 보유 데이터 현황	106
〈표 4-23〉 성·취득 방법별 보유 데이터 현황	107
〈표 4-24〉 정부기능분류별(대기능) 보유 데이터 현황	107
〈표 4-25〉 정부기능분류별(대기능) 생성·취득 방법별 보유 데이터 현황	108
〈표 4-26〉 공동활용 정책분야별 정부기능분류별(정책영역) 보유 데이터 현황	109
〈표 4-27〉 정부기능분류별(대기능) 공동활용 정책분야별 보유 데이터 현황	109
〈표 4-28〉 정책분야별 공동활용 가능 수요 데이터 현황	110
〈표 4-29〉 정책분야별 공동활용 가능 수요 데이터 세부 현황	111
〈표 4-30〉 조사 설계 개요	116

〈표 4-31〉 조사 항목	117
〈표 4-32〉 응답자 특성	118
〈표 4-33〉 공공 데이터 보유 주체	134
〈표 4-34〉 데이터 공동활용 사업명	135
〈표 4-35〉 데이터 공동활용 활성화를 위해 필요한 부분	138
〈표 4-36〉 담당 업무 수행 시 데이터 활용 중요도 및 활용도	143
〈표 4-37〉 데이터기반행정 활성화를 위해 필요한 과제	143
〈표 4-38〉 데이터기반행정 활성화를 위해 필요한 과제	144
〈표 4-39〉 조사결과 활용 체계	146
〈표 4-40〉 문체부 소속·산하기관 보유 데이터 현황(전수+공공)	148
〈표 4-41〉 정부기능분류별(대기능) 보유 데이터 현황(전수+공공)	148
〈표 4-42〉 공동활용 정책분야별 정부기능분류(정책영역)별 보유 데이터 현황(전수+공공)	149
〈표 4-43〉 공동활용 정책분야별 정부기능분류(대기능)별 보유 데이터 현황(전수+공공)	150
〈표 4-44〉 문화체육관광부 조직 분류와 정부기능분류 연계표	151
〈표 4-45〉 문화체육관광부 조직 분류 기준 보유 데이터 현황(전수+공공)	152
〈표 4-46〉 공동활용 정책분야별 문화체육관광부 조직 분류(분야)별 보유 데이터 현황 (전수+공공)	153
〈표 4-47〉 공동활용 정책분야별 문화체육관광부 조직 분류(대분류)별 보유 데이터 현황 (전수+공공)	154
〈표 4-48〉 공동활용 가능 데이터 활용 분야	155
〈표 4-49〉 정책분야별 공동활용 가능 데이터 목록(문체부 소속 및 산하기관 보유)	156
〈표 4-50〉 정책분야별 공동활용 가능 데이터 목록(타 기관 보유)	157
〈표 4-51〉 공동활용 가능 데이터 활용 분야	158
〈표 4-52〉 정책분야별 분석활용(안)	159
〈표 4-53〉 전문가 자문회의 및 FGI 의견 요약	162
〈표 5-1〉 문화체육관광부의 법·제도 정비	182
〈표 5-2〉 문화체육관광부 소속·산하기관 데이터 관리 및 분석을 위한 부서 운영 현황	185
〈표 5-3〉 표준화의 법적 근거	188
〈표 5-4〉 공공데이터의 제공 및 이용 활성화에 관한 법률 시행령 제16조	197
〈표 5-5〉 데이터 라이프 사이클에 따른 데이터 역량(대분류)	200
〈표 5-6〉 데이터 역량지표 체계	201
〈표 5-7〉 데이터 역량지표별 역량강화 추진방안(예시(안))	202
〈표 5-8〉 데이터 역량강화를 위하여 보완되어야 할 주요 요건	204

〈표 5-9〉 데이터 역량강화를 위한 교육(안)	206
〈표 5-10〉 데이터기반행정법 제24조(데이터 관련 전문인력 양성)	206
〈표 5-11〉 문화체육관광부 데이터 관리 규정 제8조(교육 및 전문인력 양성)	207
〈표 5-12〉 정책 라이프사이클 빅데이터 활용(예시)	210

---

# 그림 목차

[그림 1-1] 연구의 목적	4
[그림 1-2] 연구 수행절차	8
[그림 2-1] 데이터기반행정의 개념적 배경 및 특징	22
[그림 2-2] 비식별화 단계	42
[그림 2-3] 신혼부부 신용DB 자료연계 구성항목	43
[그림 3-1] 경찰청의 112신고 군집분석 결과	56
[그림 3-2] 데이터 기반 범죄위험도 예측	57
[그림 3-3] 인천시 △△구 ○○동 범죄위험도 5점 척도 예측 결과(예시)	58
[그림 3-4] 지역별 초등돌봄 수요예측 사례	61
[그림 3-5] 병역면탈 의심자 포착 관련 분석방안 및 분석결과	64
[그림 3-6] 서울시 심야버스 분석 프로그램 구조도	65
[그림 3-7] 서울시 심야버스 심층분석도	66
[그림 3-8] ECHILD 데이터베이스 생성을 위한 연계 프로세스	70
[그림 3-9] VIVET Programme의 통계 프로세스	75
[그림 3-10] VIVET Programme의 데이터 공동활용 개념도	76
[그림 4-1] 데이터 기본 유형	119
[그림 4-2] 단일 데이터 유형	120
[그림 4-3] 공공/민간 데이터 결합 유형	120
[그림 4-4] 공공 데이터 제공 기관	121
[그림 4-5] 공공 데이터 정책분야	121
[그림 4-6] 타 기관 공공 데이터 이용 경로	122
[그림 4-7] 타 기관의 데이터 제공 원활 정도	122
[그림 4-8] 타 기관의 데이터 제공이 원활하지 않았던 이유	123
[그림 4-9] 타 기관의 데이터 제공 시 소요 기간	123
[그림 4-10] 내부/외부 기관에 데이터 제공 경험	124
[그림 4-11] 외부 기관에 대한 데이터 제공 원활 정도	124
[그림 4-12] 외부 기관에 대한 데이터 제공이 원활하지 않은 이유	125

[그림 4-13] 외부 기관에 데이터 제공 시 소요 기간	125
[그림 4-14] 데이터 교육 경험	126
[그림 4-15] 데이터 교육 내용	127
[그림 4-16] 보유 데이터 자격증 개수 및 항목	127
[그림 4-17] 담당 업무 수행 시 데이터 활용 중요도	128
[그림 4-18] 업무 유형별 데이터 활용 중요도_정책기획	128
[그림 4-19] 담당업무 수행 시 데이터 활용 정도	129
[그림 4-20] 업무 유형별 데이터 활용 정도	129
[그림 4-21] 데이터기반행정을 위한 과제들의 중요도	130
[그림 4-22] 데이터기반행정을 위한 과제들의 실현 정도	131
[그림 4-23] 데이터 공동활용 정책 및 사업 참여 경험	132
[그림 4-24] 사업목적	132
[그림 4-25] 공공 데이터 공동활용 경험	133
[그림 4-26] 민간 데이터 공동활용 경험	133
[그림 4-27] 공동활용 정책분야	135
[그림 4-28] 데이터 확보 채널	136
[그림 4-29] 데이터기반행정 강화를 위한 데이터 공동활용 중요도	136
[그림 4-30] 데이터기반행정 강화를 위한 데이터 공동활용 실현 정도	137
[그림 4-31] 데이터 공동활용 제약조건의 심각성	138
[그림 4-32] 데이터 교육 경험	142
[그림 5-1] 정책정보센터 조직도	185
[그림 5-2] 데이터 역량강화 4단계 절차	201
[그림 5-3] 빅데이터 아카데미 과정	209



문화·관광·콘텐츠 데이터 공동활용 기반 구축 연구

제1장

서론





# 제1절 연구 배경 및 목적

---

## 1. 연구 배경

증거기반(evidence-based), 데이터기반(data-based), 통계기반(statistics-based) 등의 의미는 어떻게 표현하든 결론은 “정부·자치단체·공공기관의 정책결정 혹은 의사결정은 결정자의 주관적 의견에 기반(opinion-based)하지 말고 증거, 데이터, 통계 등 객관적 근거에 기반 해야 한다”는 것이다(정성호 외, 2020). 미국, EU 등 주요국 등이 지난 2019년 이후 국가적 차원의 데이터 전략을 수립하여 적극적으로 이러한 움직임에 대응하고 있는 가운데, 우리나라도 데이터에 기반한 과학적 행정 구현을 위해 「데이터기반행정 활성화에 관한 법률(이하 ‘데이터기반행정법’)」을 제정(20.6.9.) 하였고, 2020년 12월 10일부터 시행에 들어갔다.

「데이터기반행정법」은 △기관 간 데이터 공동활용 △데이터기반행정 활성화를 위한 제도 확립 △데이터기반의 일하는 방식 혁신 등의 추진전략을 통해 데이터에 기반한 과학적 행정을 구현하고자 하는 취지의 법이다. 동 법에서는 데이터 공동활용을 기반으로 과학적인 행정을 구현하고 이를 통해 공공기관의 책임성 및 신뢰성을 높이고 국민의 삶의 질을 향상시키는 것을 법률이 만들어진 궁극적인 목적으로 제시하고 있다(데이터기반행정법 제1조).

이처럼 공공기관이 수집·보유하고 있는 데이터를 주고받고 기관 간의 데이터를 공동 활용하는 것이 법제화되었기 때문에, 중앙부처 및 공공기관 등은 당장 무엇을 주고받을 지에 대한 고민이 시작되었다(정성호 외, 2020). 문화예술, 콘텐츠, 관광, 스포츠 분야 관련 통계·데이터를 생산하고 관리·운영하고 있는 문체부 산하 소속·공공기관도 마찬가지이다. 문체부 산하 소속·공공기관의 데이터 관리에 많은 예산이 투입됨에 따라 방대한 공공데이터가 축적되고 있으나, 최근까지도 기관별 폐쇄적·독점적 운영과 부처 차원의 전사적인 데이터 공유체계 부족으로 데이터를 공동활용 할 수 있는 근거 및 제도가 부재한 상황이다. 따라서 앞으로는 「데이터기반행정법」 시행을 맞아 기관 간 칸막이 제거

를 통해 타 기관 데이터를 요청·제공하고 공동활용하여 데이터에 기반한 과학적인 행정 추진 방안을 모색하는 것이 중요한 시점이다.

나아가서는 기존에 업무관행, 과거경험 등에 의한 일률적인 행정 추진이 아닌, 객관성·신뢰성이 담보된 과학적 행정을 추진하여 국민의 행복과 삶의 질을 개선하기 위한 기반을 조성하는 것이 중요한 과제로 떠오르고 있다.

## 2. 연구 목적

이에 본 연구는 다음과 같은 목적을 가진다.

첫째, 문화체육관광부 산하 소속·산하기관을 대상으로 공공데이터 현황조사, 보유 및 수요 데이터 전수조사, 데이터 업무 활용 실태조사를 실시한다.

둘째, 데이터 현황조사 결과를 수집·정리하여 기관별 데이터 보유 현황 및 분석업무에 필요한 타 기관 데이터 수요를 파악하여 데이터기반행정 활성화의 추진기반을 마련한다.

셋째, 데이터기반행정 공동활용 기반 구축을 목적으로 ①데이터 관리 및 공동활용 기반 마련 ②데이터 활용역량 강화 ③데이터기반행정 활성화 문화 조성 ④기관 간 데이터 공동활용을 제약하는 장애요소 발굴 및 개선사항을 마련하고자 한다.

[그림 1-1] 연구의 목적



## 제2절 연구 범위 및 방법

### 1. 연구 범위

#### 가. 대상적 범위

본 연구의 대상적 범위는 문체부 산하 소속 ·공공기관 고유의 행정업무 수행을 통해 생성, 취득하여 보유 또는 운영·관리하는 모든 데이터를 대상으로 하되, 아래의 조건을 모두 충족하는 데이터를 대상적 범위로 설정하였다.

〈표 1-1〉 대상적 범위 기준

- 
- ▶ 기관별로 보유 또는 운영·관리하고 있으나, 외부에 개방(공개)하지 않은 데이터\*\*
    - 정책수립, 집행과정 모니터링, 정책평가 등 과정에서 필요할 것으로 예상되는 데이터
    - 기관 간 공동활용 데이터로써 활용성이 높다고 판단되는 보유 데이터
    - \* (외부 개방 데이터) 메타데이터관리시스템, 공공데이터포털 등에서 기 개방(공개)된 데이터
    - \*\* (외부에 개방하지 않은 데이터) 외부 개방 데이터를 제외한 시스템을 통해 개방(공개)하지 않은 데이터를 연구대상으로 함. 단, ①기관 홈페이지 및 자체 정보시스템에서만 공개한 데이터 ②전체 데이터 중 일부만 외부에 개방(공개)된 데이터(ex. 개인정보 포함으로 일부만 개방)는 연구대상에 포함
  - ▶ 각 기관의 업무 수행 시 활용되고 있는 데이터 및 필요할 것으로 예상되는 타 기관 데이터
    - 기존 업무·분석 수행 시 이미 활용하고 있는 타 기관 데이터
    - 기존 업무·분석 수행 시 필요한 데이터가 있었으나 데이터 보유기관으로부터 제공이 되지 않았던 데이터
    - 정책수립, 집행과정 모니터링, 정책평가 등 과정에서 필요할 것으로 예상되는 타 공공기관 데이터
    - 공공기관 간 공동활용 데이터로써 활용성이 높다고 판단되는 타 공공기관 보유 데이터
- 

#### 나. 시간적 범위

연구진행을 위한 시간적 범위는 기관별로 보유하고 있는 데이터 유형, 데이터 최초 작성시점, 작성주기, 시계열 연속성 등이 상이함을 고려하여 시간적 범위를 특정시점으로 한정하지는 않는다. 다만, 기관별 보유 데이터 전수조사를 실시한 이후 취합·정리된 데이터 현황 분석을 통해 시간적 범위를 재정의 하도록 한다.

## 다. 내용적 범위

본 연구의 범위는 첫째, 이론적 논의 및 선행연구 분석을 통한 데이터기반행정 추진여건 및 환경변화 진단, 둘째, 공공데이터 공동활용 사례분석, 셋째, 문체부 산하 소속·공공기관 보유 및 활용 데이터 현황조사, 넷째, 조사결과 활용 방안 마련, 다섯째, 데이터 공동활용 활성화 방안 제시 등이다. 이에 따른 구체적인 연구의 내용은 다음과 같다.

먼저, 데이터기반행정 추진여건 및 환경분석에서는 데이터기반행정의 이론적 의의를 검토하고 공공데이터 공동활용 관련 선행연구들을 살펴본다. 공공데이터 공동 활용 사례분석에서는 공공데이터 공동활용 관련 국내외 사례를 분석하고 시사점을 제시한다. 보유 및 활용 데이터 현황조사에서는 먼저 기관별로 보유 또는 운영·관리되고 있으나, 외부에 개방(공개)되지 않은 데이터를 전수조사하고, 이어서 각 기관에서 업무 수행 시 실제 활용되는 데이터 현황 및 필요할 것으로 예상되는 타 기관 데이터 수요조사를 진행한다. 조사결과 활용에서는 조사결과를 기반으로 문화체육관광부 소속 및 산하기관이 보유 또는 운영·관리하고 있는 데이터 소재 및 현황을 파악하고, 이어서 공동활용 가능 데이터를 목록화 하여 정책분야별 데이터 분석활용 과제(안)를 제시한다. 마지막으로 데이터 공동활용 활성화 방안 제시에서는 제도측면, 기술측면, 역량측면의 기반 마련 및 강화 방안을 제시한다.

## 2. 연구 방법

### 가. 문헌분석

정부의 최근 데이터정책 기초, 데이터3법, 공공데이터법, 데이터기반행정 활성화에 관한 법률안 등 거시적 환경분석을 실시하였다. 정부의 데이터 관련 최근 정책분석과 타 부처와 비교한 문체부의 추진현황을 검토하였고, 데이터 수집·등록·연계·활용 및 공공데이터 공동활용 관련 선행연구 검토를 수행하였다.

### 나. 사례분석

국내외 연구보고서, 학술논문, 정책자료 등을 기초로 기관별로 산재되어 있는 공공데이터 활용 사례분석을 실시하였고, 이를 기반으로 시사점 도출하였다.

#### **다. 설문조사**

외부에 개방되지 않은 데이터를 조사대상으로 설정하고 관련 정책수립 및 평가에 필요함과 동시에 기관 간 공동활용 가능성 등이 높을 것으로 판단되는 데이터를 중심으로 보유데이터 전수조사를 실시하였다. 더불어, 기관별 업무 수행 시 실제 활용되고 있는 타 기관 데이터 현황을 조사하고 필요한 데이터가 있었으나, 데이터 보유기관으로부터 제공이 되지 않았던 데이터 수요조사 진행을 병행하여 공동활용 데이터 발굴에 필요한 기초 데이터를 수집하였다.

#### **라. 관계자 심층인터뷰**

문체부 산하 소속·공공기관별 데이터 담당자에 대한 심층인터뷰를 통해 업무 수행 시 필요한 데이터 수요를 파악하고 기관 간 공동활용을 제약하는 장애요소 및 개선 필요사항 등에 대한 수요와 요구를 청취하였다.

#### **마. 전문가 자문**

데이터기반행정법 시행으로 인한 거시적 환경변화 및 일하는 방식에 미친 영향, 데이터 현황조사 활용 방안 검토, 데이터기반행정법에 따른 공동활용 데이터 목록화 방안, 정책분야별 분석과제(안) 도출에 대한 의견, 향후 데이터 공동활용 방안 모색에 전문가에 대한 자문을 수행하였다.

### 3. 연구 수행절차

본 연구는 크게 I. 서론 - II. 데이터기반행정 추진여건 및 환경분석 -III. 데이터 공동활용 사례분석 - IV. 데이터 현황조사 -V. 데이터 공동활용 활성화 방안 마련 - VI. 결론 및 제언으로 총 6단계로 구분하여 연구를 수행할 계획이며, 본 연구의 수행절차는 다음 [그림 1-2]와 같다.

[그림 1-2] 연구 수행절차

I. 서론	▶ 제1절 연구 배경 및 목적 제2절 연구 범위 및 방법 제3절 연구 기대효과
II. 데이터기반행정 추진여건 및 환경분석	▶ 제1절 데이터기반행정의 의의 제2절 데이터기반행정 정책환경 분석 제3절 데이터 공동활용 관련 선행연구 분석 제4절 소결
III. 데이터 공동활용 사례 분석	▶ 제1절 사례분석 방법 제2절 국내 사례 제3절 해외 사례 제4절 소결
IV. 데이터 현황조사	▶ 제1절 데이터 현황조사 추진체계 제2절 공공데이터 현황조사 제3절 보유 및 수요 데이터 전수조사 제4절 데이터 업무 활용 실태조사 제5절 조사결과 종합 및 활용
V. 데이터 공동활용 활성화 방안 마련	▶ 제1절 데이터 관리 및 공동활용 기반 마련 제2절 데이터 공동활용 활성화 제3절 소결
VI. 결론 및 제언	▶ 제1절 결론 제2절 정책제언

## 제3절 연구 기대효과

---

### 가. 문화·관광·콘텐츠 데이터 공동활용 기반 마련

문체부 산하 소속·공공기관 통계·데이터 보유 및 활용 현황에 대해 체계적인 조사와 분석을 실시함으로써 문체부 차원의 정책분야별 데이터 소재 및 현황을 파악하고 이를 기반으로 정책분야별 중요도를 감안한 조사결과 활용 방안을 마련한다. ‘데이터기반행정 중점추진 분야 및 공동활용 가능 데이터’에 대한 목록화를 진행한다. 또한 기관별로 분산된 공공데이터 및 정보시스템 간 데이터 공동활용을 위한 데이터 제공·연계·활용 근거를 마련한다.

데이터 분석활용 과제(안) 발굴을 통해 공동활용 데이터의 주요 활용 분야, 수요가 높은 공동활용 데이터의 속성·유형 등을 파악하고, 데이터 수집·제공·연계 등의 활용과정에서 발생할 수 있는 행정적·제도적 이슈 등 데이터 거버넌스 관점의 시사점을 도출한다.

### 나. 데이터 공동활용 활성화 및 역량 강화를 위한 방안 제시

문체부 소속·산하기관의 데이터 공동활용이 원활하게 추진될 수 있도록 데이터 관리, 관련 제도 검토, 역량 강화 등이 포함된 데이터 공동활용 활성화 방안을 마련하고 이를 기반으로 데이터 기반 업무프로세스를 강화 방안을 제시한다.

범정부차원의 데이터관련 정책기조와 ‘데이터기반행정 활성화에 관한 법률’에 대응하는 실효성 있는 데이터 공동활용 기반을 마련하고, 본 연구결과를 연간 단위로 행정안전부에 제출이 필요한 문체부 ‘데이터기반행정 활성화 시행계획’ 수립 시 필요한 기초자료로 활용한다.

문체부 데이터분석센터 설치·운영 기관으로서 연구원이 지향해야할 데이터기반행정 관련 기능과 역할을 제시한다.





문화·관광·콘텐츠 데이터 공동활용 기반 구축 연구

제2장

데이터기반행정 추진여건 및  
환경 분석



# 제1절 데이터기반행정의 의의

## 1. 데이터의 정의

“데이터”의 사전적 정의는 “넓은 의미에서 데이터는 의미 있는 정보를 가진 모든 값, 사람이나 자동 기기가 생성 또는 처리하는 형태로 표시된 것”을 뜻한다. 어떠한 사실, 개념, 명령 또는 과학적인 실험이나 관측 결과로 얻은 수치나 정상적인 값 등 실체의 속성을 숫자, 문자, 기호 등으로 표현한 것이며 데이터에 특정한 의미가 부여될 때 정보가 된다(두산백과, [www.doopedia.co.kr](http://www.doopedia.co.kr)). 즉, 데이터 자체는 단순한 사실에 불과하지만, 일련의 처리과정에 따라 특정한 목적에 소용되는 정보를 만들기 위한 재료로 사용되는 것을 뜻한다(두산백과, [www.doopedia.co.kr](http://www.doopedia.co.kr)).

공공 영역에서 “데이터” 정의는 2021년 6월 9일 제정된 「데이터기반행정 활성화에 관한 법률(이하 데이터기반행정법)」 제2조제1호에서 확인할 수 있다.

〈표 2-1〉 「데이터기반행정법(제2조제1호)」 데이터 정의

“데이터”란 정보처리능력을 갖춘 장치를 통하여 생성 또는 처리되어 기계에 의한 판독이 가능한 형태로 존재하는 정형 또는 비정형의 정보를 말한다.

데이터행정기반에서의 데이터의 정의는 기존의 데이터 정의와는 차이가 있다. 기존에는 데이터를 단순한 정보의 저장이나 수집을 의미하였다. 데이터의 영역은 각종 디지털 디바이스들을 통하여 저장·수집된 데이터 속에서 가치 있는 정보를 찾아내어 알기 쉽게 전달하고, 정보를 원하는 사람이나 기관에 제공하는 전 과정을 전부 포괄하는 개념으로 확장되었다(손현, 2017).

데이터기반행정법상, 데이터는 국가경쟁력의 핵심자원이다. 이는, 행정적 차원에 국한되지 않으며, 데이터를 기반 정책 추진으로 이어져 국민생활까지 아우른다. 즉, 데이

터는 미시적인 차원에서는 국민 개개인의 생활을, 거시적인 차원으로는 국가의 경쟁력에 이르기까지 모든 영역에서 매우 중요한 시대임을 보여준다.

데이터기반행정법에서 공공기관은 기관 자체적으로 보유·관리한 데이터뿐만 아니라, 타기관의 데이터를 취득하여 활용할 수 있음을 설명한다. 공공기관이 보유·관리하는 데이터에 대해 2013년 7월 제정된 「공공데이터의 제공 및 이용활성화에 관한 법률(이하 공공데이터법)」의 제2조제1호에서는 이를 “공공데이터”라고 정의한다.

〈표 2-2〉 「공공데이터법(제2조제1호)」 공공데이터 정의

“공공데이터”란 데이터베이스, 전자화된 파일 등 공공기관이 법령 등에서 정하는 목적을 위하여 생성 또는 취득하여 관리하고 있는 광(光) 또는 전자적 방식으로 처리된 자료 또는 정보로서 다음 각 목의 어느 하나에 해당하는 것을 말한다.

가. 「전자정부법」 제2조제6호에 따른 행정정보

나. 「지능정보화 기본법」 제2조제1호에 따른 정보 중 공공기관이 생산한 정보

다. 「공공기록물 관리에 관한 법률」 제20조제1항에 따른 전자기록물 중 대통령령으로 정하는 전자기록물  
라. 그 밖에 대통령령으로 정하는 자료 또는 정보

공공데이터는 중앙정부·지방자치단체 및 공공기관이 보유·관리하는 데이터를 일반 국민이 자유롭게 이용할 수 있도록 다양한 형태로 개방·제공하고 새로운 데이터 가치를 창출하도록 지원하고자 함에 초점을 둔다.

문화체육관광부는 2020년 9월 25일 제정한 「문화체육관광부 데이터 관리 규정(문화체육관광부훈령 제430호)」에 “데이터”와 “공공데이터”, “빅데이터”의 정의를 제3조에 담았다. 데이터의 정의는 데이터기반행정법, 공공데이터에 대한 정의는 공공데이터법의 내용과 동일하며, 빅데이터에 대해서는 다음과 같이 정의하고 있다.

〈표 2-3〉 「문화체육관광부 데이터 관리 규정」 빅데이터 정의

“빅데이터”란 아날로그 또는 디지털 환경에서 생성되는 정형 또는 비정형의 수치, 문자, 영상 등의 대량 데이터의 집합과 이로부터 추출한 가치, 결과를 분석하는 기술과 환경을 말한다.

빅데이터를 정의하는 관점은 크게 두 가지로 볼 수 있다. 첫째, 데이터의 양 등 빅데이터가 가지는 특성에 따른 관점이다. 빅데이터는 전자적 기술 향상에 따른 데이터가 가지는 특성인 데이터의 양(volume), 속도(velocity), 다양성(variety)에 초점을 맞춘 개념이다. 둘째, 빅데이터의 가치와 활용 효과 측면과 그 기술을 강조하는 것이다(손현,

2017). 상기의 규정에서와 같이 “대량의 정형 또는 비정형 데이터 세트로부터 가치를 추출하고 결과를 분석하는 기술”로 정의한다.

“데이터”의 정의를 「데이터기반행정법」과 「공공데이터법」 및 「문화체육관광부 데이터 관리 규정」 상의 내용을 바탕으로 주요 특징을 3가지로 정리할 수 있다.

첫째, 데이터를 운영하는 주체가 ‘공공기관’이다.

둘째, 데이터의 형태는 ‘기계에 의한 판독 가능’해야 한다. 공공데이터법에서는 ‘광(光) 또는 전자적 방식으로 처리’된 형태라고 설명한다. 즉, 두 법 모두 공통적으로 데이터는 ‘기계판독이 가능한 형태(machine-readable)’로 제공되어야 함을 명시하고 있다.

셋째, 데이터의 범주는 ‘정형 또는 비정형의 정보’ 또는 ‘전자적 자료 또는 정보’이다.

〈표 2-4〉 데이터와 공공데이터의 차이 비교

구분	데이터	공공데이터
소관 법률	데이터기반행정 활성화에 관한 법률 (법률 제17370호, 2020. 6. 9., 제정)	공공데이터의 제공 및 이용 활성화에 관한 법률 (법률 제11956호, 2013. 7. 30., 제정, 법률 제17344호 2020. 6. 9., 개정)
목적	데이터를 기반으로 한 행정의 활성화에 필요한 사항을 정함으로써 객관적이고 과학적인 행정을 통하여 공공기관의 책임성, 대응성 및 신뢰성을 높이고 국민의 삶의 질을 향상시키는 것을 목적으로 함	공공기관이 보유·관리하는 데이터의 제공 및 그 이용 활성화에 관한 사항을 규정함으로써 국민의 공공데이터에 대한 이용권을 보장하고, 공공데이터의 민간 활용을 통한 삶의 질 향상과 국민경제 발전에 이바지함을 목적으로 함
범주	기계 판독이 가능한 정보	광(光) 또는 전자적 방식으로 처리된 자료 또는 정보

출처: 관련 법률 참조하여 연구진이 정리

상기와 같은 데이터의 정의를 바탕으로 본 연구에서 조사, 활용하고자 하는 문화체육관광 분야 데이터는 데이터 운영의 주체, 형태 및 범주 등을 고려하여 다음과 같이 정리할 수 있다.

〈표 2-5〉 문화체육관광 분야 데이터 정의

“문화체육관광 분야 데이터”란 문화체육관광부 및 소속·공공기관에서 관계 법령 및 연구·조사·사업 등의 활동을 통해 생산, 관리, 취득하여 관리되고 있는, 기계판독이 가능한(machine-readable) 정형 또는 비정형 형태의 정보를 말한다.
--

## 2. 데이터기반행정의 개념

데이터(data)에 개념을 사전적인 의미, 정책적 관점, 데이터기반행정법에서 각각 그 개념을 살펴보았다. 이어, 데이터기반, 데이터기반행정의 개념은 무엇인지에 대해서도 차례로 파악해보고자 한다.

데이터기반(DATA基盤)에서 “기반(基盤)”은 사전적 의미로 “기초가 되는 바탕”을 뜻하며, “기반하다”라는 동사는 “어떤 대상이 다른 대상에 바탕을 두다”를 뜻을 가지고 있다(고려대 한국어대사전, 2009). 데이터의 정의와 연계에서 살펴보면, 데이터기반은 “정보 처리능력을 갖춘 장치를 통하여 생성 또는 처리되어 기계에 의한 판독이 가능한 형태로 존재하는 정형 또는 비정형의 정보”인 데이터를 바탕에 두고 있다.

데이터기반행정(DATA基盤行政, Data-based administration)에 대한 개념은 데이터기반행정법 제2조제2항에서 확인할 수 있다.

〈표 2-6〉 데이터기반행정의 정의

---

“데이터기반행정”이란 공공기관이 생성하거나 다른 공공기관 및 법인·단체 등으로부터 취득하여 관리하고 있는 데이터를 수집·저장·가공·분석·표현하는 등(이하 “분석등”이라 한다)의 방법으로 정책 수립 및 의사결정에 활용함으로써 객관적이고 과학적으로 수행하는 행정을 말한다.

---

데이터기반행정법에는 상기와 같이 데이터기반행정 정의가 명시되어 있다. 이와 관련하여, 데이터기반행정 활성화에 관한 법률 제정(안) 입법예고(행정안전부, 2020.5.19.)에서 해당 법의 제정이유를 통해 데이터기반행정의 개념을 다시 한 번 살펴보면 다음과 같다. 데이터기반행정법을 제정하고자 하는 이유는, “국민 생활전반에 걸쳐 데이터의 활용이 국가 경쟁력을 좌우하는 핵심자원으로서 그 중요성이 부각되고 있는 바, 데이터를 기반으로 한 행정을 활성화하기 위하여 공공기관은 공동활용할 필요가 있는 데이터를 등록할 수 있도록 하고, 등록되지 아니한 데이터를 제공받으려는 경우 데이터 소관 공공기관의 장에게 데이터 제공을 요청할 수 있도록 … 행정의 책임성, 대응성 및 신뢰성을 높이고 국민의 삶의 질을 향상시키려는 것”이다. 요약하면, “데이터를 기반으로 한 행정을 활성화”하는 것이 데이터기반행정의 핵심이다.

“데이터기반행정”의 등장은 기존의 증거기반(evidence-based) 개념을 행정에 도입한 것에서 찾을 수 있다. 정부는 제도적으로 이 개념을 활용하고 있으며, 이는 「데이터

기반행정법」으로 법제화되어 있다. 데이터기반행정은 다음과 같은 특성을 갖고 있다.

첫째, 데이터기반행정법의 데이터는 필수적으로 공공기관과 관련이 있다. 데이터기반행정의 주체는 공공기관이기 때문이다. 따라서 데이터기반행정법의 데이터는 공공기관이 자체적으로 생산·보유하고 있는 데이터이거나, 공공기관이 타기관에서 취득하여 관리하는 데이터를 뜻한다. 데이터기반행정법에서 “공공기관은 공공기관 운영에 관한 법률상 공공기관뿐만 아니라 중앙행정기관이나 지방자치단체와 같은 행정기관, 그밖에 대통령령으로 정하는 법인·기관 및 단체”를 포함한다(데이터기반행정법 제2조제3호).

둘째, 데이터기반행정은 행정 수행에 있어 객관성과 과학성에 주안점을 둔다. 데이터기반행정에서 객관성을 강조하는 이유는, 행정 및 정책과정에 정책결정자가 주관적 직관에 의존하여 발생하는 문제를 객관적 증거로 뒷받침하여 신뢰성을 주고자 함에 있다. 과학적 방법의 중요성은 증거기반정책에서 강조되어 온 사항이다. 증거기반정책 수립은 공공정책이 실증적 자료나 증거보다는 직관, 경험, 가치 등에 바탕을 두고 이루어져왔다는 점에 대한 비판으로 등장하였다(한국정보화진흥원, 2016). 따라서 데이터기반행정은 증거기반정책의 이론적 계보를 이어 등장한 개념이라고 볼 수 있다.

셋째, 데이터가 정책과정에 활용되는 과정이 제시된다. 이를 위해서는 우선 데이터가 있어야 하므로 관련 데이터 수집 및 저장(적재)하는 작업이 필요하다. 데이터는 특정 행정 혹은 정책문제 등의 해결에 유의미한 결론을 도출하기 위한 분석을 할 수 있도록 가공의 절차를 거친다. 가공된 데이터 분석하고, 필요에 따라서는 데이터 간 연계·분석을 통해 그 결과를 정책결정에 활용하는 과정이 이어지게 되는 것이다. 데이터를 정책결정에 활용하는 과정을 통해, 데이터기반행정에서 부각되는 주요 사항으로 데이터를 통한 과학적 융합에 대한 사고와 역량임을 알 수 있다.

### 3. 데이터기반행정과 기존의 데이터 관련 제도와의 관계

데이터기반행정 논의를 이해하기 위해서는 그 전에 논의되고 도입되었던 정책과 제도 및 관련되는 핵심 기술을 이해할 필요가 있다. 이에, 공공데이터, 증거기반정책 및 전자정부 등과 데이터기반행정과의 관계를 제도 및 개념 등을 통해 정리해보고자 한다.

### 3.1. 공공데이터

「공공데이터의 제공 및 이용 활성화에 관한 법률」(이하 공공데이터법)은 2013년 7월 30일 제정되었고, 2020년 6월 9일 타법개정을 거쳐 현재에 이르고 있다.

법 제정의 맥락에서, 정부3.0 추진계획에서는 “소통하는 투명한 정부” 전략의 일환으로 “공공데이터의 민간 활용 활성화”를 중점 추진과제로 선정하였다(관계부처 합동, 2013:2). 공공데이터 활용을 통해 신성장동력을 창출하는 것이 직접적 목적이었으며, 이를 위한 공공데이터 개방·활용 인프라로서 공공데이터포털로의 제공 창구 일원화, 국무총리와 민간 공동위원장으로 하는 공공데이터전략위원회 구성, 공공데이터법 제정 등이 이루어졌다(관계부처 합동, 2013: 13-15).

공공데이터 개방과 관련하여 국민과 기업이 원하는 공공데이터를 쉽게 이용할 수 있도록 2013년 공공데이터포털([www.data.go.kr](http://www.data.go.kr))을 구축하였다. 각 기관별로 흩어져있는 공공데이터를 한 곳에서 통합 제공하고, 파일데이터, 오픈API 등 다양한 형태로 제공하고 있다(행정안전부 홈페이지, [www.mois.go.kr](http://www.mois.go.kr)).

데이터기반행정법과 공공데이터법의 차이는 다음과 같다.

〈표 2-7〉 데이터기반행정법과 공공데이터법의 차이 비교

구분	데이터기반행정법	공공데이터법
소관법률	「데이터기반행정 활성화에 관한 법률」	「공공데이터의 제공 및 이용 활성화에 관한 법률」
목적	과학적 행정을 통하여 공공기관의 책임성, 대응성 및 신뢰성 향상	민간 활용을 통한 신규 비즈니스와 일자리 창출, 국민편익 향상
활용	데이터 제공·연계 및 공동활용하여 정책 수립과 관련 의사결정에 활용	공공데이터의 공공 및 민간 활용
범주	기계 판독이 가능한 정보	기계 판독이 가능한 형태의 공공데이터
제공창구	공동활용데이터 등록관리시스템 ( <a href="http://insight.go.kr">insight.go.kr</a> )	공공데이터 포털 ( <a href="http://data.go.kr">data.go.kr</a> )

출처: 관련 법률 참조하여 연구진이 정리

공공데이터법은 정부 및 공공기관에 축적된 공공데이터를 민간에서 활용할 수 있도록 제공하여 데이터 이용 경계를 없애고, 이용자 확대를 이끄는 데 주안점을 두고 있다. 즉, 공공데이터법은 공공에서 제공하는 개방된 데이터의 대상이 정책입안자에만 국한하지 않는, 이용 제한 없는 이용자의 활용에 초점을 둔다. 반면, 데이터기반행정법은 정부 및 공공기관 간 혹은 민간과 공유·연계·활용하여 기존의 데이터를 다각적인 시각에서 사용



하고, 정책적 함의를 찾고자 하는 데이터 정책이다. 공공 영역에서 생산·관리하고 있으나, 공개하지 않는 데이터를 지속적으로 찾아내고, 이 데이터들을 어떠한 방법을 통해 보다 더 정책적 목적성과 결과적 함의한 데이터로 활용할 수 있는 지에 주안점을 둔다.

### 3.2. 증거기반정책

증거기반정책(evidence-based policy)은 체계적 연구에서 도출된 외적 증거와 전문지식을 통합하여 결정하는 것을 의미하며, 정책 결정 과정에서 과학적 증거(scientific evidence)의 정책 활용을 강조한다. 경험이나 직관도 중요한 의사결정의 근거가 될 수 있지만 체계적 오류를 피하는 것은 확률적으로 어렵기 때문에, 정책 과정의 증거기반성이 낮다는 것은 정책결정의 주요 행위자들이 과학적 증거보다는 개인의 주관적 경험이나 직관에 의존한다는 것을 의미한다(윤건, 2019). (윤건, 2019). 과학적 증거는 과학적 방법에 의해서 획득된 증거를 의미한다. 이러한 증거는 전문가의 지식과, 연구 간행물, 과거의 정책평가자료, 통계적 모델을 통한 산출물 등이 포함될 수 있다. 실제 행정현장에서는 이러한 증거의 질을 어느 정도 담보되었는지가 그 잣대가 되며, 그러한 기준은 증거의 품질, 신뢰성, 관련성과 비용이 포함될 수 있다(손현, 2015).

데이터의 수집 및 축적 기술의 누적과 발전으로, 증거기반(evidence based)을 넘어서 데이터 주도(data driven)의 정책 수립의 기능과 역할이 증대되고 있다. 다만, 증거기반정책은 정책 수립 시 데이터와 같은 과학적 증거를 근거로 판단해야 한다는 정책의 정당화 측면이 강조되는 반면, 데이터 주도 정책 수립은 축적 가능한 포괄적인 데이터를 활용하여 정책이슈와 대안을 발굴할 수 있다는 점에서 정책 수단 혹은 정책 환경의 복잡성 이론에 더 부합하는 것으로 볼 수 있을 것이다(오세영 외, 2017).

데이터기반행정은 증거기반정책의 영향을 받았다. 데이터기반행정법상 객관적이고 과학적인 행정을 위한 논의는 증거기반정책의 주요 골자이다. 따라서 데이터기반행정도 증거기반정책과 크게 다르지 않다는 논리적 귀결이 가능하다. 그러나 데이터기반행정에서 활용하는 데이터가 주로 빅데이터라는 점이 다르다. 증거기반은 방법론적 엄밀성과 인과성을 더 중요시하며, 데이터기반행정은 데이터 활용 시 데이터 간 상관성과 기술적 패턴 파악에 목적이 있다. 정리하면, 데이터기반행정은 기술적 진보로 새롭게 형성되고 분석이 가능해진 빅데이터를 사용한 정책결정을 의미하는 점에서 차이가 있다. 여기서 전제조건은 이러한 빅데이터의 사용이 문제를 해결하는데 더 유용하다는 것이다(윤건 외, 2019).

〈표 2-8〉 데이터기반행정과 증거기반 비교

구분	데이터기반행정	증거기반
핵심 키워드	데이터 주도(data driven)	증거기반(evidence based)
정책 주안점	추적 가능한 포괄적인 데이터를 활용하여 정책이 슈와 대안을 발굴하기 위한 정책 수단	정책 수립 시 데이터와 같은 과학적 증거를 근거로 판단해야 한다는 정책의 정당화 측면 강조

출처: 연구진 정리

### 3.3. 전자정부

전자정부는 “정보기술을 활용하여 행정기관 및 공공기관의 업무를 전자화하여 ... 상호 간의 행정업무 및 국민에 대한 행정업무를 효율적으로 수행하는 정부”를 의미한다(전자정부법 제2조제1호). 전자정부는 행정의 생산성과 투명성, 민주성이 제고되고 궁극적으로 국민의 삶의 질 향상을 목표로 하기 때문에, 정부 내부적으로는 전자문서 사용으로 비용이 절약되고 빠른 의사소통으로 생산성이 높아지며, 원스톱 서비스를 통해 대국민 행정서비스의 질 제고를 도모할 수 있다(윤건, 2019).

전자정부는 정보기술과 초고속 통신기술 등의 접목으로 행정업무의 혁신으로 가능한 일이다. 모바일 기술 발달은 모바일 정부를 탄생시켰고, 사물인터넷 기술 발달로 유비쿼터스 정부, 스마트 기술을 접목한 스마트 정부로 지속적인 변화를 거듭하였다. 나아가 2020년 7월, 정부는 블록체인·인공지능(AI) 등 신기술과 5G·클라우드 등 디지털 기반을 활용한 지능형 정부로 행정혁신을 추진하고 있다(정책브리핑, 2020.7.15.)

이외에, 지식정부는 기술적 맥락 이외에 국가경쟁력 관점에서 논의가 이루어졌다. 이러한 맥락에서 데이터는 아무 의미가 없지만, 정보는 데이터를 분석한 결과로 일정한 의미가 부여되고, 지식은 여기에 부가가치의 의미를 담아내기 때문에 지식정부는 국가 차원에서 지식관리시스템을 작동하여 부가가치를 향상시키는 한편 국가경쟁력을 극대화하는 정부를 지향한다(윤건, 2019).

결과적으로, 데이터기반행정이 추구하는 목표는 전자정부 및 지식정부와 동일하다. 데이터기반을 통한 데이터 연계 및 활용을 통한 분석의 결과로 정책 결정에 반영하고, 이 결과로 행정서비스의 질이 향상되면, 결과적으로 국민의 삶에도 긍정적 효과가 발휘되기 때문이다. 그럼에도 불구하고, 기존의 개념들이 지속성을 갖고 개념적 보완을 이어가기 보다는 또 다른 개념으로 등장하는 이유를 찾아보면 다음과 같다. 전자정부의 도입은 국

민들에게 직접적으로 영향을 미치게 되는 영향에 초점을 둔다. 지식정부는 공무원들이 축적한 업무 처리 노하우를 조직 내 공무원 구성원들과 공유하여 정부조직 생산성을 극대화시키고, 민간과도 해당 정보를 공유하여 민간 차원 모두에서 실질적인 경쟁력 제고를 지향한다. 반면, 데이터기반행정은 공공기관이 중심이 되어 데이터를 기반으로 객관적이고 과학적인 의사결정과 정책수립을 통한 행정서비스 구현과 정책결정 향상성 제고를 추구한다. 정책결정 향상성 제고 측면에서 볼 때, 전자정부의 개념과 데이터기반행정의 개념은 밀접한 관계가 있다. 전자정부와 데이터기반행정의 차이는 다음과 같이 정리할 수 있다. 전자정부가 자료 제공에 초점을 두고 있고, 데이터기반행정도 보유 데이터를 타 기관에 제공하는 점은 유사하나, 데이터기반행정은 데이터 제공에만 국한하지 않고 타 공공·민간기관 데이터와의 연계 및 공동활용을 통해, 방법론으로 데이터를 분석하고 해당 결과를 행정과정에 활용함으로써 보다 효용성 높은 행정서비스를 실현함에 있다.

### 3.4. 데이터기반행정 관련 법·제도 개념의 특징

데이터기반행정 관련 개념의 정의와 특징 및 한계를 요약하면 <표 2-9>와 같다.

<표 2-9> 데이터기반행정 관련 개념의 정의와 특징

	정의	특징	한계
데이터기반행정	데이터 기반의 객관적·과학적 행정체계를 구축하기 위한 제도적 기반	사회·환경변화를 대응해 행정의 책임성, 대응성, 신뢰성 향상	데이터기반행정의 범위가 포괄적이기 때문에 기능과 목적이 겹칠 수 있는 기존의 법 제도들과의 중복성이 있음
공공데이터	정부, 지자체, 공공기관이 전자적으로 생성 또는 취득하여 관리하고 있는 모든 전자화된 파일(DB)	다양한 공공기관이 생성 및 관리해왔던 공공 데이터의 통합적 활용을 통해 데이터기반행정의 주요 재원의 역할	부처별로 데이터 공유가 미진할 경우, 데이터가 구조화되지 않을 경우 활용도가 떨어짐
전자정부제도	행정기관 및 공공기관의 업무에 정보기술을 활용, 전자화하여 행정업무를 효율적으로 수행하는 정부	행정의 생산성과 투명성, 민주성 제고를 통해 대국민 행정서비스의 질적 향상 도모	하드웨어에 치중, 중복투자와 공동 활용 미흡으로 인해 비효율성 및 통합서비스의 어려움
증거기반정책	과학적 증거에 기반을 두어 정책을 결정하려는 시도	신뢰성과 타당성 있는 정책과정을 통해 생성된 정보와 지식이 축적되어 더 나은 정책 형성 가능	과학적 증거 부재로 인한 적용범위의 제한성, 담당자의 이해부족에 따른 기술적 해석의 한계

출처: 윤건 외(2020), 「한국지방자치학회보」 32권 4호(통권112) 재인용, 연구진 재정리

데이터기반행정은 기존의 전자정부제도, 증거기반정책, 빅데이터가 가지는 핵심적 특징과 가치에 영향을 받고 있다. 데이터기반행정이 기존의 제도와 다른 점은 정책과정의 특정한 과정의 의사결정에만 사용되는 것이 아니라 정책과정의 모든 과정에 데이터를 활용하는 것을 의미한다. 정리하면, 데이터기반, 빅데이터 그리고 증거기반정책은 주관적 판단이 아닌 객관적으로 수집된 자료 및 자료에 바탕을 둔 의사결정을 시도한다는 측면에서 유사점이 있다. 그러나 데이터기반행정은 분석의 기반이 되는 데이터의 적용범위가 증거기반정책과 차이가 있다.

[그림 2-1] 데이터기반행정의 개념적 배경 및 특징



출처: 윤건 외(2020), 「한국지방자치학회보」 32권 4호(통권112) 재인용

## 제2절 데이터기반행정 정책환경 분석

### 1. 데이터기반행정 정책환경

#### 1.1. 데이터 경제의 가속화

정부는 공공데이터법 시행 이후, 사회·경제적으로 파급효과가 큰 국가중점데이터를 과감히 개방하여 데이터경제 활성화를 지원 하고 있다. 다만, 데이터 공동활용을 위한 법제도, 전략 등 추진기반 미흡으로 정책 수립·의사결정 지원을 위한 공공데이터의 분석·활용은 미흡하였다.

데이터는 디지털 대전환 시대의 국가경쟁력 확보의 핵심 원천이다. 따라서 데이터는 정부의 책임성과 투명성을 증진시키기 위한 전략적 국가자산이다. 데이터 시장의 성장과 함께 데이터 경제의 가속화가 진행됨에 따라, 미국, EU 등 주요국은 데이터 전략을 수립 하여, 적극적으로 디지털 시대에 대응하고 있다.

〈표 2-10〉 주요국 데이터 전략 현황

- 미국 : 미국 연방 데이터 전략(FDS) 시행계획 수립('19.12)
- EU: EU 디지털전략 정책안 內 유럽 데이터전략 수립('20.2)
- OECD: The Path to Becoming a Data-Driven Public Sector('19)

출처: 행정안전부, 제1차 데이터기반행정 활성화 기본계획(안) 재인용, 2021.2.

데이터 경제는 혁신성장 정책의 중요한 과제다. 혁신성장은 소득주도 성장과 함께 새로운 경제성장을 위한 정부의 핵심전략이다(대한민국 정책브리핑, 2020.3.13.). 우리나라는 2017년 10월 대통령 직속 ‘4차산업혁명위원회’를 신설하는 등 대응체계를 마련했으며, 2017년 11월에는 4차 산업혁명을 위기가 아닌 새로운 성장과 일자리 창출의 기회로 삼기 위한 범국가적 ‘4차 산업혁명 대응 계획’을 수립했다(대한민국 정책브리핑, 2020.3.13). 문재인 대통령은 2018년 8월 31일 ‘데이터경제 활성화 규제혁신 현장방

문'에서 '데이터경제로의 전환'을 선언하고 '데이터를 가장 잘 다루면서, 가장 안전하게 다루는 나라'의 비전을 제시한 바 있다(청와대, 2018).

## 1.2. 범정부 데이터 공동활용 기반의 확충

2020년 코로나19 확산에 따른 위기상황 극복, 사회현안 해결에 대한 선제적 대응 등을 위해 정부의 맞춤형 서비스의 요구와 필요가 증대되었다. 이를테면, 확진자의 동선 파악, 공적마스크 제공, 긴급재난지원금 지급 등 코로나 19에 대한 정부의 신속한 대응 추진이 이루어졌고, 이와 같은 정책 추진에는 데이터가 기반이 되었다. 아울러, 개개인의 상황에 맞는 공공서비스 제공 요구 또한 증가하였다. 의료정보, 복지수급정보, 신용정보 등을 융합 분석하여 복지 사각지대, 중복 수혜 등 기존 제도의 문제점 개선 및 신규 복지정책 발굴에도 데이터기반의 정책이 실제 영향력을 미치고 있다. 따라서 보다 효율적인 수립, 실행, 모니터링 등 공공서비스를 위한 소 과정에서 데이터기반의 혁신을 적용할 필요가 있다. 즉, 데이터 연계 및 민관 협력을 통해 혁신적 공공서비스 요청의 증가에 따라 정부의 정책도 빠르게 상응할 수 있는 여건 조성이 마련되어야 한다.

이를 위해서, 범정부 데이터 공동활용 기반의 확충이 필요하다. 기관이 보유한 데이터, 존재하는 데이터 위주의 활용으로 분석주제가 한정적이고 활용분야가 제한적이다. 그러므로 한정된 데이터의 한계를 넘어서 기관 간 데이터 공동활용을 활성화하고 개별 기관의 지식을 상호 공유할 필요가 있다. 현실적으로 다양한 분야의 행정업무와 데이터를 접목시키는 작업이 익숙하지 않고, 데이터 활용의 법적·윤리적 이슈를 파악하기 어렵다. 따라서 데이터를 올바르게 이해하고 활용할 수 있는 기반환경 조성이 수반되어야 한다.

데이터에 기반한 행정혁신을 촉진하기 위해서는 당면한 문제들을 해결해야 한다. 기관별 현안 및 이슈 해결을 위한 단발적인 데이터 분석에만 그치고 있어 다부처 공통현안 및 장기적 국가과제 분석에 한계가 있기 때문이다. 개별이슈 분석과 더불어 다부처 협업 과제, 국가 미래이슈 등에 대한 전략적 대응방안 마련이 필요하다. 데이터기반행정이 강조되고 있으나, 선례·경험으로만 판단하는 기존 의사결정과정의 틀을 벗어나기 어렵다. 범정부 차원의 일하는 방식 개선을 촉진하여 행정 효율성 제고 및 양질의 대국민서비스 제공하는 방법을 모색해야 한다. 이와 같은 데이터기반행정 활성화를 위해서는 데이터

공동활용을 제한하는 장애요소 및 법적 제약사항을 발굴·개선하여야 한다(행정안전부, 2021). 이와 같은 배경에서 정부는 데이터기반행정의 추진과 활성화를 도모하고 있다. 현재 데이터기반행정을 추진하기 위한 법 제도, 정부 정책의 계획, 정부부처 및 관련 공공기관의 조직 현황에 대해서 살펴보고자 한다.

## 2. 데이터기반행정 관련 제도 현황

### 2.1. 법 제도

데이터기반행정법은 총 25개 조항으로 구성되어 있다. 제1장은 목적, 정의 등에 대한 내용을 담고 있으며, 제2장은 추진체계, 제3장은 데이터의 등록 및 제공 절차, 제4장은 데이터기반행정 기반구축 등으로 구성되어 있다. 이 가운데 데이터기반행정법의 핵심 사항은 제4장(제16조~제25조)이라고 볼 수 있다.

제4장은 데이터관리체계의 구축(제16조), 데이터기반행정 표준화(제17조), 데이터통합관리 플랫폼(제18조), 데이터기반행정 책임관(제19조), 데이터분석센터(제20조), 데이터기반행정 전문기관(제21조), 데이터기반행정 실태 점검 및 평가(제22조), 데이터기반행정 우수사례의 발굴·보급(제23조), 데이터 관련 전문인력 양성(제24조), 민간 및 국제 협력(제25조) 등 데이터의 제공과 연계, 공동활용 등과 직접적으로 관련 있는 내용들이다. 특히, 제16조, 제17조, 제18조는 데이터에 대한 메타데이터, 데이터 관계도, 데이터의 표준화 등의 체계적 관리를 통해 데이터 간 유사성이나 관련성을 확보하여 데이터 공동활용의 가능성을 제고하는 것에 힘을 싣고 있다. 현재 데이터기반행정법을 기반으로 데이터기반행정 업무 관련 주무부처 및 부서는 행정안전부 빅데이터분석활용과이다.

데이터기반행정법과 공공데이터법은 각각 공공기관의 데이터에 관한 내용을 담고 있기 때문에 상호 밀접한 관련성이 있다. 2013년 제정된 공공데이터법에 의거, 공공데이터 관련 업무의 현재 주관 부처 및 부서는 행정안전부 공공데이터정책과이다. 공공데이터법은 제2장에서 공공데이터 정책수립, 제3장에서 공공데이터 등록 등 제공 기반 조성, 제4장에서 공공데이터 제공절차 등에 관한 내용을 다루고 있다. 공공데이터의 개방은 특정 개인과 집단을 대상으로 한정하지 않기 때문에, 공공데이터 개방성이 높아질수록, 데이터 접근성도 높아지기 때문에, 데이터의 다양한 활용 활성화에 기여할 수 있다. 그

러나 현실적으로 개방 및 제공되고 있는 공공데이터 가운데 기관의 행정업무 목록 등과 같이 정책적 활용 가치에 대해서 고민이 필요한 데이터 등도 다수 존재하기 때문에 정책 문제 해결을 위한 데이터 분석에 필요한 정도가 되지 못한다는 비판이 제기되고 있다.

전자정부법은 2001년에 제정되었다. 전자정부법은 “이 법은 행정업무의 전자적 처리를 위한 기본원칙, 절차 및 추진방법 등을 규정함으로써 전자정부를 효율적으로 구현하고, 행정의 생산성, 투명성 및 민주성을 높여 국민의 삶의 질을 향상시키는 것을 목적으로 한다(전자정부법 제1조)”. 정부조직 내에서 데이터의 활용의 전제조건이 되는 것은 컴퓨터와 네트워크로 구성되는 ICT에 기반한 정부, 즉 전자정부이다(윤건, 2019). 전자정부란 “정보기술을 활용하여 행정기관 및 공공기관(이하 “행정기관 등”이라 한다)의 업무를 전자화하여 행정기관등의 상호 간의 행정업무 및 국민에 대한 행정업무를 효율적으로 수행하는 정부”, 전자정부서비스란 “행정기관등이 전자정부를 통하여 다른 행정기관 등 및 국민, 기업 등에 제공하는 행정서비스”를 의미한다(전자정부법 제2조제1,5호). 전자정부에서의 ‘정보’는 행정기관 등이 직무상 작성하거나 취득하고 관리하고 있는 행정정보를 뜻한다(전자정부법 제2장제6호). 한편, 정보자원은 행정기관 등이 보유하고 있는 행정정보이고, 관련 정보시스템, 관련 기술과 예산, 인력 등을 포함한다(전자정부법 제2조제11호). 전자정보법에서는 시스템의 연계와 통합, 행정정보 공동이용 등과 같이 공공의 데이터 활용과 관계되는 내용을 포함한다. 정보시스템을 활용한 행정정보의 연계와 통합, 공동이용 등은 데이터기반행정법에서 언급하는 데이터 연계와 공동활용을 활성화할 수 있는 기초 환경을 제공한다. 전자정부법 시행 관련 주무부처는 행정안전부이며, 디지털정부정책과에서 총괄을 맡고 있다.

전자정부법에서는 지역정보화와 관련하여 한국지역정보개발원을 규정하여 지역 이슈를 포괄하고 있다(전자정부법 제72조). 그러나 지역정보화 관련한 전자정부 정보화사업은 실질적으로는 스마트도시법을 중심으로 논의되고 있다. 스마트도시법은 유비쿼터스법(2008년 제정)을 2017년에 개정한 것이다. 스마트도시법에서 정의하는 스마트도시는 “도시의 경쟁력과 삶의 질의 향상을 위하여 건설·정보통신기술 등을 융·복합하여 건설된 도시기반시설을 바탕으로 다양한 도시서비스를 제공하는 지속가능한 도시(스마트도시법 제2조제1호)”이다. 스마트도시서비스는 “스마트도시기반시설 등을 통하여 행정·교통·복지·환경·방재 등 도시의 주요 기능별 정보를 수집한 후 그 정보 또는 이를 서로 연계하여 제공하는 서비스”를 말한다(스마트도시법 제2조제2호). 이 법에서는 정보 간



연계 제공의 내용을 포함하며, 관련 주무부처 및 부서는 국토교통부 도시경제과이다.

1995년에 제정된 정보화추진기본법은 2009년 명칭 변경을 통해 국가정보화기본법, 2020년에 전부 개정(2020.6.9.)을 통해 지능정보화 기본법으로 변경되었다. 지능정보화 기본법은 국가 및 지자체, 공공기관 및 국민 등 공공기관의 정보화 사업에 대한 포괄적인 내용을 담고 있는 대표적인 법이다. 이 법은 “지능정보화 관련 정책의 수립·추진에 필요한 사항을 규정함으로써 지능정보사회의 구현에 이바지하고 국가경쟁력을 확보하며 국민의 삶의 질을 높이는 것을 목적”으로 한다(지능정보화 기본법 제1조). 국가 차원의 정보화 전자정부법과 공공데이터법은 지능정보화 기본법 관할에 포함될 수 있다. 지능정보화법에서 정의하는 정보화는 “정보를 생산·유통 또는 활용하여 사회 각 분야의 활동을 가능하게 하거나 그러한 활동의 효율화를 도모하는 것(지능정보화 기본법 제2조제2호)”이다. 그리고 동법에서 정의하는 지능정보기술은 “데이터(부호, 문자, 음성, 음향 및 영상 등으로 표현된 모든 종류의 자료 또는 지식을 말한다)를 전자적 방법으로 수집·분석·가공 등 처리하는 기술(지능정보화 기본법 제2조제2호나항)”을 뜻한다. 지능정보화법에서 정의하는 데이터는 데이터기반행정법의 데이터 대비 그 범위가 확장된 영역이 있다. 지능정보화 기본법의 주무부처 및 부서는 과학기술정보통신부 정보통신정책과이다. 데이터기반행정법, 공공데이터법 및 전자정부법의 주무부처가 행정안전부이고, 스마트도시법의 주무부처가 국토교통부인데, 각 법은 유사한 내용들을 일부 담고 있기 때문에 향후에는 상호 법 기능의 유사성과 중복에 대해서는 여러 논의의 여지가 발생할 수 있는 상태이다.

데이터 이용 활성화 관련해서는 데이터 3법이 무엇인지 확인할 필요가 있다. 데이터 3법은 개인정보 보호법, 정보통신망 이용촉진 및 정보보호 등에 관한 법률(정보통신망법), 신용정보의 이용 및 보호에 관한 법률(신용정보법) 등 3가지 법률을 통칭한다(대한민국 정책브리핑, 2020.3.20.).

데이터 3법 개정안은 2020년 1월 9일 국회 본회의를 통과했다. 법률 개정안 주요내용으로는 ① 데이터 이용 활성화를 위한 가명정보 개념 도입, ② 관련 법률의 유사·중복 규정을 정비하고 추진체계를 일원화 하는 등 개인정보 보호 협치(거버넌스) 체계의 효율화, ③ 데이터 활용에 따른 개인정보 처리자의 책임 강화, ④ 모호한 ‘개인정보’ 판단 기준의 명확화 등이다(대한민국 정책브리핑, 2020.3.20.).

데이터 3법 개정의 개정으로 데이터 이용에 관한 규제 혁신과 개인정보 보호 협치(거버넌스) 체계 정비 등의 문제가 해결되어 안전한 데이터 이용을 위한 사회적 규범 마련

의 단초가 되었다(대한민국 정책브리핑, 2020.3.20.).

〈표 2-11〉 데이터 3법의 개정 주요 내용

법안	해당부처	목적	주요 내용
개인 정보 보호법	행정안전부	개인정보의 처리 및 보호에 관한 사항을 정함으로써 개인의 자유와 권리를 보호하고, 나아가 개인의 존엄과 가치를 구현	<ul style="list-style-type: none"> <li>•개인정보 관련 개념체계 개인정보·가명정보·익명정보로 명확화</li> <li>•가명정보 데이터를 제품·서비스 개발에 활용, 특정 개인을 알아보는 행위를 금지하고 이를 위반시 형사처벌·과징금 등 벌칙 부과</li> <li>•개인정보 관리·감독 기능을 개인정보보호위원회로 일원화, 중앙행정기관으로 격상</li> </ul>
신용 정보 보호법	금융위원회	신용정보 관련 산업을 건전하게 육성하고 신용정보의 효율적 이용과 체계적 관리를 도모하며 신용정보의 오용·남용으로부터 사생활의 비밀 등을 적절히 보호함으로써 건전한 신용질서를 확립하고 국민경제의 발전에 이바지함	<ul style="list-style-type: none"> <li>•데이터결합 및 데이터전문기관의 법적근거 마련</li> <li>•가명정보 금융분야 빅데이터 분석·이용, 가명정보의 주체 동의 없이 이용·제공 허용</li> <li>•전문개인신용평가업·개인사업자신용평가업·본인신용정보관리업 도입</li> <li>•신용정보집중기관이 공공기관에 요청할 수 있는 정보범위 확대</li> </ul>
정보 통신망법	과학기술 정보통신부 · 방송통신 위원회	정보통신망의 이용을 촉진하고 정보통신서비스를 이용하는 자를 보호함과 아울러 정보통신망을 건전하고 안전하게 이용할 수 있는 환경을 조성하여 국민생활의 향상과 공공복리의 증진에 이바지함	<ul style="list-style-type: none"> <li>•개인정보보호 관련 사항을 개인정보보호법으로 이관</li> <li>•온라인상 개인정보보호와 규제·감독 주체를 방통위에서 개인정보보호위원회로 변경</li> </ul>

출처: 대한민국 정책브리핑(2020.3.30.), 국가법령정보센터 내용을 기반으로 연구진 재정리

## 2.2. 관련계획

데이터기반행정법 도입에 따른 기본계획 및 실행계획 등은 2021년 최초 작성되었다. 따라서 관련하여 데이터기반행정법의 데이터 연계 및 공동활용 활성화를 위한 세부적인 정부부처의 발표는 현재 시점에서는 확인이 용이하지 않다. 이에, 기존의 데이터 연계 관련하여 최근 5년 이내의 공공데이터 부분 정부 계획 3가지를 비교분석하였다.

공공데이터 연계 관련 3개 계획안을 포함한 이유는, 공공데이터에 연계 관련 정부 주요 계획에서 언급하고 있는 주요 데이터가 빅데이터이며, 데이터 공유 및 활용 체계 구축 등과 같이 현 시점 기준에서 확인했을 때 데이터기반행정법 상의 내용과 충돌이 없고 유사한 내용들이기 때문이다.

〈표 2-12〉 공공데이터 연계 관련 정부 주요 계획

작성 주체	구분	시기	공공데이터 융합 관련 주요 과제
공공데이터 전략위원회 (행정자치부)	공공 빅데이터 활용 활성화 추진계획(안)	2016년 2월	<ul style="list-style-type: none"> <li>• (공공분야 빅데이터 추진체계 강화) 공공 빅데이터 협의회 확대, 분석모델 개발 위해 공공·민간 협의체 구성</li> <li>• (빅데이터 관련 법제도 개선) 기간간 데이터 공유·활용·조정 협의, 개인정보 활용</li> <li>• (개인정보보호 비식별화 기술 기준 마련) 빅데이터 개인정보보호 기술가이드, 개인정보 보호법 개선 방안</li> <li>• 빅데이터 분석 표준모델 정리, 적용</li> <li>• 공공 빅데이터 교육강화·인력양성</li> </ul>
공공데이터 전략위원회 (행정자치부)	제2차 ('17~'19) 공공데이터 기본계획	2016년 12월	<ul style="list-style-type: none"> <li>• (민·관 융합 공공데이터 개방 확대) 민·관 데이터 결합 활성화 지원</li> <li>• (공공데이터 통합 거버넌스 구축) 데이터 생성, 관리, 제공에 있어 통합적 체계로 개편</li> <li>• 민간데이터 범정부 공동 이용체계 마련</li> <li>• 데이터 품질관리 범위에 데이터융합분석 포함</li> </ul>
관계부처합동 (과학기술정보통신부)	데이터·AI경제 활성화 계획	2019년 1월	<ul style="list-style-type: none"> <li>• (양질의 데이터 축적 및 개방 확대) 빅데이터 플랫폼, 센터 구축, 빅데이터 네트워크 협의회</li> <li>• (연구데이터 공유·활용 체계 구축) 국가연구데이터플랫폼</li> <li>• (공공데이터 개방 및 효율적 관리) 데이터맵 구축, 국가중점데이터 지정·개방, 범정부 데이터플랫폼 구축</li> </ul>

출처: 해당 계획별 연구진 재정리

“공공 빅데이터 활용 활성화 추진계획(안)”에서는 제도 차원에서 추진체계와 법제도 기반의 강화, 빅데이터 활용을 위한 개인정보 비식별화 기술의 기준 마련, 빅데이터 분석 모델 마련, 공공 빅데이터 교육 등을 주요 과제로 제시하였다(공공데이터전략위원회, 2016.2). “제2차 공공데이터 기본계획”에서는 민·관 데이터 결합, 데이터의 생성과 관리, 제공에 있어 통합적 체계로의 변화, 민간데이터 범정부 공동이용 체계화, 데이터품질관리 범위 확대 등이 주요 과제로 제시되었다(공공데이터전략위원회, 2016.12). “데이터·AI경제 활성화 계획”에 의하면 양질의 데이터 개방을 위해 빅데이터플랫폼, 센터, 빅데이터 네트워크 협의회 등을 마련하고 국가연구데이터플랫폼과 같은 연구데이터 공유 및 활용 체계를 구축하며, 데이터맵, 국가중점데이터, 범정부 데이터 플랫폼을 제시하는 등, 공공데이터의 개방과 효율적 관리를 주요 과제로 제시하였다(관계부처합동, 2019.1.).

## 2.3. 관련 조직과 기능

데이터기반행정법을 기반으로 한 데이터 연계 및 공동활용은 해당 법이 2020년 12월에 시행된 관계로, 해당 기능을 수행하고 있는 조직에 대해서는 향후 우리 연구원의 별도 연구를 통해서 진행될 계획에 있다. 따라서 데이터기반행정 업무의 총괄 주무부처인 행정안전부와 해당 업무를 전담하고 있는 한국지능정보사회진흥원, 문화체육관광부의 데이터기반행정 전담부서와 문화체육관광부 산하 데이터전담기관의 주요 역할과 기능을 중심으로 살펴보도록 하겠다.

데이터기반행정법에 근거, 해당 업무의 총괄 주무부처는 행정안전부이다. 행정안전부에서 데이터기반행정업무 전담 부서는 공공데이터정책관 내 빅데이터분석활용과이다. 공공데이터정책관은 총 5개 부서로 구성되어 있다. 5개 부서는 데이터기반행정 실태점검 계획수립 및 시행업무, 빅데이터 분석 및 활용, 빅데이터 플랫폼 구축 운영 등을 주관하는 빅데이터분석활용과(데이터기반행정법 등), 공공데이터 업무를 주관하는 공공데이터정책과, 공공데이터유통과(공공데이터법), 이외에 공공지능정책과, 지능행정기반과로 구분할 수 있다. 데이터기반행정업무의 핵심 사업인 데이터 연계 활용 업무의 추진을 위해서는 빅데이터 분석이 필요한 사항이기 때문에 두 개의 업무를 별도로 분리하기 보다는 하나의 부서에서 통합적으로 관할하고 있음을 확인할 수 있다.

행정안전부의 데이터기반행정업무를 전담하고 있는 공공기관은 한국지능정보사회진흥원이다. 한국지능정보사회진흥원은 공공데이터 본부(공공데이터활용지원센터)를 두고, 하부 조직으로 데이터기반행정팀과 공공데이터 관련 4개 팀(공공데이터기획팀, 공공데이터개발팀, 공공데이터품질팀, 공공데이터활용팀) 등 5개 팀으로 운영하고 있다. 해당 본부의 경우, 공공데이터 대비 데이터기반행정업무의 비중이 다소 적은 편이나, 이는 향후 공공데이터 관련 팀을 세분화하여 운영하고 있는 것처럼 향후에는 데이터기반행정이 활성화될 경우 해당 업무에 대해서도 세분화가 될 수 있는 가능성을 예측해볼 수 있다. 한국지능정보사회진흥원의 데이터기반행정팀은 행정안전부의 빅데이터분석활용과와 업무 협조체제를 통해, 데이터기반행정법상의 주요 업무들을 수행하고 있다. 데이터기반행정 법제도 및 정책기획, 데이터기반행정 활성화 지원, 데이터기반행정 책임관 협의회 운영, 공동활용 데이터 등록관리시스템 등록 및 확대 구축 지원 등이 데이터기반행정법상에서 명시한 주요 업무를 전담관리하고 있다.

관련하여, 문화체육관광부의 데이터기반행정 업무를 전담하고 있는 부서는 기획조정실 내 정책분석팀이다. 문화체육관광부 정책분석팀의 주요 역할과 업무는 ① 문화체육관광부와 그 소속기관·공공기관의 통계·공공데이터·빅데이터에 관한 기획, 예산 검토·조정·작성, 활용 기반 구축 및 관리 총괄, ② 문화체육관광부와 그 소속기관·공공기관의 통계·공공데이터·빅데이터에 관한 품질관리 총괄, ③ 통계 및 수치자료가 활용된 연구결과물의 정확성 검토 등 문화체육관광부 통계기반 정책평가의 운영계획 수립 및 시행, ④ 국내외 다른 기관과의 통계·공공데이터·빅데이터에 관한 협력·지원에 관한 사항 총괄, ⑤ 그 밖에 문화체육관광부 관련 통계·공공데이터·빅데이터에 관한 사항 등으로 정리할 수 있다.

문화체육관광부에서는 문화데이터기반 문화정책 수립 및 활성화 지원 역할 수행을 위해 문화체육관광부 데이터분석센터를 한국문화관광연구원에 설치·운영하기로 하였다(2021.12.10.). 이에, 한국문화관광연구원은 정책정보센터 내에 데이터분석팀을 신설하였다(2021.3). 데이터분석팀은 주요 역할과 기능은 다음과 같다. ① 문화관광콘텐츠 분야 데이터정책 연구, ② 문화관광콘텐츠 관련 데이터의 수집, 등록, 제공, 관리 분석, 활용, ③ 문화관광콘텐츠 분야의 소비, 산업, 고용 등 동향 및 전망 분석, ④ 문화관광콘텐츠 관련 빅데이터의 가공, 분석, 활용, ⑤ 문화관광콘텐츠 분야의 데이터 거버넌스 구축, ⑥ 공공데이터의 수집, 제공, 활용, 관리, ⑦ 데이터 관련한 정책사업 및 연구보고서 발간 등이다.

문화체육관광부와 한국문화관광연구원은 향후 문화체육관광부 산하 소속·공공기관 및 타부처, 외부 민간업체와의 데이터 연계 사업 추진을 위해 우선적으로 문화체육관광부 데이터실무협의회를 개최하였다(2021.6, 2021.7). 데이터 연계 및 공동활용을 위해서는 각 기관별 보유 데이터 및 수요 데이터 등에 대한 현황화가 필요하며, 추가적으로 연계하고자 하는 데이터의 정비 및 표준화 또한 필수적으로 진행되어야 하기 때문에 단계적인 추진계획이 필요한 상태이다. 이와 관련해서는 부처 및 각 기관 간 지속적인 업무 공조를 지속해야 하며, 데이터 제공에 대한 협조 또한 원활히 수행되어야 하는 조건이 필수불가결하다.

마지막으로, 정부부처 및 관련 공공기관은 데이터기반행정 책임관을 임명하도록 되어 있는데, 2021년 9월 현재 총 812개(중앙 48, 지자체 260, 공공 504) 기관 책임관 임명 완료된 상태이다. 또한, 데이터기반행정 활성화 위원회에 참여하는 기관은 총 20개 기관으로 파악된다(제3차 데이터기반행정 책임관 협의회, 2021.9).

〈표 2-13〉 데이터기반행정 활성화 위원회에 참여 기관

구분	기관명	
중앙행정기관(14개)	기획재정부	보건복지부
	교육부	환경부
	과학기술정보통신부	고용노동부
	행정안전부	국토교통부
	문화체육관광부	해양수산부
	농림축산식품부	중소벤처기업부
	산업통상자원부	통계청
광역시도(2개)	경기도	경상남도
기초자치단체(2개)	서울특별시 도봉구	충청남도 논산시
공공기관(2개)	한국지능정보사회진흥원	한국지역정보개발원

출처: 제3차 데이터기반행정 책임관 협의회 개최 계획(2021.9) 인용

## 2.4. 요약

데이터기반행정을 중심으로 관련 법, 계획, 조직 등을 살펴보았다. 공공데이터법과 데이터기반행정법을 명확하게 구분하기 어려운 점이 존재한다. 공공데이터법은 제공과 개방에 초점을 맞추고 데이터기반행정법은 수집과 분석에 초점을 맞춘다. 데이터의 제공과 수집, 활용은 데이터 개방성과 데이터 개방 대상(개인 중심, 기관 중심 등)이냐에 따라 달라질 수는 있으나 그 경계는 상당히 모호하다.

데이터 연계와 관련된 계획은 기존의 공공데이터 기본계획과 공공 빅데이터 추진계획, 데이터·AI경제 활성화 계획 등을 살펴보았다. 각각의 계획에는 데이터 연계 활성화를 위한 추진체계 구축, 인력 양성 등에 대해 제시한다. 무엇보다 추진체계 구축이 중요한 이유는 데이터를 보유하고 있는 기관 간 협조체계에 따라 데이터 활용 정도에도 큰 편차를 나타낼 수 있으므로, 이 부분은 매우 중요한 사항이다. 향후 데이터기반행정법을 근거로 하는 데이터기반행정 업무가 기존의 공공데이터법 및 업무 등과 같이 다 년간 데이터 및 업무 노하우를 축적한다면, 2021년 현재 시점 대비 보다 구체적인 데이터 활성화 방안을 확인할 수 있을 것으로 예상된다.

조직적인 측면에서 볼 때, 정부부처 및 공공기관 내 데이터기반행정관이 812명 임명된 것은 고무적인 일이다. 기관별 특성에 따라 데이터기반행정관이 실제로 데이터 관련 분야 전문가가 아닐 수도 있으나, 관련해서는 조직 내 데이터 전담부서를 설치하고, 데이터 관련 전문 인력 보강 및 기존 직원 대상으로 데이터 업무 교육 직원 등을 통해 전문적인 데이터 업무를 수행할 수 있는 기반 마련이 필요하다.

## 제3절 데이터 공동활용 관련 선행연구 분석

---

### 1. 데이터 연계의 개념 및 의의

#### 1.1. 데이터 연계의 개념

데이터의 공동활용을 위해서는 서로 다른 데이터를 어떠한 방법으로 사용하는 지에 대한 부분에 대해서도 개념적 정리가 필요하다.

데이터기반행정법에서는 “공공기관은 데이터의 제공, 연계 및 공동활용을 적극적으로 수행하고, 그 성과가 유용하게 활용될 수 있도록 노력하여야 한다(데이터기반행정법 제3조제3호)”, “데이터의 연계·제공 및 공동활용에 필요한 체계 구축(데이터기반행정법 제6조제4호)”, “다른 공공기관 등과의 데이터 연계협력 방안(데이터기반행정법 제7조제5호)” 등과 같이 데이터의 공동활용은 데이터의 연계가 담보된 것으로 사용되고 있다.

데이터 연계와 관련해서는 다양한 용어가 사용되고 있다. 국내적으로는 연계, 융합, 통합, 매칭 등의 용어가 혼용되고 있고, 국외적으로는 linkage, fusion, integration, matching 등을 확인할 수 있다.

데이터 연계(Data Linkage)란 두 개 이상의 출처로부터 동일인이나 동일한 사건, 기관, 장소에 연관된 정보를 함께 가져오는 것을 의미하며, 식별자 등을 이용해 정보를 결합함으로써 단일 출처의 정보만으로는 알기 힘든 정보 요소 간의 관계가 밝혀질 가능성이 있다(개인정보보호위원회, 2017). 즉, 서로 다른 복수의 데이터 파일을 결합하여 보다 풍부한 정보를 제공해 줄 수 있는 하나의 완전한 통합데이터를 만드는 방법으로 정의될 수 있다(오미애, 2015).

데이터 연계는 레코드 연계(record linkage), 데이터 매칭(data matching), 데이터 통합(data integration) 등으로도 불린다(개인정보보호위원회, 2017). 데이터 연계 관련하여 사용하는 용어들은 유사한 특징이 나타난다. 통합은 연계와 유사한 의미로 활용

하기도 하지만(이민호 2015), 연계를 과정으로 통합을 연계 과정에서 발생하는 문제를 완화하는 것으로 다르게 정의하기도 한다(송명원 외 2006). 통합은 물리적 측면이 강조되기도 한다. 연계는 연결하는 과정적 측면이 강조되고, 매칭은 데이터 파일이나 데이터베이스를 결합하는 기술적 측면이 강조된다(배종면, 2016).

데이터 매칭은 기록(record), 개체 식별(entity resolution), 목표 식별(objective identification), 필드 매칭(field matching), 데이터 통합(data integration), 데이터 퓨전(data fusion) 등의 개념으로도 사용된다(오미애, 2015). 데이터 매칭의 개념은 정확 매칭(exact matching)과 통계적 매칭(statistical matching)으로 나누어서 정의할 수 있다. 서로 다른 자료에서 동일한 개체(same individual, family, event or place)를 결합하는 것을 정확 매칭(exact matching) 또는 기록 연계(record linkage)라고 하며, 서로 다른 자료에서 유사한 개체(similar individual, family, event or place)를 결합하는 것을 통계적 매칭(statistical matching)이라고 한다(윤건, 2019).

정확 매칭은 고유식별정보가 존재하여 동일한 개체를 연결하므로 행정데이터 간의 연계나 행정데이터와 조사데이터의 연계에 이용될 수 있고, 통계적 매칭은 고유식별정보가 존재하지 않아 이와 유사한 개체를 찾아 그 데이터를 결합하므로 조사데이터와 조사데이터의 연계, 행정데이터와 조사데이터의 연계(조사데이터에서 개인에 대한 정확한 정보를 얻지 못할 경우)에서 이용될 수 있다(오미애, 2015).

data fusion은 좀 더 거시적인 틀로 이해되며, data integration은 linkage와 유사하지만 linkage가 데이터 간 연결 측면이 강조된다. data matching은 데이터 통합 data integration의 일부로서 좀 더 기술적 측면이 부각된다(윤건, 2019).

〈표 2-14〉 데이터 연계 개념 정의 관련 국내외 선행연구

선행연구	활용개념	개념 정의
배종면(2016)	연계	다른 목적으로 구축된 전산자료원들 간에 정보들을 전산처리를 통해 상호 연결하는 과정
이은우 외(2017)	연계	두 개 이상의 출처로부터 동일인이나 동일한 사건, 기관, 장소에 연관된 정보를 함께 가져오는 것을 의미함. 데이터 연계는 레코드 연계(record linkage), 데이터 매칭, 데이터 통합(data integration) 등으로도 불림
김학래(2017)	연계	기준정보를 활용한 이종 데이터 간 연결을 뜻하며 이는 데이터의 통합을 위한 기반이 될 수 있음
박선우(2018)	융합	상이한 특성을 가진 데이터들을 상호 결합하여 데이터의 활용 범위를 넓히고 보다 깊은 함의를 도출



선행연구	활용개념	개념 정의
송명원·김상범(2006)	통합	분산되어 있는 각 정보시스템을 폐쇄하고 단일한 시스템으로 물리적 통합을 시도하는 것이며 시스템 간 연계과정에 복잡한 연계방식으로 인해 발생할 수 있는 장애요인들을 완화하는 것
오철호·윤상오(2008)	통합	개별기관의 업무처리 및 대국민서비스를 위하여 각자 보유하고 있는 데이터를 연계·공유하거나 통합하는 것
이민호(2015)	통합	정보의 요청과 제공이 전자적으로 용이하게 이루어질 수 있도록 데이터베이스의 연계 등이 가능한 수준
이영섭 외(2007)	매칭	별개의 데이터 파일을 결합하여 하나의 데이터 파일을 만드는 방법
오미애 외(2014)	매칭	서로 다른 복수의 데이터 파일을 결합하여 보다 풍부한 정보를 제공해 줄 수 있는 하나의 완전한 데이터를 만드는 방법
이성건 외(2017)	매칭	두 개 이상의 서로 다른 데이터베이스를 맞추어 결합하는 것으로, 데이터를 하나로 통합하는 데이터 퓨전에 사용됨. 즉 공통적인 속성을 기반으로 서로 다른 데이터를 맞추어 하나로 만드는 것
Christen & Goiser (2007)	Linkage	데이터의 질과 무결성을 향상시키고, 새로운 연구를 위해 기존 데이터를 재사용하도록 하며, 데이터 수집의 비용과 노력을 줄여주는 기술로서, 데이터 마이닝 시스템에서 패턴 감지에 이용
Palamuthusingam et al.(2019)	Linkage	동일한 사람, 장소, 가족이나 사건과 관련하여 다양한 소스로부터의 정보를 통합(uniting)하는 가치 있는 기술
The U.S. Department of Defense(1991)	Fusion	다양한 소스로부터의 데이터와 정보를 조합하여 자동적으로 감지하고 연관성과 상관성, 인과성을 분석하는 다층적, 다면적 과정
Mangolini(1994)	Fusion	서로 다른 특성을 가진 다양한 소스로부터의 데이터들을 활용하여 요청된 정보의 질을 향상시키는 방법, 도구, 수단의 세트
Wald(1998)	Fusion	서로 다른 소스로부터 기원하는 데이터를 연합(alliance)하는 수단과 도구로 표현되는 공식적 프레임워크
Lenzerini(2002)	Integration	서로 다른 소스에 있는 데이터를 연결(combining)하여 사용자에게 데이터를 통합된(unified) 시각으로 제공하는 문제
IBM(2019)	Integration	분리된 소스의 데이터를 연결(combine)하여 의미 있고 가치 있는 정보를 만들어내는 기술적/업무적 프로세스의 조합
Zhang & Meng(2007)	Matching	서로 다른 데이터셋을 상호운용 가능하도록(interoperable) 하는데 도움을 주는 근본적 측정 방식 중 하나
Christen(2012)	Matching	데이터 융합(data integration) 과정에서 공통적으로 요구되는 절차

출처: 윤건(2019), 데이터기반행정 강화 방안 연구, 한국행정연구원, 재인용

데이터 연계는 새롭거나 발전된 통계의 생산, 현재 일부 정보가 존재하는 측정값들에 대한 추가적인 정보의 생산, 단일 데이터 소스로부터 얻을 수 있는 것보다 더 많은 단위의 폭넓은 변수 활용, 기존 데이터 소스의 개선이나 검증 가능성, 응답자 부담감소 등의 측면에서 연구에 이익이 된다(개인정보보호위원회, 2017). 그러나 데이터 연계는 개인 식별에 관련된 데이터가 더욱 많아진다. 따라서 데이터 연계 시에는 관련 주요 정보에 대한 유출 및 공개 시의 위험성에 대해서 주의가 필요하다(개인정보보호위원회, 2017).

이처럼 여러 용어들 간에는 경우에 따라서는 위계가 있고, 유사성과 상이점 등이 존재한다. 그럼에도 불구하고 데이터의 연계에서 필요충분한 조건이 이루어지기 위해서는 서로 다른, 2개 이상 복수 데이터가 있어야 한다. 데이터의 원천, 종류, 특성 및 데이터 생성의 최초 목적 등이 다를 수 있다. 그리고 이러한 서로 다른 데이터를 일정한 목적 하에 연결시켜 하나로 만드는 것이다. 일정한 목적에 이용되기 때문에 가치가 있고, 의미가 있는 것이다. 데이터기반행정에서 데이터의 연계 및 활용 시에는 반드시 공공기관의 데이터가 있어야 한다. 복수의 데이터 중 한 가지 이상이 공공기관의 보유 데이터여야 한다. 이 경우, 연계 데이터가 민간 보유 데이터라고 하더라도 데이터기반행정에서의 의미하는 연계 및 공동활용에 해당된다.

## 1.2. 데이터 연계의 의의

2.1.절에서 데이터 연계에 대한 개념 파악을 파악하였다. 오미애(2015)는 연구에서 데이터 연계 및 활용 시 어떠한 유용한 점이 있으며, 그 결과를 어떻게 활용할 수 있을지를 정리하였는데, 해당 내용을 재인용하여 정리하면 다음과 같다.

첫째, 데이터 연계는 기존의 데이터를 보다 다양하게 활용 가능하도록 데이터 정보를 확장·제공해준다. 우리가 가지고 있는 데이터에 추가적인 열(column)정보를 더하여 더 많은 정보를 얻을 수 있게 해주어 더 복잡한 연구 문제를 해결할 수 있게 해준다.

둘째, 개인 이력을 통해 종단 분석을 가능하게 해 준다. 서로 다른 시간에 서로 다른 유형의 데이터를 매칭하는 이 메커니즘을 통해 시계열 데이터를 생성도 가능하다.

셋째, 데이터 연계는 설문 조사, 행정 자료의 정확성과 신뢰성을 검사하는 방법이 될 수 있다. 이는 연구자가 분석하고자 하는 데이터 셋에 대한 정보를 연계된 다른 데이터 셋의 정보와 비교해봄으로써, 자료에 대한 신뢰성을 검사할 수 있다.

넷째, 데이터 연계는 조사데이터에서 발생하는 결측치에 대한 정보를 제공해 줄 수 있다. 이는 다양한 데이터 연계 방법론으로 적용 가능하다.

마지막으로, 데이터 연계는 조사 데이터의 응답자 부담을 줄여주고, 설문 조사 비용을 감소시켜 준다. 행정데이터와 조사데이터의 연계는 응답자의 응답에만 의존하는 경우보다 정확한 정보를 얻을 수 있으며 이에 따라 데이터의 품질을 유지하는데 들어가는 시간과 비용을 절감할 수 있는 장점이 있다.

### 1.3. 데이터 연계 이슈

데이터기반행정법에서 규정된 데이터 연계 및 공동활용 관련 조항에서, 공공데이터 연계 관련 이슈가 제기될 수 있다. 먼저, 데이터기반행정에서 공공데이터 연계의 체계 부분이다. 데이터기반행정법에서는 활용하는 데이터는 공공기관이 생성하고 수집하는 데이터가 대상이다. 다시 말해, 데이터기반행정법 상의 데이터는 공공데이터라는 것이다. 따라서 상기 법에서 언급하는 데이터 연계는 공공기관의 데이터 연계를 의미한다. 정리하면, 데이터기반행정법에서 핵심 사항은 공공데이터를 필수조건으로 1개 이상 활용한 연계 및 공동활용이라는 것이다.

데이터 연계에서 중요한 이슈 중 하나는 데이터 연계 대한 목적과 필요에 대한 이해이다. 왜 데이터 연계가 필요한가? 이유는 보다 나은 분석 결과를 얻기 위한 것이다. 보다 나은 분석 결과는 결국 보다 나은 정책결정으로 이어질 것이기 때문에 공공데이터 연계가 중요하게 부각되는 것이다. 공공데이터 연계는 반드시 그 목적과의 관계를 끊임없이 고민해야 한다. 그렇지 않으면 실제 정책결정에 활용되지 않을 가능성이 높아진다. 수집된 데이터가 많더라도 그것이 정책에 별로 쓸모없는 데이터일 수 있듯이, 데이터 연계의 경우에도 데이터를 연계하고 결합하는 것이 실제 정책문제 해결에 도움이 되는 정보를 제공함에 있어 큰 도움이 되지 않는다면 데이터 연계의 의미가 없는 것이다(윤건 2019).

데이터기반행정법에서 데이터 연계는 데이터 제공 및 분석과 밀접한 관계가 있다. 데이터의 수집·저장·가공·분석·표현(데이터기반행정법 제2조제2항) 과정이 데이터기반행정법에서 데이터를 활용하는 절차적 단계이다. 즉, 데이터기반행정을 추진하기 위해서는 우선적으로 데이터의 존재가 필요사항이다. 그리고 이를 분석에 필요한 형태로 가공한다. 가공된 최종데이터를 분석하여 정책결정자가 활용할 수 있도록 표현하며, 최종적으로 정책결정자는 이를 반영하여 정책결정을 하게 된다.

데이터의 수집은 데이터의 제공과 직접적으로 연결되어 있다. 정책결정자는 여러 중앙행정기관이나 지방자치단체, 공공기관 재직 중에 하나이고, 자신이 가진 데이터만으로도 정책결정에 필요한 분석이 가능하다면 문제가 없겠지만 대체로 다른 기관의 데이터를 필요로 하고, 결국 수집을 위해서는 다른 기관의 제공 행위가 있어야 하는 것이다. 만일 해당 기관의 데이터 제공이 원활하지 않은 경우에는 데이터 수집이 어려워지는 것이다. 데이터 연계 및 공동활용을 위해서는 기존에 가지고 있는 데이터 외에 공공 또는

민간기관이 보유한 데이터가 필요하다. 이를 위해서는 데이터 보유 기관 간 상호 적극적인 데이터 제공과 교류가 이루어져야 한다.

데이터 연계 시 민간데이터 활용 이슈가 있다. 공공데이터는 행정기관이나 공공기관의 데이터이고, 민간데이터는 민간기업 및 시민단체, 일반 개인이 보유한 데이터이다. 데이터 생산·관리 주체에 따라 공공데이터와 민간데이터로 구분할 수 있다. 신용카드 데이터 혹은 통신사 데이터는 국민 개개인의 생활방식에 대한 정보를 담고 있다. 이에, 공공데이터와 연계·활용할 경우, 업무의 활용 영역이 확대된다. 데이터 연계는 공공기관 간의 데이터 공동활용만을 뜻하는 것이 아니다. 2개 이상의 공공데이터 간 연계, 혹은 공공데이터와 민간데이터 사이의 연계 등을 모두 포함하며, 그 중요도는 서로 다르지 않다.

마지막으로, 데이터 수집 확보, 기술 요건 검토와는 별도로 데이터 공개 기준 및 교환 방식 표준을 마련하여 데이터 연계 및 데이터 공동활용을 위한 선행 체계 수립이 필요하다. 다양한 리소스들과 데이터 유통을 위해서는 데이터 관리를 위한 체계적인 거버넌스 도입이 중요하다. 데이터 제공 기관 담당자가 공개한 정보를 관리할 수 있는 운영 관리 환경을 제공해야 한다.

데이터의 연계의 효율성 제고를 위해 데이터 연계 방식에 대한 기술표준 개발 및 배포가 필요하다. 데이터의 공개 및 연계에 관한 기술적 표준 개발을 통해 공공 및 민간 기관 간 상호 연계 효율성 제고를 해야 한다. 이를 통해 민간 기업의 데이터 개방 및 연계 관련 기술적 장벽을 낮추어 참여 확대 유도해야 한다. 데이터 개방에 따른 표준 가이드 뿐 아니라 민간 데이터 연계 활용을 위한 가이드라인 제공이 필요 하다. 해외 가이드라인은 공개된 데이터의 이용, 활용 및 활성화를 위한 다양한 솔루션 제공해야 한다. 국내 공공데이터 개방 표준은 정부기관이 공공데이터 개방 시 따라야 할 표준 가이드가 필요하다. 데이터의 연계 활용이 활성화 되어있는 해외 선진사례와 같이 공공 데이터의 개방을 위한 표준 제시와 함께 이용, 활용 및 활성화를 위한 솔루션의 제공도 중요하다.

## 2. 데이터 공동활용 개념

데이터 공동활용은 데이터기반행정법에서 핵심적 내용이다. 공공기관은 데이터의 상호연계성이 유지되도록 노력해야 하고, 데이터의 제공과 연계, 공동활용을 적극적으로 수행할 책무가 있다(데이터기반행정법 제3조). 3년마다 수립되는 데이터기반행정 활성화 기본계획에는 데이터 연계, 제공, 공동활용에 필요한 체계 구축과 전문인력 양성에 대한 내용이 반드시 포함되어야 하고(동법 제7조), 매년 수립되는 시행계획에도 다른 공공기관 등과의 데이터 연계, 협력에 대한 사항이 포함되어야 한다(동법 제8조).

또한 공공기관의 장은 데이터관계도를 체계적으로 관리할 책무가 있고, 행정안전부장관은 각 공공기관의 데이터관계도를 통합·연계하여 관리할 수 있으며(동법 제17조), 공공기관의 데이터 제공, 연계, 공동활용을 위한 데이터통합관리 플랫폼(이하 데이터플랫폼)을 운영하고 경우 데이터플랫폼의 효율적 운영을 위해 데이터 관리 시스템을 상호연계 혹은 통합할 수 있도록 하였다(동법 제19조). 또한 공공기관의 장은 데이터기반행정 책임관을 임명해야 하는데, 이 직책의 주요 임무가 데이터 연계, 제공, 공동활용에 관한 업무 총괄 및 지원이다(동법 제20조). 마지막으로 공공기관의 장은 데이터기반행정 실태 점검 및 평가를 자체적으로 수행해야 하는데, 이 경우에도 평가의 주요 내용 중 하나가 데이터 연계, 제공, 공동활용 성과이다(동법 제23조제1항).

데이터기반행정법이 본격적으로 시행(2020년 12월 10일)됨에 따라 행정기관 등 공공기관의 데이터를 활용할 수 있도록 공공서비스의 설계가 새롭게 진행되고 있다. 2021년 3월, 행정안전부는 데이터기반 과학적 행정을 통한 지능형 정부서비스 제공을 위해 향후 3년간의(2021~2023) 범정부 추진전략과 정책방향을 담아 「데이터기반행정 활성화 기본계획」을 수립·시행함을 밝혔다. 데이터기반행정법 시행 이후 처음 수립된 이번 기본계획에서는 ‘정책결정에 데이터를 적극 활용하는 과학적 행정 구현’, ‘국민이 신뢰하고 공감하는 지능형 행정서비스 제공’이라는 추진목표를 설정하고, 주요 추진전략(4대) 및 세부 추진과제(11대)를 마련하였다. 4대 주요 추진전략에서, 데이터의 활용 부분은 ‘데이터 공동활용’ 축진이 핵심사항이다.

〈표 2-15〉 데이터 공동활용 촉진

데이터 통합기반을 구축하여 데이터 공동활용을 촉진한다.

- 공동활용 가능한 공공데이터를 선별하여 기관 간에 공유하고, 카드매출·유동인구 등 민간데이터도 함께 활용할 수 있는 여건을 만든다.
- 또한, 정부 내 데이터 분석 및 공동활용을 지원하기 위해 데이터 통합관리 플랫폼을 구축하고, 교통·금융·문화·에너지 등 분야별 빅데이터 플랫폼과의 데이터 연계도 추진한다.

출처: 행정안전부 보도자료(2021.2.20.), “데이터로 국민체감형 공공서비스 새롭게 설계한다”

「데이터기반행정 활성화 기본계획」에 이어 2021년 5월 4일에 발표된, 「2021년 데이터기반행정 활성화 시행계획」은 데이터 관리 및 공동활용 기반 강화, 데이터 분석·활용으로 공공서비스 개선, 데이터기반행정 추진기반 및 역량 확충 등의 내용으로 구성되어 있다.

〈표 2-16〉 공공기관의 공동활용 데이터 등록 활성화

공공기관의 공동활용 데이터 등록 활성화

- 각 기관이 보유한 안전·복지·교통 등 분야별 공동활용 데이터 등록 추진
- 공동활용이 필요한 데이터를 지정하여 각 기관에 등록 요청
- 인터넷에 개방된 활용성이 높은 데이터를 선별하여 등록

출처: 행정안전부 보도자료(2021.5.4.), “데이터 기반 행정 활성화 위해 각 부처 데이터 칸막이 없앤다”

데이터기반행정에서 데이터 활용 부분의 주요 골자는, 데이터 통합기반을 구축하여 데이터 공동활용을 촉진하는 것이다. 정부 및 공공기관 상호 간에 데이터를 공동활용할 수 있도록, 공공분야 데이터인프라를 구축하고, 범정부적인 데이터관리체계를 구축이 포함되어 있다. 그동안 공공분야에는 방대한 데이터가 축적되었으나, 공동활용이 가능하지 않았던 관계로 데이터가 사장되거나 데이터가 비효율적으로 활용되어 왔다. 즉, 해당 시행계획의 데이터 공동활용 방안은 “공동활용 가능한 공공데이터를 선별하여 기관 간에 공유하고, 카드매출·유동인구 등 민간데이터도 함께 활용할 수 있는 여건을 만들겠다”는 것이다.

데이터를 공동활용하기 위해서는 선행적으로 진행되어야 할 사항들이 있다. 이에 앞서, 데이터 공동활용의 기초가 되는 기본 개념들에 대해 살펴보도록 하겠다. 이어, 기존에 진행되었던 데이터 공동활용에 대한 연구 결과들에 대해서 알아보고자 한다.

### 3. 데이터 공동활용 선행연구

앞 절에서 데이터 연계 등의 개념을 살펴보았다. 데이터는 다른 데이터와의 연계와 결합을 통해 보다 가치 있는 통찰력을 도출할 수 있다. 데이터기반행정에서 데이터는 그 어느 때보다 정책 의사결정의 중요 요소로 꼽히고 있다. 이러한 환경에서 다양한 분야의 데이터 간 상호 연계 필요성 증대되고 있다. 데이터기반행정의 핵심자원인 데이터의 공동활용을 위한 순환형 생태계 조성을 위해서는 데이터 간 연계를 통해 새로운 데이터 자원의 생성 및 제공이 필요하다. 데이터기반행정에서 데이터의 공동활용은 데이터 연계가 전제되는 조건이다.

데이터 연계는 다양한 새로운 가능성을 제시하고 있다. 본 절에서는 데이터 간 연계 연구 및 관련 방법론 적용 사례 등을 파악하여 데이터 연계의 시사점과 필요성을 되짚어 보고, 데이터 공동활용을 위한 함의를 제시하고자 한다. 기존의 선행연구는 첫째, 통계청의 공공데이터와 민간데이터 간 연계 방법 연구와 사례, 둘째, 기존의 데이터 연계 관련 연구 분석을 통한 시사점 도출로 나누어 살펴보하고자 한다.

#### 3.1. 공공데이터와 민간 빅데이터 연계 방법 및 사례 분석

앞서 ‘데이터 연계(Data Linkage)’에 대해 설명한 바와 같이, 데이터 연계는 두 개 이상의 출처로부터 동일인이나 동일한 사건, 기관, 장소에 연관된 정보를 함께 가져와서 식별자 등을 이용해 정보를 결합함으로써 단일 출처의 정보만으로는 알기 힘든 정보 요소 간의 관계를 밝혀내는 것을 뜻한다.

통계청에서는 새롭고 높은 부가가치 창출을 위해서 개별 데이터 활용 보다는 데이터 간 연계가 필요하다는 인식 확산에 따라 공공데이터와 민간데이터 간 연계 분석 사업을 진행하였다(통계청, 2016). 연계를 위한 전제로서 비식별화 조치에 관해 미국일본유럽 등은 물론 국내도 관계부처 공동으로 개인정보 비식별 조치 가이드라인을 발간 및 공포(행정안전부 보도자료, 2016. 6. 30.)하였다.

통계청의 데이터 연계 사업에서는 현행 법령 하에서 개인정보 침해가능성을 차단한 비식별 분석을 진행하였다. 비식별화는 데이터값 삭제, 범주화 등 개인정보의 일부 또는 전부를 삭제 또는 대체하는 것을 뜻한다.

〈표 2-17〉 개인정보 비식별 가이드라인과의 차이점

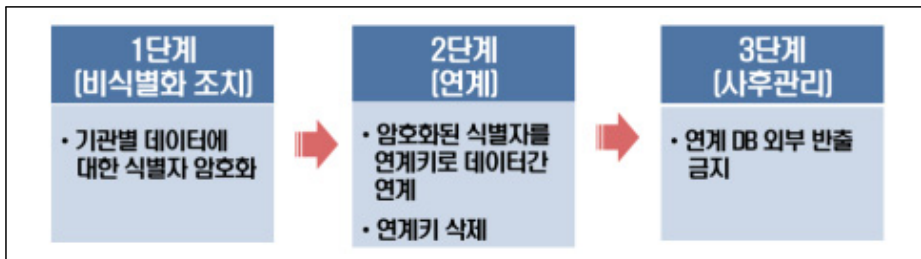
구분	통계청 연계 방식	가이드라인 방식
비식별화 형태	<ul style="list-style-type: none"> <li>식별자의 비식별화 조치</li> <li>속성 정보의 비식별화 미적용 ⇒ 구체적 분석 가능</li> </ul>	<ul style="list-style-type: none"> <li>식별자의 비식별화 조치</li> <li>속성 정보의 비식별화 적용 ⇒ 활용 범위 제한</li> </ul>
연계DB 관리방식	<ul style="list-style-type: none"> <li>물리적 공간 및 자료 활용 형태 제한 (예시: 집계표)</li> </ul>	<ul style="list-style-type: none"> <li>공유 기관 간 자료의 개별 활용 (제3자 제공 금지)</li> </ul>

출처: 통계청(2016.10), 공공 데이터와 민간 빅데이터 연계 방법론 및 사례

식별자에 대한 비식별 조치 방법에는 암호화, 총계처리, 데이터 삭제, 데이터 주화, 데이터 마스킹 등을 단독 또는 복합적으로 사용하였다. 암호화 처리는 복원이 불가능한 암호화된 가상번호로 대체한 후 주민등록번호, 외국인등록번호 등은 즉시 삭제하는 방식을 사용하였다. 속성자의 경우, 재식별화 위험 차단을 위해 특정 장소에서 처리 등 사후 관리 강화, 속성자 비식별화 등을 조치하였다.

데이터 공유의 한 주체가 정부임을 감안하여 완전하고, 실질적인 비식별화 조치를 시행하였다. 그간의 비식별화 경험과 각국 통계청의 비식별화 방식을 반영하여 단계적인 비식별화 조치를 마련하였다.

[그림 2-2] 비식별화 단계



출처: 통계청(2016.10), “공공 데이터와 민간 빅데이터 연계 방법론 및 사례”

연계를 위한 데이터 제공은 관련법상 제3자 제공 금지 규정으로 추가적인 보완 조치가 필요하였다. 통계청이 국내 금융회사의 신용정보를 공유하는 KCB(Korea Credit Bureau, 코리아크레딧뷰)에 데이터를 제공하면, KCB에서 데이터 간 연계 및 연계키 삭제 후 분석용 DB를 작성한다. 해당 DB는 양 기관이 공유하여 활용하는 방식이다. 이때 제공되는 분석용 DB는 개인정보는 아니다. 데이터 관리에 있어 데이터 보유 기관 간에



소관 데이터의 유출이 우려가 되는 부분은 자료 접근권자에게 인가를 주어, 통계청의 독립된 제한 공간(데이터센터)에서 자료 분석 및 통계를 작성하여 데이터 접근성을 최소화하였다.

통계청은 공공데이터와 민간기관 빅데이터 연계방법을 연구하고, 이 방법을 적용하여 아래와 같이 공공데이터(통계조사 및 공공행정자료)와 민간데이터(신용정보회사 데이터) 등 간 연계·분석을 진행하였다.

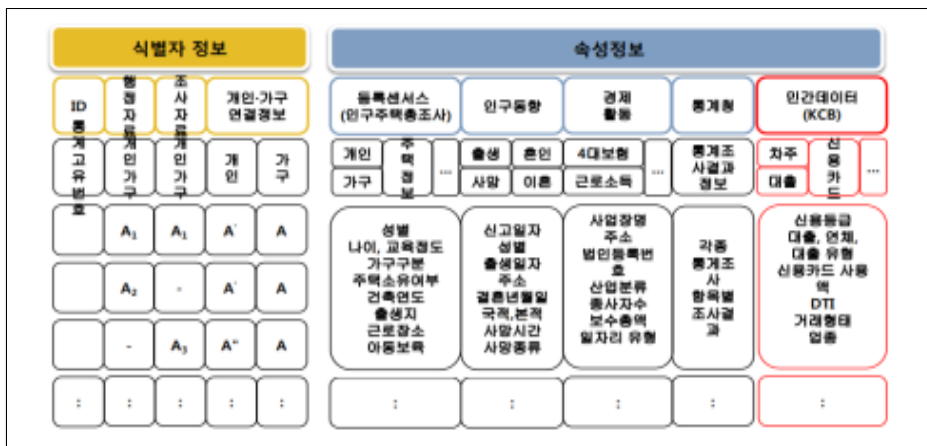
〈표 2-18〉 신흥부부 신용DB 구축 시 사용한 주요 데이터

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

출처: 통계청(2016.10), 공공 데이터와 민간 빅데이터 연계 방법론 및 사례

신흥부부 신용 DB 자료 연계 구성항목은 식별자 정보, 속성 정보로 나눌 수 있다.

[그림 2-3] 신흥부부 신용DB 자료연계 구성항목



식별자 정보는 ID통계고유번호, 행정자료(개인가구), 조사자료(개인가구), 개인·가구 연결정보(개인, 가구)로 구성되었다. 속성정보는 등록센서스(성별, 나이, 교육정도, 가구

구분 …), 인구동향(신고일자, 성별, 출생일자 …), 경제활동(사업장명, 주소, 법인등록 번호 …), 통계청(각종 통계조사 항목별 조사 결과 …), 민간데이터(신용등급, 대출, 연체, 대출유형, 신용카드 사용액 …) 등의 세부 항목들로 구성되어 있다.

통계청의 공공데이터와 민간 빅데이터 연계 연구 및 사례 분석의 결과는 다음과 같다. 5만 신혼부부 가구 단위의 부채 및 신용 유를 구축하였다. 개인별 신용 데이터와 가구 데이터를 연계하여 가구 단위의 신용평가 모델 활용이 가능하다.

해당 DB 분석 결과, 신혼부부의 세대구성에 따른 소득, 부채, 주택수, 자녀수 등 유의미한 결과를 도출하였다. 이 결과는 저출산 대책 등 신혼부부 지원을 위한 정책에 활용 가능하다.

〈표 2-19〉 신혼부부 횡단분석 및 종단분석에 따른 주요 결과

	구분	소득	부채	주택수	자녀수
횡단표본 (‘14년 기준)	전체 신혼부부	5,123만 원	4,273만 원	0.49채	0.68명
	부모와 함께 거주하는 신혼부부(부모 세대주)	4,431만 원	3,282만 원	0.35채	0.73명
종단표본 (‘14년 기준)	전체 신혼부부	5,348만 원	4,962만 원	0.58채	1.05명
	부모와 함께 거주하는 신혼부부(부모 세대주)	4,590만 원	3,645만 원	0.41채	1.18명

출처: 통계청(2016.10), 공공 데이터와 민간 빅데이터 연계 방법론 및 사례

## 3.2. 데이터 연계 관련 연구 분석 및 주요 시사점

앞서 3.1절에서는 통계청에서 진행한 공공데이터와 민간데이터 간 연계 연구·분석을 통해 데이터 공동활용에 대한 구체적인 사례를 살펴보았다. 본 절에서는 데이터 연계 활용 관련 기존의 선행 연구들을 요약하고, 데이터 연계 및 공동활용을 위해 필요한 제반 여건 및 필요사항 등에 관하여 정리하고자 한다.

오철호(2017)는 데이터시대의 정책분석평가 방법의 일환으로, 정부의 특정한 정책에 대하여 여러 부처가 관여하여 추진할 경우, 각 부처별 평가결과만으로는 범정부적 상황을 파악하는데 한계가 있다고 보았다. 이럴 때 각 부처별 평가결과, 즉 데이터를 연계하거나 통합하여 범정부적 평가결과 데이터를 만든다면 평가결과에 대한 종합적인 분석과 해석이 실질적으로 가능할 것으로 보았다. 데이터 연계는 새롭거나 발전된 통계의 생산,

현재 일부 정보가 존재하는 측정값들에 대한 추가적인 정보의 생산, 단일 데이터 소스로부터 얻을 수 있는 것보다 더 많은 단위의 폭넓은 변수 활용, 기존 데이터 소스의 개선이나 검증 가능성, 응답자 부담감소 등의 측면에서 연구에 이익이 될 수 있을 것이다(개인정보보호위원회, 2017).

둘째, 김근령(2018)은 빅데이터 분석과 관련하여 개인정보보호와 관련된 쟁점으로서 개인정보 개념의 포괄성 및 불명확성, 통지 및 동의 원칙 적용의 어려움, 빅데이터 구축이 개인정보 최소화 원칙, 목적 명확성 원칙에 배치되는 문제, 익명화 불가능과 재식별 가능성 문제, 익명화로 인한 데이터 가치 저하 문제 등을 꼽았다. 현재의 개인정보보호 관련 법제가 빅데이터 활용을 어려울 수 있으며, 따라서 데이터 연계를 위해서는 개인정보 침해가능성을 차단해야 한다는 관점이다. 즉, 데이터 연계를 위해서는 가장 먼저 개인정보보호 문제를 해결해야 한다. 데이터 연계가 수행되기 위해서는 데이터 보유기관으로부터 데이터 접근을 위한 승인을 받는 것에서부터 연구자 등 실제 데이터 이용자에게 개인정보 침해 위험을 최소화하는 방식으로 데이터를 제공하는 일련의 과정이 정비되어야 한다(개인정보보호위원회, 2017). 우리나라의 개인정보보호 관련 규제는 미국, 유럽, 일본이 보다 강력한 상태로 평가된다. 이로 인하여 데이터 활용과 개인정보보호 관련 법률이 상충하여, 데이터 연계 및 데이터 공동활용에 어려움 또한 존재하고 있다. 이에, 정보공유와 정보공개가 구별될 필요가 있으며, 공공부문 내에서 행정업무를 수행하는 업무 담당자 간의 정보공유에 대해 민간부문에 대한 정보공개와 동일한 수준의 보안조치나 개인정보보호조치를 요구하는 부분은 정보공유 활성화에 큰 저해요인으로 작용하고 있으며, 예외적 조항 및 완화된 법제도의 개선 등이 필요하다는 관점의 연구도 있다(이민호, 2015).

셋째, 박선우(2018)는 민-관 데이터 연계 사례로서 해외유입 감염병 차단, 서울시 심야버스노선 개발 사례, 한국교통연구원의 View 플랫폼 등을 제시하였다. 데이터 연계는 민간기업이 보유하고 있는 데이터를 정책에 활용하여, 그 가치와 활용성을 높일 수 있다. 따라서 데이터 연계는 공공데이터와 민간데이터 연계 활용이 중요하다. 우리나라 데이터 시장에서 민간데이터의 점유율은 공공데이터를 앞서며, 많은 부분에서 민간 영역이 데이터 시장을 주도하고 있다고 볼 수 있다. 공공데이터의 경우, 대부분 각 기관별로 데이터를 보유·관리하고 있기 때문에, 공공과 민간 간 데이터 연계는 민관의 업무협약 등을 통해 이뤄졌던 특정 교류의 형태가 많았다.

넷째, 윤희정(2014)은 빅데이터 활용과 비즈니스 효과에 미치는 조직적, 기술적, 환경적 요인을 분석하였다. 해당 연구에서 독립변수는 조직적 요인(CEO 및 구성원의 인식, 교육 및 인재육성), 기술적 요인(시스템 도입, 데이터 유형), 환경적 요인(투자 환경, 인프라 환경)이고, 종속변수는 기능별 활용성 및 비즈니스 효과이다(윤희정, 2014). 조직적 요인으로 CEO와 구성원 인식과 교육 및 인재육성, 기술적 요인으로 시스템 도입과 데이터 유형, 환경적 요인으로 투자 환경과 인프라 환경을 설정하였으며, 분석결과 모든 요인이 통계적으로 유의미하게 나타났다(윤희정, 2014). 조직적 요인의 가설 검정 결과, 'CEO 및 구성원의 인식'이 '기능별 활용성'에 영향을 미쳤고, '교육 및 인재육성'은 '기능별 활용성'과 '비즈니스 효과'에 영향을 미쳤다. 기술적 요인의 가설 검정 결과, '시스템 도입'은 '기능별 활용성'과 '비즈니스 효과' 모두에 영향력이 있었으며, 환경적 요인의 가설 검정 결과, '인프라 환경'이 '기능별 활용성' 및 '비즈니스 효과'에 영향을 미친다는 가설이 모두 채택되었다(윤희정, 2014). 데이터 활용 활성화와 그에 따른 분석 효과를 다양하고, 다각적으로 도출하기 위해서는 결과적으로 데이터 활용에 대한 조직 내 의사결정문화가 정착되어야 한다. 새로운 시스템을 도입에 대한 기관의 적극적인 지원, 임직원들을 대상으로 한 지속적인 교육을 통해 구성원들이 데이터 활용 이해와 수준을 향상시키고, 전문인력을 양성해야 한다.

## 제4절 소결

---

「데이터기반행정 활성화에 관한 법률」은 2020년 법제화되었다(2020년 6월 9일 제정, 동년 12월 10일 시행). 공공기관이 행정 및 연구, 정책 추진 과정상의 결과를 통해 보유·관리하고 있는 데이터를, 공공 부분의 행정과 정책 수립 및 의사결정에 활용하고자 함에 그 목적을 두고 있다. 궁극적으로는 데이터 기반의 과학적인 행정체계 구축·운영에 관한 제도적 기반을 마련하고 데이터 분석을 통해 국가적 현안, 사회문제 해결에 최적화된 대안과 미래 사회 변화의 선제적 대응 및 맞춤형 행정서비스를 제공 가능한 ‘데이터 기반의 과학적 행정 구현’을 목표로 한다.

본 연구의 ‘제2장 데이터기반행정 추진여건 및 환경 분석’에서는 데이터기반행정과 관련된 주요 사항에 대해 개념적 정의를 살펴보고, 데이터기반행정의 정책환경 분석과 관련 선행 연구에 대한 분석을 진행하였다.

첫째, 데이터기반행정법에서 정의하는 “데이터”는 “기계에 의한 판독이 가능해야 하며, 정형 또는 비정형의 정보”를 뜻한다. 이와 같은 맥락에서 공공데이터법과 문화체육관광부 데이터 규정법 등을 기반으로 “문화체육관광 분야 데이터”를 정의하였다. “문화체육관광 분야 데이터는 문화체육관광부 및 소속·공공기관에서 관계 법령 및 연구·조사·사업 등의 활동을 통해 생산, 관리, 취득하여 관리되고 있는, 기계판독이 가능한 정형 또는 비정형 형태의 정보”를 뜻한다.

둘째, “데이터기반행정”은 “공공기관이 생성하거나 다른 공공기관 및 법인·단체 등으로부터 취득하여 관리하고 있는 데이터를 수집·저장·가공·분석·표현 등의 방법으로 정책 수립 및 의사결정에 활용함으로써, 객관적이고 과학적으로 수행하는 행정”을 의미한다. 데이터기반행정은 ‘공공기관이 생성, 취득하여 관리하고 있는 데이터’를 활용하는데, 이와 같은 정의는 민간 데이터를 활용한 결과물 또한 데이터기반행정의 범위에 포함 가능하다. 데이터기반행정 관련하여 기존의 개념들과 비교하면 다음과 같다.

〈표 2-20〉 데이터기반행정 관련 개념의 정의와 특징

	정의	특징	한계
데이터기반 행정	데이터 기반의 객관적·과학적 행정체계를 구축하기 위한 제도적 기반	사회·환경변화를 대응해 행정의 책임성, 대응성, 신뢰성 향상	데이터기반행정의 범위가 포괄적이기 때문에 기능과 목적이 겹칠 수 있는 기존의 법 제도들과의 중복성이 있음
공공데이터	정부, 지자체, 공공기관이 전자적으로 생성 또는 취득하여 관리하고 있는 모든 전자화된 파일(DB)	다양한 공공기관이 생성 및 관리해왔던 공공 데이터의 통합적 활용을 통해 데이터기반행정의 주요 재원의 역할	부처별로 데이터 공유가 미진할 경우, 데이터가 구조화되지 않을 경우 활용도가 떨어짐
전자정부 제도	정보기술을 활용하여 행정기관 및 공공기관의 업무를 전자화하여 행정업무를 효율적으로 수행하는 정부	행정의 생산성과 투명성, 민주성 제고를 통해 대국민 행정서비스의 질적 향상 도모	하드웨어에 치중, 중복투자과 공동 활용 미흡으로 인해 비효율성 및 통합서비스의 어려움
증거기반 정책	과학적 증거에 기반을 두어 정책을 결정하려는 시도	신뢰성과 타당성 있는 정책과정을 통해 생성된 정보와 지식이 축적되어 더 나은 정책 형성 가능	과학적 증거 부재로 인한 적용 범위의 제한성, 담당자의 이해 부족에 따른 기술적 해석의 한계성 가능

셋째, 데이터기반행정의 정책적 환경은 국내외 환경적 영향에서 살펴볼 수 있다. 정부는 공공데이터법 시행(2013년) 이후, 사회·경제적으로 파급효과가 큰 국가중점데이터를 과감히 개방하여 데이터경제 활성화를 지원해왔다. 다만, 데이터 공동활용을 위한 법제도, 전략 등 추진기반이 미흡하였던 관계로, 정책 수립·의사결정 지원을 위한 공공데이터의 분석·활용도 여러 제약적 여건에 따라 목표만큼의 성과를 얻지는 못 하였다. 이러한 상황 가운데, 2020년 코로나 19 확산에 따른 위기상황 극복, 사회현안 해결에 대한 선제적 대응 등을 위해 정부의 맞춤형 서비스의 요구와 필요가 증대되었다. 그러나 데이터 기반의 과학적 행정의 필요성 대비, 현실적으로 다양한 분야의 행정업무와 데이터를 접목시키는 작업이 익숙하지 않고, 데이터 활용의 법적·윤리적 이슈를 파악하기 어려운 점이 있다. 따라서 데이터를 올바르게 이해하고 활용할 수 있는 기반환경 조성이 수반되어야 한다.

넷째, 데이터기반행정 추진을 위한 조직 정비가 필요하다. 데이터기반행정법이 2020년 12월에 시행되었기 때문에, 아직까지는 공공기관에서 이 법에 따른 데이터기반행정 업무를 전담하는 조직체계 구성이 미약하다. 기존의 공공데이터 연계 업무와 데이터포털 운영 현황 등을 통해 데이터기반행정 추진을 위한 조직 현황과 기능은 다음과 같이 정리할 수 있다. 다양한 행정기관들이 데이터포털을 운영하고 있다. 아직까지는 데이터 개방

중심으로 운영되는 데이터포털이 대다수이다. 데이터기반행정의 취지 상, 데이터는 일차원적 활용과 개방에서만 끝나는 것이 아니라, 다양한 분석과 연계 등을 통해 다면적인 인사이트를 도출할 수 있는 체계로 변화가 필요하다. 향후에는 데이터포털의 역할이 데이터 제공에만 한정하지 않고, 행정안전부의 ‘혜안(insight.go.kr)’과 같이 데이터 공동 활용과 전문 분석 서비스를 제공 가능한 시스템 구축으로 전환이 될 필요가 있다.

다섯째, 데이터 연계 관련한 이론적 분석 결과는 다음과 같이 정리하였다. 데이터기반 행정법에서는 ‘데이터의 연계·제공 및 공동활용’을 강조한다. 데이터 연계와 관련해서는 국외적으로는 fusion, integration, linkage, matching, 국내적으로는 융합, 통합, 매칭(결합), 연계 등의 용어가 혼용되고 있음을 확인할 수 있다(윤건, 2019). 용어들이 서로 위계가 있거나 약간의 차이가 있지만, 공통적인 특징은 서로 다른 데이터가 있어야 한다는 것을 전제로 하고 있다(윤건 2019). 데이터 연계(data linkage)는, 서로 다른 복수의 데이터 파일을 결합하여 보다 풍부한 정보를 제공해 줄 수 있는 하나의 완전한 통합데이터를 만드는 방법으로 정의될 수 있다(오미애, 2015).

여섯째, 데이터 연계 관련 기존 연구를 통해 몇 가지 해결해야 할 부분들을 확인할 수 있었다.

① 데이터 연계에서 중요한 이슈 중 하나는 데이터 연계 대한 목적과 필요에 대한 이해이다. 보다 나은 분석 결과는 결국 보다 나은 정책결정으로 이어질 것이기 때문에 공공데이터 연계가 중요하게 부각되는 것이다. 공공데이터 연계는 반드시 그 목적과의 관계를 끊임없이 고민해야 한다. 그렇지 않으면 실제 정책결정에 활용되지 않을 가능성이 높아진다. 수집된 데이터가 많더라도 그것이 정책에 별로 쓸모없는 데이터일 수 있듯이, 데이터 연계의 경우에도 데이터를 연계하고 결합하는 것이 실제 정책문제 해결에 도움이 되는 정보를 제공함에 있어 큰 도움이 되지 않는다면 데이터 연계의 의미가 없는 것이다(윤건 2019).

② 데이터행정법에서 데이터 연계는 데이터 제공 및 분석과 밀접한 관계가 있다. 데이터의 수집·저장·가공·분석·표현(데이터기반행정법 제2조제2호) 과정이 데이터행정법에서 데이터를 활용하는 절차적 단계이다. 즉, 데이터기반행정을 추진하기 위해서는 우선적으로 데이터의 존재가 필요사항이다. 그리고 이를 분석에 필요한 형태로 가공한다. 가공된 최종데이터를 분석하여 정책결정자가 활용할 수 있도록 표현하며, 최종적으로 정책결정자는 이를 반영하여 정책결정을 하게 된다.

③ 민간데이터와의 연계에 대해서도 해결해야 할 이슈들이 있다. 공공데이터는 행정 기관이나 공공기관의 데이터이고, 민간데이터는 민간기업 및 시민단체, 일반 개인이 보유한 데이터이다. 데이터 생산·관리 주체에 따라 공공데이터와 민간데이터로 구분할 수 있다. 그러므로 공공데이터의 연계를 위해서는 공공데이터에만 한정하지 않고, 민간데이터의 적극적인 연계도 중요하다.

제2장에서는 데이터기반행정 추진여건 및 환경분석을 통해 개념적 정리, 제도 및 조직 현황, 선행연구 분석 등에 대한 연구를 진행하였다. 데이터기반행정을 설명함에 있어 가장 기초적인 단위는 데이터이다. 본 연구의 핵심 키워드는 데이터 연계와 공동활용을 목적으로 하는 데이터기반행정이 중심이다. 데이터 공동활용의 기반 마련을 위한 연구 분석틀은 데이터, 데이터기반행정, 제도, 기술, 역량으로 구분하였다. 데이터 요인에서는 메타데이터 정보를 획득하고, 데이터 개방, 보유 및 수요 현황을 파악한다. 데이터기반에서는 데이터 연계 및 공동활용 사례 분석, 데이터 제약조건 및 상호 관계를 분석한다.

제도 요인의 경우 근거규정(법, 제도, 업무규정), 데이터 제공기관과 협력, 상위 차원의 조정체계를 살펴본다. 기술 요인은 데이터 표준화, 데이터품질관리, 데이터관리체계를 분석한다. 조직 역량은 데이터리더십, 데이터조직, 데이터예산, 데이터인력, 데이터문화 등을, 개인 역량의 경우 데이터경험, 데이터지식, 데이터인식, 데이터교육 등에 대한 역량을 분석한다.

〈표 2-21〉 연구 분석틀 설정

요인	세부항목	비고
데이터	데이터, 메타데이터, 데이터 개방 현황, 데이터 보유 현황, 데이터 수요 현황	
데이터기반	데이터기반, 데이터기반행정, 데이터 연계, 데이터 공동활용, 데이터 업무활용, 데이터 제약조건, 데이터 간 상호관계	
제도	근거규정(법, 제도, 업무규정), 데이터 제공기관과 협력, 상위 차원의 조정체계, 데이터 등록 및 제공 가이드라인	
기술	데이터표준화, 데이터품질관리, 데이터관리체계	
역량	조직 역량	데이터리더십, 데이터조직, 데이터예산, 데이터인력, 데이터문화
	개인 역량	데이터경험, 데이터지식, 데이터인식, 데이터교육



문화·관광·콘텐츠 데이터 공동활용 기반 구축 연구

제3장

데이터 공동활용 사례 분석



## 제1절 사례 분석 방법

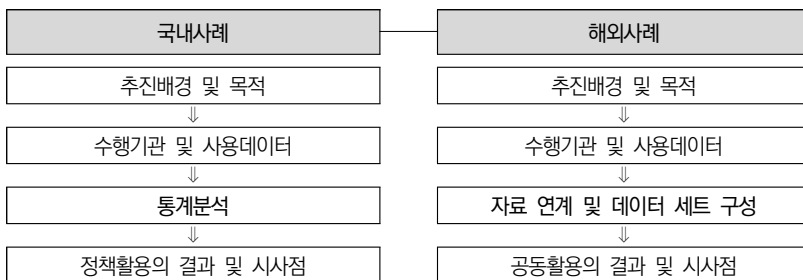
제2장에서는 데이터기반행정의 의의 및 데이터기반행정 정책환경 분석을 통해 데이터기반행정의 추진 여건 및 환경 분석을 실시하였다. 이러한 정책환경을 고려하여 문화관광콘텐츠 관련 데이터 공동활용을 실현하기 위해서는 유사한 정책환경의 데이터기반행정 및 데이터 공동활용의 사례들을 찾아보고 검토할 필요가 있다.

따라서 제3장에서는 국내외 데이터 공동활용 사례를 살펴보고 문화관광콘텐츠 데이터 공동활용 기반 구축을 위해 정책적으로 개선점을 도출하고자 한다.

사례분석은 크게 해외사례와 국내사례로 구분되는데, 국내사례는 정부부처 및 지방자치단체가 실시한 빅데이터 활용사업 및 데이터기반행정의 대표적인 사례들을 중심으로 살펴보면, 각 사업별로 해결하고자 한 문제가 무엇이었는지, 어떤 기관에서 어떤 데이터를 가지고 활용분석하여 문제점을 해결하였는지 등을 살펴보면 그 과정에서 겪는 어려움이 있다면 그것을 해결하고자 어떤 노력을 기울였는지를 살펴보고자 한다.

반면, 해외사례는 데이터 개방성이 높은 나라로 평가받고 있는 영국, 스웨덴, 네덜란드를 중심으로 데이터기반행정을 실현하기 위해 제도적, 환경적으로 어떻게 기반을 마련해 왔는지 살펴보고자 한다. 특히 각국의 정부부처가 협력하여 각자 보유한 행정 데이터를 통합 또는 연계한 데이터 세트를 만들었던 과정 및 데이터 개방 노력을 토대로 국내 사례와의 차이점을 살펴보고자 한다.

〈표 3-1〉 사례분석 방법



## 제2절 국내 사례

우리나라는 중앙정부부처와 광역자치단체 주도의 데이터기반행정 실현과 데이터 공동활용으로 행정의 효율성을 높이고 현장수요에 부합하는 과학적 행정의 제도적 기틀이 마련하고 있다. 특히 교통, 범죄예방·안전, 보육·교육과 같이 우리의 일상과 밀접한 관련이 있는 공공영역을 중심으로 데이터 공동활용을 통한 실효성 있는 정책들이 개발되어 국민의 삶의 질을 높이는 커다란 변화를 만들어가고 있다.

본 절에서는 정부부처와 지방자치단체가 데이터를 공동활용하여 정책 개발 및 개선을 한 데이터기반행정의 우수사례를 중심으로 살펴보고자 한다.

〈표 3-2〉 국내 데이터공동활용 주요 사례

기관	사업명	사용 데이터	분석과정	정책활용
경찰청·행정안전부	112 신고데이터 기반 범죄 위험도 분석	<ul style="list-style-type: none"> <li>•(경찰청)112신고 362만 건, 범죄통계</li> <li>•(소상공인진흥공단) 소상공인데이터</li> <li>•(인천시)항공사진</li> <li>•(SKT)유동인구</li> <li>•(카드사)카드매출자료</li> </ul>	<ol style="list-style-type: none"> <li>1. 범죄·무질서 위험도 예측모델 제시</li> <li>2. 월(月),일(日),2시간 단위 범죄·무질서 발생 위험지역 예측</li> <li>3. 범죄·무질서 발생에 영향 미치는 주요 환경적 요인</li> </ol>	(경찰청) 범죄예방시스템에 반영하고 치안자원을 효율적으로 분배하여 범죄 발생의 사전예방에 활용
교육부·보건복지부·국토교통부·창원시·청주시	신규 공동주택 초등돌봄시설 수요예측	<ul style="list-style-type: none"> <li>•(국토부)공동주택 구조·환경 전용면적, 주차대수, 난방종류 등</li> <li>•(자치단체)주민등록 데이터, 가계동향 조사 등</li> </ul>	<ol style="list-style-type: none"> <li>1. 아파트 단지의 초등학교 수와 맞벌이 비율을 산출하고,</li> <li>2. 돌봄이 필요한 초등학교 수 예측</li> </ol>	(국토부) 500세대 이상 신규 공동주택 단지 내 다함께 돌봄센터를 의무 설치하는 규정 개정 (복지부) 다함께 돌봄 센터 운영 가이드라인에 예측모델을 활용 명시
병무청·법무부·경찰청 등	병역면탈 의심자 포착	<ul style="list-style-type: none"> <li>•(병무청) 병역면탈자의 질병별 치료경과</li> <li>•(법무부) 출입국자료, 자격증 취득현황</li> <li>•(경찰청 등) 취업현황, 소셜데이터</li> </ul>	<ol style="list-style-type: none"> <li>1. 자격면허 제한 질환, 비정형적 증가 질환, 계속치료 등 분류,</li> <li>2. 병역면탈 의심 대상 추출</li> </ol>	(병무청)병역면탈자에 대한 기획수사 활용 및 병역면탈 방지를 위한 프로세스, 시스템 등 보완

기관	사업명	사용 데이터	분석과정	정책활용
서울특별시	심야 전용 시내버스 '올빼미 버스'	<ul style="list-style-type: none"> <li>• (서울시) 120 다산콜센터 60만 건, 유동인구, 교통수요량</li> <li>• (KT) 휴대전화 이력</li> <li>• (카드) 택시스마트 카드자료</li> </ul>	<ol style="list-style-type: none"> <li>1. 버스노선, 유동인구 및 교통수요 연구</li> <li>2. 지리정보시스템(GIS)에 시각화한 후,</li> <li>3. '심야버스 노선 수립 지원시스템'에 분석</li> </ol>	(서울시) 시민들의 야간활동을 지원하는 심야교통 수요가 많은 일부 지역에 심야버스 확대 운영

## 1. 경찰청, 112 신고데이터 기반 범죄 위험도 분석

### 1.1. 개요 및 추진 배경

전 세계적으로 경찰의 범죄 대응은 범죄수사와 같은 사후대응에서 범죄 예방을 위한 사전 예방 중심으로 패러다임이 바뀌고 있다(방준성 외, 2019). 시민의 안전과 범죄 예방 차원에서 설치한 지능형 CCTV의 증가로 그 효과가 나타나고 있으나 점차 지능화되는 신종 범죄가 출현하는 상황에서 보다 실효성 있는 범죄 예측을 통한 범죄예방에 대한 요구도 커지고 있다.

이에 경찰은 경찰 데이터 및 공공 데이터를 활용하여 통계 정보를 활용하여 범죄의 패턴과 추세를 분석한 범죄 예측 활동으로 스마트 행정 및 치안을 위한 노력을 기울이고 있다. 범죄 예측 기반의 경찰 활동은 자료 수집과 자료 분석, 경찰 수사 그리고 범죄대응의 4단계를 거쳐 이루어진다. 데이터 분석 결과 범죄 유형, 위험성 등에 따라 경찰 투입 인력을 조정하며 상황·단계별 맞춤 정보를 경찰에 제공한다(방준성 외, 2019).

행정안전부·경찰청·인천광역시는 이와 같은 맥락에서 주요 데이터를 공동활용한 무질서·범죄 위험도 예측모델 구축사업을 실시하였다. 대상지역은 송도, 청라 등 인천광역시를 대상으로 하며, ① 무질서·범죄<sup>1)</sup> 위험도 예측모델 설계, ② 월(月), 일(日), 2시간 단위의 범죄·무질서 발생 위험지역 예측, ③ 무질서나 범죄 발생 위험도에 영향을 미치는 주요 환경요인을 파악하는 과정으로 이루어졌다.

1) 범죄: 살인·강도·성폭력·절도·폭력(5종), 무질서: 주취자·시비 등 관련 112 신고 10종

## 1.2. 추진 내용

이 사업은 경찰청의 치안데이터와 해당 지역 인구수, 날씨, 요일, 실업률과 고용률과 같은 공공의 자료와 건물 노후 정도나 공원 숫자, 관광지 및 소상공인 업소의 숫자 등 민간의 다양한 자료가 공동활용되었다.

특히, 경찰청의 112신고·범죄통계, 소상공인시장진흥공단 보유의 소상공인 데이터(8만 건), 인천광역시가 보유한 항공사진(16.2GB), 그리고 SK텔레콤의 유동인구 자료(530만 건)·신용카드사 매출정보(521만 건) 등 다양한 데이터가 주요자료로 공동활용되었다.

[그림 3-1] 경찰청의 112신고 군집분석 결과

군집	군집 특성	통계 정보	대표 지역		
			항공사진	스트리트뷰	특징
1	연중 신고가 많으며, 주말과 심야(0시~2시대) 신고 증가	격차 개수			신고 건수 최상위 지역, ○○역 남부 유흥가
		17			
		평균 신고 건수			
		1,139			
2	연중 신고가 많으며, 23시~04시에 신고 증가	격차 개수			신고 건수 상위 지역, 주택가 중심 상권
		146			
		평균 신고 건수			
		386			
3	6월~9월에 신고 증가	격차 개수			신고 건수 중위 지역, △△공원
		992			
		평균 신고 건수			
		113			
4	8시~19시에 신고 증가	격차 개수			신고 건수 하위 지역, ◇◇단지
		6,218			
		평균 신고 건수			
		11			

출처: 경찰청, 2019

분석단계에서는 인천 지역을 가로 200m×세로 200m 크기의 2만 3천여 개 격자로 나누고 범죄·무질서 건수 및 비율을 기반으로 자동으로 군집화하는 K-means 알고리즘을 활용하여 인천지역을 5개로 나누고 군집분석을 실시하였다[그림3-1].

그 결과, 범죄·무질서 발생 상위 군집 지역은 유흥가 인접 지역으로 0~2시 사이, 범죄·무질서 발생 하위 군집 지역은 8~19시 사이에 112신고가 급증하는 것을 확인하였다.

한편, 경찰청은 LSTM<sup>2)</sup> 딥러닝 알고리즘, 그리고 그라디언트 부스팅<sup>3)</sup>기반의 Catboost 알고리즘을 활용하여 신고범죄 건수뿐만 아니라 환경적 요인을 결합된 범죄·무질서 발생 위험도를 예측모델 [그림3-2]을 개발하였는데, 이를 통해 최소 2시간 단위로 범죄 발생 건수 및 범죄·무질서 위험도 [그림3-3]을 산출할 수 있게 되었다.

[그림 3-2] 데이터 기반 범죄위험도 예측



출처: 경찰청, 2019

예측결과를 바탕으로 인천광역시에는 범죄·무질서 발생 위험지역 16개에 경찰관과 순찰차를 6주간 집중 배치하였다. 그 결과, 신고 건수는 전년 동기 대비 23.7% 감소하였으며, 범죄발생건수는 9.7% 감소한 것으로 나타났으며, 현장에 적용한 결과로도 실제 범죄 예방에 효과를 보였다. 이에 경찰청은 범죄위험도 예측모델을 인천 지역을 대상으로 시범 운영 후 전국적으로 확대 실시하였다.

2) Long Short Term Memory. 시계열 데이터를 처리하기 위한 딥러닝 모델

3) 약한 예측 모델을 결합하여 예측도를 향상시키는 기계 학습 모델

[그림 3-3] 인천시 △△구 ○○동 범죄위험도 5점 척도 예측 결과(예시)



출처: 경찰청, 2019

### 1.3. 의의

이 사업은 다양한 공공데이터의 공동 활용으로 국민이 삶의 변화를 체감할 수 있는 분석과제를 수행한 과학적인 데이터 분석과 활용의 대표적인 사례이다. 경찰청은 경찰력을 효율적으로 활용하기 위해 무질서·범죄 위험도 예측모델 구축사업이라는 목표를 가지고 관련 범죄율에 영향을 미치는 관련 데이터를 수집, 분석하였으며 이 결과를 토대로 높은 시간대·지역에 순찰차·경찰인력을 집중 배치하여 제한된 경찰력을 효율적으로 활용하고 범죄발생률을 낮추는 긍정적인 결과를 가져왔다.



## 2. 교육부·보건복지부 등, 신규 공동주택 초등돌봄시설 수요예측

### 2.1. 개요 및 추진 배경

행정안전부·교육부·보건복지부·국토교통부는 아파트 시공 전 초등학생 돌봄수요를 예측해 이에 부합하는 ‘다함께돌봄센터’를 설치하는 ‘초등돌봄 수요예측 모델’ 사업을 실시하였다. 신규 공동주택 초등돌봄시설 수요예측은 신규 아파트를 지을 때에 그 지역의 공공데이터를 적극 활용하여 돌봄이 필요한 초등학생 수를 예측하고 그에 따른 돌봄서비스를 제공하기 위한 사업이다.

### 2.2. 추진 내용

교육부는 '19년 10월 2020학년도 초등학생들의 돌봄수요를 파악하여 지역내 초등학교 대상 돌봄서비스를 확대 및 구축을 위한 목적으로 범부처 합동의 '2020학년도 범정부 초등돌봄 수요조사'를 실시하였다. 수요조사는 학교, 유치원, 어린이집을 통해 실시되었으며, 전국 초등학교 재학생(1~5학년), 그리고 이듬해 초등학교 입학예정인 유치원/어린이집에 재원 중인 예비취학 아동(만5세 아동)을 둔 학부모(보호자포함)를 대상으로 진행되었다. 또한 교육부는 창원시와 청주시로부터 전용면적, 가구당 주차대수, 건폐율 등 아파트의 구조와 관련한 데이터, 초등학교까지의 거리 등 아파트 주변환경 데이터 등 지자체 내의 아파트 구조, 아파트 환경과 관련된 다양한 데이터<표3-3>를 제공받았다(행정자치부, 2017).

이후 행정안전부는 교육부의 데이터와 통계청의 동별 국내인구이동통계, 가계동향조사, 행정안전부의 행정구역의 인구데이터 등 총 290종의 변수를 공동활용하여 '20년 1월부터 약 6개월간 신규로 만들어질 아파트 단지의 초등학생 돌봄 수요를 예측하기 위해 공공데이터 분석을 수행했다.

초등학생 수와 상관분석을 수행한 결과, 아파트의 전용면적이 크고 가구당 주차대수가 많을수록, 주변 종합병원 등 인프라를 갖춘 환경일수록 아파트 거주 초등학생 수가 증가하며, 초등학교가 멀거나 아파트의 건폐율이 클수록 초등학생 수가 감소하는 것으로 나타났다(행정자치부, 2017).

〈표 3-3〉 공동주택 구조·환경 및 인구 관련 데이터 목록(1)

제공기관	데이터명	변수 개수	주요항목	출처	기준연도
공동주택	아파트 구조	15종	건폐율, 층수, 주차대수, 방 개수, 난방 종류, 시공사, 전용면적 등	지자체	2020
	아파트 환경	20종	종합병원까지 거리, 편의시설 유무 등		2020
인구	국내인구이동 통계(동별)	240종	전입, 전출입, 연령세대별 전입, 전출입	통계청(MDIS)	2018
	행정동 인구통계	15종	동별 전체 인구수 및 초등학교 인구수	행정안전부	2019

출처: 관계부처합동, 2020

〈표 3-4〉 공동주택 구조·환경 및 인구 관련 데이터 목록(2)

종속변수 관련	데이터명	변수 개수	주요항목	출처	기준연도
초등학교 수	주민등록인구	1종	초등학교인구(만 6~11세)	지자체	2020
맞벌이 비율	가계동향조사	1종	전용면적, 자녀유무, 맞벌이 여부 등	통계청(MDIS)	2019

출처: 관계부처합동, 2020

〈표 3-5〉 입력변수 및 출력 예시

ID	건물명	주소	공공연도	행정동명	인구, 6세, 11세	초등학교	건설사	거리	주차대수
1	우미컨에듀파크1단지	충청북도 청주시 상당구 상당구 수영로 360(충청동 1286, 우미컨에듀파크)	2017	충청1동	26	신성초등학교	우미건설	236	142
2									
3									
...									
※본예시 공동주택단지 기준 1. 층도 : 아파트 2. 세대수 : 100세대 이상 3. 공공연도 : 2010년 ~									

출처: 관계부처합동, 2020

〈표 3-6〉 가계동향조사 가구유형별 맞벌이 비율

초등학교	18평 이하	18평 초과 - 24평 이하	24평 초과 - 26평 이하	26평 초과
무	19.8%	26.6%	32.6%	30.2%
유	38.2%	43.5%	47.2%	40.2%

출처: 관계부처합동, 2020

이처럼 초등학생 수 예측을 위한 유의미한 것으로 확인된 변수를 머신러닝 알고리즘<sup>4)</sup>에 적용한 결과, 초등학생 아이의 보육환경에 중요한 특성을 알고리즘이 자동으로 도출하여 아파트 단지의 초등돌봄 수요<sup>5)</sup>를 예측했다(그림3-4).

그 결과, 500세대 규모의 신규 공동주택 단지의 경우 25명 내외의 돌봄이 필요한 초등학생 수가 산출되어 최소 1개 규모의 다함께 돌봄센터가 필요한 것으로 분석되었다.

[그림 3-4] 지역별 초등돌봄 수요예측 사례



출처: 관계부처합동, 2020

4) RandomForest, SVM(Support Vector Machine) 등의 알고리즘을 적용하였고, 예측치와 실제값의 상관계수는 0.94로 유의미하게 나타남.

5) 예측된 초등학생 수와 통계청 가계동향조사 데이터를 이용해 추정된 맞벌이 비율로 최종 산출.

또한, 다함께 돌봄센터 규모산정을 위해 세대수·아파트 면적 등 해당 단지의 주요 데이터를 등록하면 해당 지역의 돌봄이 필요한 초등학생의 수를 산출하는 예측모델을 개발하였다.

이러한 분석결과를 바탕으로, 교육부는 ‘온종일 돌봄 확대를 위한 아파트 주민공동시설 활용방안’을 보건복지부 등 관련 부처와 함께 적극적으로 추진할 계획이다(행정자치부, 2017).

그 결과를 토대로 국토교통부는 500세대 이상의 신규 공동주택 단지 내 다함께 돌봄센터를 의무 설치하는 「주택건설기준 등에 대한 규정」을 개정하였고, 보건복지부는 지자체에 제공하는 다함께돌봄센터 운영 가이드라인에 예측모델을 활용하도록 명시하여, 지자체에서 ‘초등돌봄 수요예측모델’을 활용하여 다함께돌봄센터를 설치·운영하도록 권고하고 있다.

## 2.3. 의의

이 사업은 다부처가 연계된 국정과제 추진을 공동으로 지원하여, 정확성이 높은 수요예측을 통해 사회비용 및 정부예산집행의 효율성까지 얻을 수 있었다는 면에서 데이터 기반행정의 긍정적인 사례라고 볼 수 있다.

# 3. 병무청, 병역면탈 의심자 포착

## 3.1. 개요 및 추진 배경

병무청은 병역면탈자에 대한 기존의 기획수사 및 병역면탈 방지를 위한 제도 등을 보완하기 위해 공공데이터의 공동활용을 통한 빅데이터 분석을 추진하였다. 비정형증가질환, 신규수법 증가 등으로 면탈범죄 적발 어려움이 커지는 상황에서 제보에만 의존하다 보니 의심자 포착까지 긴 시간 소요되었기 때문이다.

게다가 병무청 병역조사과에는 베테랑 조사관이 30여명 있었으나 조사대상만 33만여명이 되다보니 전체 병역 면탈자 조사에 한계가 있었다.

특히 병역면탈자로 의심되는 사람의 질병, 취업, 자격증과 같은 다양한 정보를 복합적으로 분석해야 하는데 기획수사 명단 수기 추출(엑셀)과 병역면탈 데이터의 실시간 공유 불가 및 이력관리 미비로 수십만 건에 달하는 자료를 일일이 대조할 수 없기 때문이다. 이에 병무청은 '17년 축적된 병역면탈 자료를 분석하여 병역면탈자의 질병별 치료경과, 출입국자료, 자격증 취득현황, 취업현황, 소셜데이터 등 다양하게 축적되어있는 데이터를 복합적으로 분석할 수 있는 모델을 개발하였다(행정자치부, 2017).

### 3.2. 추진내용

병무청은 과거 7년 동안의 34만 건의 병역면탈자 데이터를 기반으로 병역감면자에 대한 신체검사(연도별, 질환별, 의사별 증감추이), 자격/면허 정보 등을 분석하였다. 특히, 자격면허 제한 질환, 비정형적 증가 질환, 계속치료 등 병역감면사유를 분류하고 병역면탈 의심 대상을 추출하였다. 이렇게 만들어진 병역면탈 의심자 분석 모델은 법무부, 국민건강보험공단, 국민연금관리공단, 경찰청 등으로부터 제공된 데이터를 공동활용하여 이루어졌다(표 3-6).

분석방법은 약 31만 명의 병역의무자를 대상으로 질환, 병원, 지역 정보 간 군집분석 및 카이제곱검증을 실시하여 상관관계 및 변수를 도출하였으며, 의사결정나무분석 및 Random Forest, Gradient Boosting 방법론을 적용하였다(그림 3-5).

〈표 3-7〉 병역면탈 의심자 포착 관련 데이터 목록

제공기관	데이터 종류	주요속성
병무청	기본정보	- 주소, 학력
	병역감면정보	- 병역감면정보
	적성정보	- 자격, 면허
	의무자 자원상태	- 주소, 직업
	민원출원	- 민원명, 출원일자
	민원상당	- 상담유형, 상담일자
	병역면탈 조장 불건전 정보	- 계시일자, 병역처분코드_과목, 병역처분코드_질병사유
법무부	출입국정보	- 출입국일시, 법무부출입국구분코드
국민건강보험공단	진료이력/자격정보	- 건강보험가입이력요청번호, 진료일자, 요양기관명
국민연금관리공단	취업정보	- 국민연금가입이력요청번호, 취득일자, 상실일자
경찰청	운전면허취득정보	- 자격면허코드, 자격면허취득일자, 자격면허취소일자
산업인력관리공단	국가기술자격취득정보	- 자격면허코드, 자격면허취득일자, 자격면허취소일자

출처: <https://www.bigdata.go.kr/analysis17.html>

[그림 3-5] 병역면탈 의심자 포착 관련 분석방안 및 분석결과



출처: <https://www.bigdata.go.kr/analysis17.html>

### 3.3. 의의

병역면탈 의심자 분석 모델은 조사관의 업무를 획기적으로 향상시키는 등 병무청의 능동적 기획수사와 내부 프로세스, 시스템, 제도 보완에서 성과를 거두었다. 또한 병역면탈 범죄 예측 및 포착 능력을 제고하여 우리 사회의 병역면탈 부정을 감소시키는 성과를 거두었다.

이 사례는 공공데이터를 공동활용하면서 빅데이터 분석을 통해 병역면탈 위험군 도출을 통한 효과적인 기획수사 성과를 가져온 사업으로 향후 병역면탈범죄 및 타 기관 소관 범죄 사실 발견, 공조수사에 적극 향상될 것으로 기대되며, 궁극적으로 공정병역 이행 및 병역이행자가 존중 받고 우대받는 사회적 분위기를 조성하는데 기여할 것으로 보인다.

한편, 병무청 병역면탈자 데이터는 ‘2017년도 공공 빅데이터 신규 분석 사업’의 일환으로 추진되었으며, 자료의 연계 작업 없이 자료의 통합을 통한 전통적인 방식의 통계분석을 통해 상관관계 등을 확인하는 방식으로 이루어졌다는 특징을 가진다.

## 4. 서울특별시, 심야 전용 시내버스 ‘올빼미 버스’

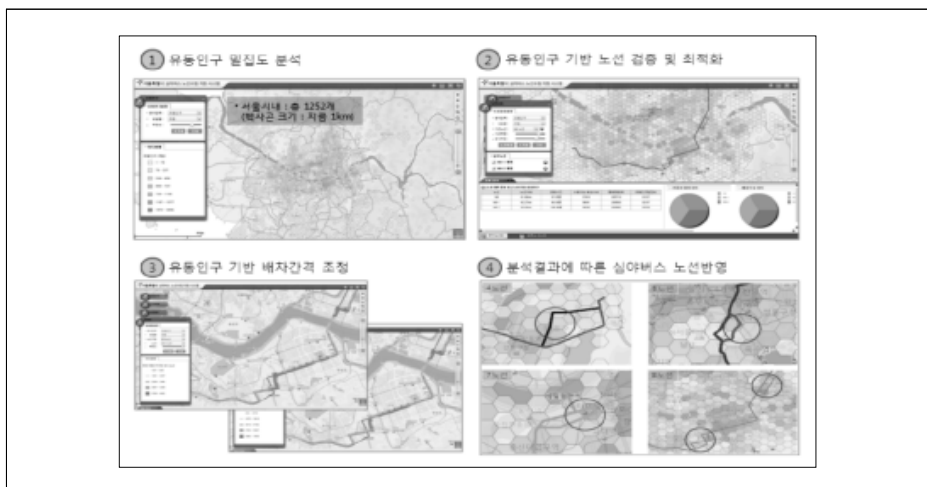
### 4.1. 개요 및 추진 배경

우리나라 정부의 데이터기반행정 사례 중 대표적인 사업으로 서울시의 교통정책인 ‘올빼미 버스’가 있다. 2013년 서울시는 ‘120다산콜센터’에 수집된 데이터 60만 건을 분석하였는데, 그 결과 교통(25.5%)이 시민들이 가장 많은 관심을 보인 정책분야라는 것을 확인하였다. 이에 서울시는 시민들의 야간활동을 지원하고 시민 만족도 향상시키는 것을 목표로 심야교통 수요가 많은 일부 지역에 버스노선을 개설하여 시범운행을 결정하였다.

### 4.2. 추진 내용

서울시는 관련 정보 수집을 위해 2013년 3월 한 달간의 주식회사 KT 휴대전화 통화 이력 데이터(30억 건)와 심야택시 승하차 데이터(500만 건)를 결합하였다. 그 결과 심야의 유동인구는 가장 많은 지역은 홍대이고 동대문, 신림역, 강남, 종로, 가락시장, 신촌, 남부터미널 등이 뒤를 이었다. 심야 택시 승·하차 데이터를 분석한 결과에선 강남이 교통수요가 가장 높았다. 이어 신림역, 홍대, 건대입구, 동대문 등 순이었다(이신 외, 2017).

[그림 3-6] 서울시 심야버스 분석 프로그램 구조도



출처: 서울특별시, 2013

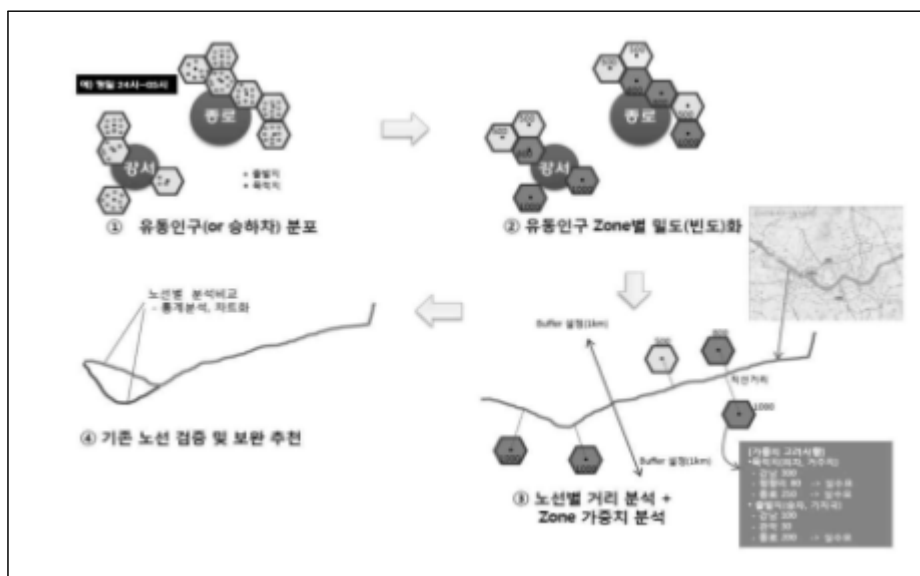
그리고 서울 전체 지역을 1km 반경을 기준으로 1,250개 셀 단위로 나눈 후 각각의 유동인구·교통수요량을 표시하였다. 그리고 기존의 일반 버스노선과 시간·요일별 유동인구와 교통수요 패턴을 분석하였다. 이후 분석된 유동인구 패턴을 지리정보시스템(GIS)에 시각화한 후 개발된 ‘심야버스 노선수립 지원시스템’으로 분석하였다[그림3-6].

그 결과 통화량 데이터와 교통카드 데이터를 분석한 결과 값이 실제 심야 유동 인구와 매우 일치하는 것을 확인하였다. 이에 노선 부근 유동인구 가중치를 계산하는 등 재분석 과정을 거쳐 최적의 노선과 배차간격을 도출했다(서울시, 2013)[그림3-7].

버스노선은 통신사의 휴대전화 통화이력으로 ‘출발지’, 택시 스마트카드 데이터로 ‘목적지’를 선정하여 유동패턴 분석을 통해 2개 노선을 결정되었다. 3개월간 동안 2개 노선을 시범 적으로 운행한 결과, 2개 심야버스 노선 이용객이 하루 평균 2,098명, 주요 이용 시간대는 오전 1시~3시로 전체 승객의 60%가 집중되었다.

한편, 두 차례의 서비스 만족도를 실시한 결과 1차 조사시 일반 시내버스 74.3점, 올빼미버스 80.15점(승객 500명 대상), 2차 조사시 응답자의 88.4%가 '확대운행을 원한다'(1,000명 대상)고 응답하여 시민들의 만족도가 상당히 높은 것을 확인하였다(이신 외, 2017).

[그림 3-7] 서울시 심야버스 심층분석도





이 결과를 바탕으로 서울시는 2013년 6월 시민공모를 통해 서울시 심야버스 고유브랜딩으로 '올빼미버스'를 선정하였고, 2013년 9월부터 9개 노선으로 확장하여 정식운행을 시작하였다. 정식으로 확대 운행이 시행된 이후에는 50일 동안 총 30만 4천명, 하루 평균 6,079명이 이용이 집계되었는데 이는 하루 평균 버스 1대 당 138명이, 정류소 1개소 당 5.8명이 이용한 셈으로 주간에 운행되는 일반 시내버스 1대 당 일 평균 이용승객인 110명과 비교하면 25% 이상 많은 수준으로 나타났다(성지은 외, 2014).

### 4.3. 의의

이 사업은 시민이 필요로 하는 심야시간 대에 버스서비스를 제공한 사업으로 민관의 데이터를 공동으로 활용하여 이를 활용하는 서비스 주체의 실제 행동을 이해하고 이들의 수요를 반영하는데 활용했다는 점에서 의의가 있다(이신 외, 2017).

한편, 서울시는 이러한 성과의 성공요인으로 데이터 보유 부서와 데이터 다루는 부서의 협력을 꼽는다. 특히 교육과 워크숍을 수차례 실시하여 부서간의 공감대를 이끌어내고 적극적인 협력의지를 이끌어내는 것이 매우 중요했다고 지적한다.

또한 심야버스 기획 단계부터 전문가 회의를 통해 데이터 활용가능성을 검토하여 공동 추진전략을 마련하였는데, 이때 민간기관, 학계, 연구기관 등 외부기관 전문가들과 의견을 나눔으로써 더욱 효율적인 해결점을 찾아갈 수 있었다고 평가하고 있다.

한편, 서울시는 공공데이터를 활용하는 과정은 IT 기술 외에도 행정 절차가 더 복잡하고 어려울 수 있다고 지적한다. 다양한 이해관계를 가진 시민을 대상으로 하는 정책을 위한 행정절차는 본질적으로 훨씬 복잡하기 때문이다. 서울시도 초기에는 어려움이 많았으나 심야버스 노선, 노인 복지시설 입지 분석 등 빅데이터 분석 경험을 바탕으로 데이터 기반행정절차를 효율적으로 관리할 수 있는 행정프레임워크를 개발하였다. 이는 41개 작업 아이템(tasks)과 14개 활동 아이템(activities)으로 분류되었으며, 총 5단계로 이루어져 있다(이신 외, 2017). 이 행정프레임워크는 실제 서울시가 추진 중인 사업을 통해 검증될 것이며, 앞으로도 다양한 데이터 관련 사업들에서 지속적으로 활용될 예정이다.

## 제3절 해외 사례

주요 선진국들은 통계개혁 추진을 통한 통계생산 기반 구축을 통해 데이터 공동활용을 준비하는 노력을 기울여왔다. 그 중에서도 영국, 스웨덴, 네덜란드는 공공부문에서 열린 정부와 오픈데이터를 표방하여 각 부처의 주요 데이터를 연계한 데이터세트를 구성하였고 이를 활용하여 데이터기반행정을 위한 정책연구 및 정책발굴을 추진해왔다.

본 절에서는 영국, 스웨덴, 네덜란드의 각 부처의 데이터를 연계 노력, 그리고 이를 공공의 정책연구에 활용한 결과를 살펴보고 정책적 시사점을 얻고자 한다.

〈표 3-8〉 해외 데이터공동활용 주요 사례

기관	기관	사업명	연계 데이터
영국	<ul style="list-style-type: none"> <li>영국 행정데이터 연구 (ARD UK)</li> <li>통계청(ONS),</li> <li>유니버시티 칼리지 런던(UCL),</li> <li>국가의료서비스 디지털 (NHS Digital),</li> <li>교육부(ED)</li> </ul>	<ul style="list-style-type: none"> <li>ECHILD 데이터 세트</li> </ul>	<ul style="list-style-type: none"> <li>(ED) 국가 학생 데이터(NPD)</li> <li>(ONS) PDS</li> <li>(NHS) HES</li> <li>(ONS) 사망률데이터</li> <li>※ 영국의 어린이 및 청소년 (25세 이하) 약1,600만 명</li> </ul>
스웨덴	<ul style="list-style-type: none"> <li>통계청(SCB),</li> <li>보험청(RFV),</li> <li>혁신청(VINNOVA)</li> </ul>	<ul style="list-style-type: none"> <li>LISA(노동시장 장기통합 데이터)</li> <li>*구LOUISE(교육, 소득과직업 정보가 포함된 시계열데이터)</li> </ul>	<ul style="list-style-type: none"> <li>(RAMS) 등록 기반 노동 시장 통계</li> <li>(RTB) 총인구</li> <li>(IoT) 소득세</li> <li>(RFV) 연금,사회부조와사적연금</li> <li>(SCB) 인구및주택센서스</li> <li>※ 16세 이상(2010년부터는15세 이상) 모든 스웨덴 국민</li> </ul>
네덜란드	<ul style="list-style-type: none"> <li>통계청(CBS),</li> <li>기업청(RVO.nl),</li> <li>환경평가청(PBL),</li> <li>토지등록부(Land Registry)</li> </ul>	<ul style="list-style-type: none"> <li>VIVET Programme</li> </ul>	<ul style="list-style-type: none"> <li>(PBL) 정책 분석</li> <li>(Kadaster) 토지등기부</li> <li>(RVO) 보조금</li> <li>(RWVS) 정보포털</li> <li>(CBS) 데이터 및 통계</li> </ul>

## 1. 영국, ECHILD 데이터 세트

### 1.1. 개요 및 추진 배경

영국 정부는 개인의 정보인권을 보호하면서 행정 데이터를 공동활용하여 정책 연구에 활용할 수 있도록 안전한 데이터 이용 환경을 만들기 위해 노력해 왔다. 그 노력의 일환으로 2013년 영국은 행정 데이터 연구 네트워크(ADRN: Administrative Data Research Network) 발족시키고, 대학, 정부부처, 대표적인 통계기관, 그리고 정책 연구소 간의 파트너십을 꾸리고 정부 데이터를 제공함으로써 연구자들이 행정 데이터를 공공의 목적을 위한 정책연구 및 발굴에 활용할 수 있도록 지원하였다. 이후 2018년 ADRN이 영국 행정데이터 연구(ARD UK)로 대체된 후에도 고용, 주거, 보건, 아동, 노인, 불평등 및 사회적 배제(social exclusion), 기후와 환경, 범죄 등 다양한 영역의 공공데이터를 연계하는 데이터 연계 사업과 이미 구축된 통합 행정 데이터를 공동활용하는 연구를 지원하고 있다(한은희, 2020).

그 중 대표적인 사업으로 영국 행정데이터 연구(ARD UK)와 영국의 통계청(ONS), 유니버시티 칼리지 런던(UCL: University College London), 국가의료서비스 디지털(NHS Digital), 그리고 교육부(ED)의 합동프로젝트인 'ECHILD 데이터 세트'가 있다.

### 1.2. 추진 내용

'ECHILD 데이터 세트'는 영국 교육부(ED)가 보유하고 있는 국가 학생 데이터(NPD: National Pupil Database)와 국가의료서비스(NHS)병원의 사고 및 응급 서비스, 외래, 입원 기록 등 병원 에피소드 통계(HES: Hospital Episode Statistics) 데이터, 그리고 통계청(ONS)의 사망률 데이터 등을 통합하여 만들어진 데이터베이스이다.

데이터 연계는 NHS Digital에 의해서 수행된다. 1단계로 환자식별정보를 합법적으로 수집할 권한을 가진 법정 기구인 NHS Digital이 교육부의 학생 데이터(NPD)를 토대로 개인의 식별 정보가 없는 가명 정보(aPMR: Anonymised Pupil Matching Reference)를 생성한다. 2단계로 NHS Digital은 HES를 토대로 환자식별정보의 다양한 조합을 사용하는 매칭 알고리즘을 사용하여 가명 정보인 HES-ID를 생성한다. 그리고 3단계로 NPD와 HES의 각각의 가명정보인 aPMR와 HES-ID에 일치하는 NHS 환자 인구 통계

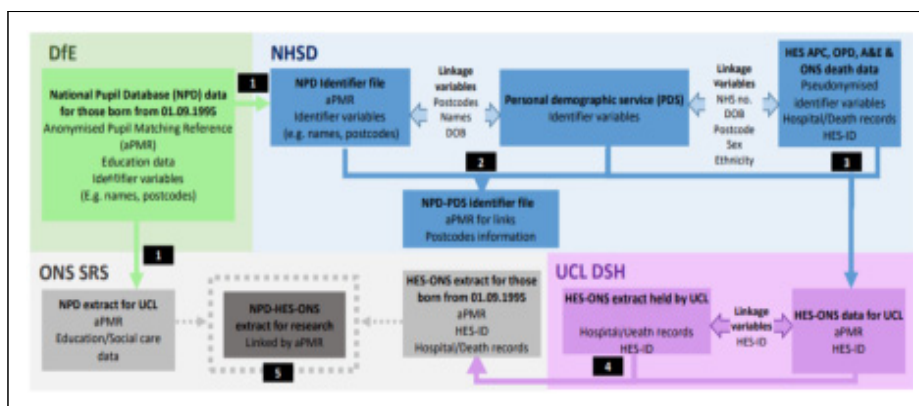
국가 데이터베이스인 PDS(Personal Demographics Service)와도 연계한다. 4단계는 유니버시티 칼리지 런던(UCL: University College London)이 이미 보유하고 있던 ONS의 사망률 데이터와 이미 연계된 NPD-HES 데이터베이스를 HES-ID를 기준으로 연계한다. 이렇게 연계된 데이터베이스 세트는 NHS Digital에서 ONS의 연구서비스(SRS:National Statistics Secure Research Service)에 이관한다. 마지막 5단계에서는 국가통계청(ONS)은 NPD-HES-ONS의 연계자료를 토대로 약 14개의 연결된 데이터를 포함하는 ‘ECHILD 데이터 세트’를 구성한다.

한편, ‘ECHILD 데이터세트’의 안전하고 책임있는 데이터 이용을 위해 ONS(Office for National Statistics)는 재식별이 어렵도록 만드는 '5가지 금고' 모델에 따라 안전한 연구 환경 서비스(SRS)를 실시하고 있다.

이렇게 만들어진 데이터베이스는 1995년 9월 1일에서 2020년 8월 31일 사이에 태어난 영국의 모든 어린이 및 청소년(25세 이하) 약 1,600만 명의 20년 이상의 교육과 사회복지, 그리고 건강정보를 포함하고 있다.

‘ECHILD 데이터세트’는 연구용 목적이기 때문에 이 데이터를 이용하기 위해서는 국가통계청(ONS)의 연구서비스(SRS)에 연구 제안서를 제출하고 승인을 받아야 한다(한은희 2020).

[그림 3-8] ECHILD 데이터베이스 생성을 위한 연계 프로세스



주: DfE = 교육부; NHSD = NHS 디지털; ONS SRS = 국가통계안전연구청; UCL DSH = 유니버시티 칼리지 런던 데이터 세이프 헤이븐; DOB = 생년월일 ; HES-ID = 병원에피소드 통계 ID.

출처: Mc Grath-Lone 외(2021), The Education and Child Health Insights from Linked Data(ECHILD) Database: An Introductory Guide for Researchers

### 1.3. 의의

‘ECHILD 데이터세트’는 건강, 교육 및 사회 복지 데이터를 연결함으로써 사회과학 분야와 보건 분야의 데이터 통합을 통한 융합연구를 가능하게 하고 아동, 청소년의 건강과 교육, 그리고 삶의 질 향상을 위한 정책 지원을 가능하게 하였다는 평가를 받고 있다(한은희, 2020).

NPD와 HES는 영국의 건강, 교육 및 사회 복지 분야의 대표적인 행정 데이터이다. 이렇게 큰 표본 크기와 긴 추적 기간 동안 축적된 자료는 행정데이터가 아니라면 별도로 수집할 수 없는 귀한 자료이다. 그리고 ‘ECHILD 데이터세트’라는 엄청난 데이터베이스는 정부 주도의 합동프로젝트이기에 가능했다고 볼 수 있다.

이와 같이 영국은 행정 데이터를 일괄적으로 관리하고 접근성을 보장하기 위해 대학, 정부부처, 대표적인 통계기관, 그리고 정책연구소 간의 파트너십을 구축해 공인된 연구자에게 가명정보로 안전하게 관리된 중요한 행정데이터를 제공함으로써 사회 및 경제 연구를 적극적으로 지원하였다.

그러나 영국 정부가 연구와 정책 개발을 위해 광범위하게 데이터를 수집해 왔음에도 불구하고, 부처 간 생산적인 활용 사례는 적은 편이다. 주된 이유는 부처와 데이터 사용을 위한 합의에 장시간이 필요하거나, 데이터 접근에 대한 규칙이 정부 부처들 내에서도 일관성 없이 결정되며, 무엇보다 부처 간 데이터이용을 허용하기 위해 제공하는 입법 체계는 번거롭고 비효율적으로 운영되고 있기 때문이다(윤건, 2019).

## 2. 스웨덴, 노동시장 장기통합 데이터(LISA)

### 2.1. 개요 및 추진 배경

스웨덴이 행정 데이터를 이용한 사회과학 연구는 오랜 전통을 지니고 있다. 스웨덴은 1950년부터 등록 자료와 센서스 자료가 구축되기 시작했는데 1960년대 중반부터 전국민 등록 기반 행정자료 통합, 활용 체계가 구축을 시작하였다. 그리고 1990년대 이후로 공공서비스를 디지털화하고 스마트정부의 역량을 집중해 왔다.

스웨덴의 통계청은 각 행정 부처에서 수집한 행정 등록(administrative register) 데이터를 통합하여 통계등록(statistical register) 데이터로 전환하는 행정 데이터 통합을 추진하였다. 행정 데이터 통합에서 중요한 역할을 담당하는 세 가지 핵심적인 등록 데이터가 있는데 비즈니스 등록 데이터(Business Register), 총인구등록 데이터(Total Population Register)와 부동산 등록 데이터(Real Property Register, RPR)이다. 이 세 가지는 핵심데이터는 각각 고용, 직업, 복지, 건강, 교육과 주거 등의 하위 영역에서 등록 데이터와 결합되어 사용된다(Jansson, 2012: 2, 신광영, 2017 재인용).

한편, 그 과정에서 스웨덴 정부는 각기 다른 정부부처와 정부 공공기관들이 정보를 중복해서 수집하는 것을 막기 위해 효율적으로 정보를 수집, 관리 및 활용한 행정데이터 통합을 시도하였다. 그 중 대표적인 사업으로 1990년부터 통합된 개인 수준의 교육, 소득과 직업에 관한 등록행정자료를 활용해 시계열 데이터(Longitudinell databas kring utbildning, incomst och sysselsättning, LOUISE)를 구축하였다.

LOUISE는 2003년 시작된 사업으로 개인의 삶의 상황, 노동 시장과의 관계, 직장 생활 및 질병에 관한 기존 등록 데이터를 더 잘 활용하기 위한 목적이었다. 이후 2004년부터 스웨덴 통계청(SCB), 보험청(RFV)과 혁신청(VINNOVA)이 함께 데이터 구축을 시작하면서 통합 행정 데이터의 명칭을 LOUISE에서 LISA로 변경하였다. 노동시장 장기 통합 데이터(Longitudinell integrationsdatabas för sjukförsäkrings-och arbetsmarknadsstudier, LISA)는 노동시장과 복지에 관한 정보를 담고 있는 행정 데이터이다.

## 2.2. 추진 내용

LISA는 사회보험청의 건강보험 데이터와 노동부의 노동시장 데이터를 통합한 것이다. 1990년부터 매년 12월 31일 기준으로 16세 이상(2010년부터는 15세 이상) 모든 스웨덴 국민의 고용상태, 소득, 직업, 경제활동, 질병, 실업과 퇴직, 연금, 사회부조와 사적 연금, 출생 국가, 부모 출생지, 이민, 거주지, 고용지역, 최종학력에 관한 개인의 정보와 일하는 기업의 주소, 산업, 피고용자 수, 성별, 교육수준별 피고용자 수, 연봉 정보와 같은 기업 정보까지 연계하여 등록하도록 하고 있다(SCB, 2017).

주요 자료는 등록 기반 노동 시장 통계(RAMS), 총인구등록부(RTB), 소득세등록부(IoT), 인구 및 주택 센서스 등으로부터 얻어진 자료를 공동 활용하였다. 이는 기존의

서베이를 통한 횡단 데이터나 종단 패널 데이터에서는 수집하기가 불가능한 양적, 질적 수준의 자료로서 스웨덴 정부가 개인, 가족 및 사업체의 등록행정자료를 개인식별번호를 바탕으로 통합자료를 구축하였기에 가능한 것이었다.

한편, 스웨덴 통계청은 행정데이터와 서베이를 통해 수집된 데이터의 연계, 통합 서비스도 제공한다. 서베이 조사 응답자들로부터 개인정보 사용에 대한 동의를 얻은 후 이들의 개인번호를 통계청으로 보내면, 통계청이 필요한 등록 데이터와 서베이 자료를 통합하여 비식별화한(de-identified) 자료를 자료 신청자에게 보내준다(신광영, 2017).

## 2.3. 의의

LISA 데이터는 노동시장, 교육과 사회복지에 관한 개인 자료이지만, 가족에 관한 정보와 회사와 일하는 곳에 관한 정보를 포함하고 있는 포괄적인 패널 데이터라는 점에서 매우 독특한 데이터이다. 통상적인 서베이 조사에서 소득에 관한 정보는 모집단의 규모가 제한되어 있고, 소득과 같은 질문에 대한 응답을 회피하거나 거부하는 경우 때문에 조사를 통해서 응답에 대한 신뢰도 문제가 있지만, 행정 데이터는 그러한 문제가 거의 없어서 통계분석의 정확도를 높이는데 기여를 할 수 있다(Donnelly 2016; Kim and Tamborini 2014, 신광영, 2017 재인용). 이는 장기적으로 서베이 조사도 더 저렴하고 더 신뢰할 수 있는 데이터를 활용할 수 있는 장점을 가진다.

한편, 스웨덴 정부는 모든 공공분야의 디지털화를 종합적인 관점에서 일괄적으로 감독하고 협력을 조성하기 위해 2019년 9월 전담조직인 DIGG(Agency for Digital Government)를 신설했다. 이를 통해 전체 정부조직이 데이터 정부 의제의 이행을 위해 정책을 추진하고 각 부처가 서로 협력할 수 있도록 데이터 관리 집행을 실행해 왔다.

그럼에도 불구하고, 스웨덴 공공분야의 데이터 지원체계는 다른 OECD국가에 비교하여 상대적으로 부족하다고 평가받는다. 스웨덴 정부의 데이터 기반의 궁극적인 의미는 정부가 데이터 관리 및 사용을 총괄하여 공공 정책과 서비스에 대한 모든 정보를 제공, 설계, 전달하고 모니터링하는 정도를 의미하는데 그친다. 더불어 데이터 활용과 관련된 규제와 표준, 부처 간 상호운용구조, 공유 데이터 인프라 등 기본 인프라의 부족 때문에 통합된 공공 서비스를 제공하거나 부처 간 조정된 의사결정을 지원하는 데이터전략 수립을 어렵게 했다고 평가받고 있다(윤건, 2019).

### 3. 네덜란드, VIVET Programme

#### 3.1. 개요 및 추진 배경

네덜란드는 공공데이터의 온라인 서비스 등 디지털 정부와 관련하여 EU 국가들 사이에서 최상위권에 해당되는 국가로서 정부의 디지털화를 중점으로 추진하고 있으며, 우리 사회에서 점점 증가하는 데이터가 다시 우리 삶에 영향을 미치는 상태를 데이터화(datafication)라고 정의하고 데이터를 최대한 활용하고자 노력하고 있다.

2018년 전체 내각에 의해 합의에 의해 NL DIGIbeter의 디지털 행정의제가 통과된 후, 2019년 중앙 및 지방 정부의 적절하고 책임 있는 데이터 이용방법을 제시한 「NL DIGItaal: Data Agenda Government」가 발표되었다. NL DIGItaal은 5가지 접근방식을 권고하고 있는데, ① 데이터기반 방식으로 문제 해결 ② 입법과 공공 가치에 초점 ③ 정부 데이터 품질의 향상과 효과적 이용 ④ 데이터 기반 접근방식에 관한 지식수집 및 공유 ⑤ 사람, 조직 및 조직문화 변화에 투자이다(윤건, 2019).

한편, 네덜란드는 풍력 등 신재생에너지 분야 강점을 바탕으로 기후변화 대응을 위한 에너지 전환 정책을 적극적 추진하고 있으며, 신재생에너지 소비 비중 또한 꾸준히 증가하는 추세이다. 그러나 EU 신재생에너지 목표에 따라 네덜란드에 설정된 신재생에너지 소비비중 2020년 목표달성 가능성에 대한 회의적인 시각이 증가하면서 좀 더 강화된 정부의 노력을 요구가 지속적으로 제기되었다.

2019년 네덜란드 주요부처는 「온실가스 감축을 위한 주요 계획」을 수립·발표하였다. 네덜란드 경제기후정책부(EZK), 농업자연식품품질부(LNV) 등 주요부처는 1990년 CO2 배출량 대비 2030년까지 51% 감축, 2050년까지 95% 감축을 목표로 온실가스 배출량의 감축을 위한 계획을 수립하였다. 해당 계획은 CO2주요 배출원인 에너지시설, 에너지 시설을 제외한 산업시설, 도로교통 등을 대상으로 세부적인 연도별 감축 계획과 지원 프로젝트 등으로 이루어져 있다.

이러한 프로젝트의 일환으로 2019년 네덜란드 통계청(CBS: Statistics Netherlands), 네덜란드 기업청(RVO.nl) 및 네덜란드 환경평가청(PBL: Netherlands Assessment Agency), 토지등록부(Land Registry) 등의 합동프로젝트인 VIVET Programme이 시작되었다.



### 3.2. 추진 내용

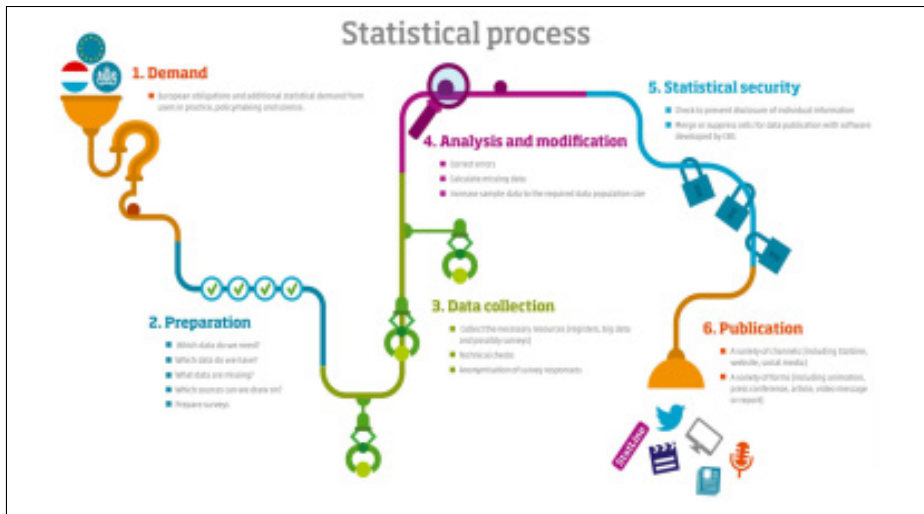
VIVET Programme은 데이터의 융합과 공유를 통해 에너지 전환을 위한 정보규정을 개선하기 위한 프로그램으로 통계청(CBS), 토지등록부, 네덜란드 기업청, 그리고 환경평가청 등의 부처는 서로의 에너지 관련 정보를 이용할 수 있다(Otto Swertz, 2020).

이 정보는 개별 가구의 에너지 소비에서 에너지 네트워크 및 에너지 생산에 대한 데이터에 이르기까지 필요한 에너지 전환을 실현하기 위해 많은 양의 데이터를 제공한다.

주요 자료는 등록 행정데이터와 빅데이터, 서베이자료 등 필수적인 수집자료와 함께 기술적인 자료들이 포함되었는데, 이러한 자료들을 토대로 토지등록부는 토지등록, 기업청(RVO.nl)은 보조금 수여, 환경평가청은 정책분석 등을 처리하였다. 특히, 통계청은 데이터와 통계를 수행하였는데 주된 업무에는 지방자치단체별 태양에너지 자료 발간 등의 업무를 수행하였다.

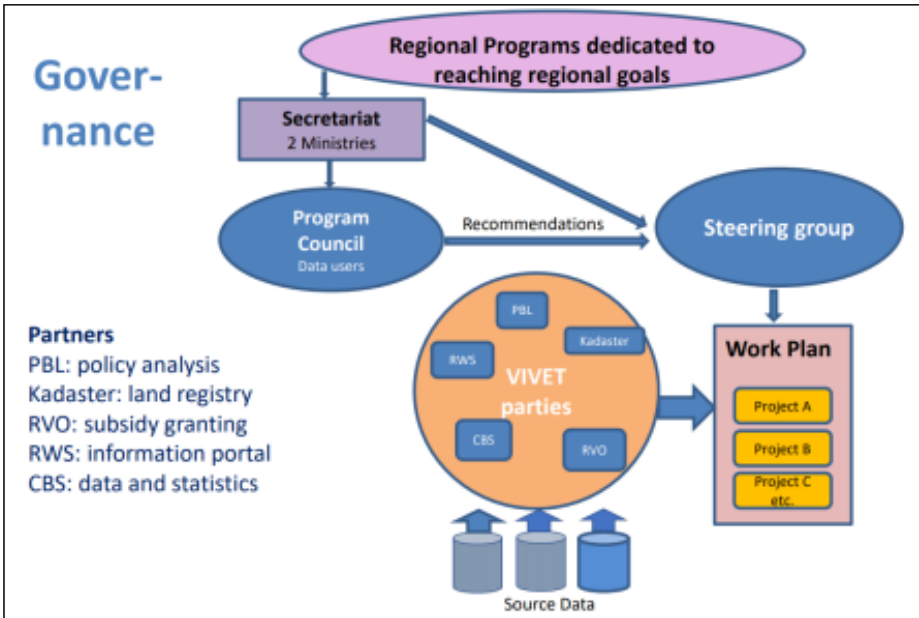
또한 통계청(CBS)은 탄소 회계 금융 파트너십을 위한 파트너십(PCAF)과의 협력 사업으로 6개의 네덜란드 은행과 1개의 보험사의 모든 모기지과 관련된 CO<sub>2</sub> 배출량을 측정했다. 이 7개 은행의 모기지 포트폴리오가 적용되는 모든 가구의 CO<sub>2</sub> 배출량에 대한 측정이 가능해지면서 대출 및 투자 활동과 관련된 GHG 배출량의 측정도 보다 신뢰도가 높아지는 결과를 가져왔다.

[그림 3-9] VIVET Programme의 통계 프로세스



출처: Otto Swertz, Head Energy, Statistics Netherlands(2020)

[그림 3-10] VIVET Programme의 데이터 공동활용 개념도



출처: Otto Swertz, Head Energy, Statistics Netherlands(2020)

### 3.3. 의의

VIVET Programme은 네덜란드의 에너지 전환을 위한 정보규정을 개선할 목적으로 만들어진 프로그램이다. 개별 가구의 에너지 소비 데이터, 에너지 네트워크, 에너지 생산에 대한 데이터 등 각 부처가 개별적으로 생산 및 관리하고 있던 다양한 종류의 데이터가 VIVET Programme을 통해 이를 효과적으로 결합하고 공유되었다.

한편, 네덜란드는 「NL DIGitaal: Data Agenda Government」를 기반으로 정책 개선 및 사회문제 해결을 위해 중앙 및 지방 정부의 책임 있는 데이터 이용을 위해 노력해 왔다. 내무부(BZK)는 행정의제를 추진하는 총괄기관으로서 예산 책정부서 정책협의 회 운영 등 네덜란드 정부 디지털화를 이끌어가고 있다.

이를 중심으로 정부 차원의 정책협의회(OBDO)를 구성하였는데, 중앙 정부 최고 정보 사무소(CIO-Rijk), 네덜란드 지방 협회(IPO), 내무부(BZK), 경제기후정책부(EZK), 인프라 수자원 관리부, 법무부, 교육문화과학부, 사회고용부, 보건복지부, 네덜란드 수도 당국, 지방자치단체 연합(VNG)은 물론 물 관리 사무국, 통계청(CBS) 등이 참여하여 8

주에 한 번 회의를 통하여 주요한 디지털 정책 의제를 논의하여 결정하고 있다.

또한, 내무부가 위탁하고 자체 예산을 확보하여 만들어진 데이터 기반의 공공부문 전문 센터(LED)를 통해 개방형 데이터 전문가, 개인 정보보호전문가, 정보 보안 책임자, 기록 보관소, 빅데이터 전문가 등을 지원하고 있다.

오픈 소스 플랫폼 Pleio를 운영하여 모든 정부 이해 관계자가 데이터 관련 업무를 추진하는데 적극 활용하도록 하고 있으며, 중앙 데이터 포털 data.overheid.nl을 통해 정부에서 축적한 15,000개의 데이터 세트를 제공하고 있다.

반면에, 공공의 가치 보호를 위한 윤리적 데이터 사용을 권장하기 위하여 내무부(BZK)가 네덜란드 지방자치단체연합과 함께 데이터의 책임있는 사용에 대한 일반 원칙인 DGPR을 개발하여 준수 및 이행을 감독하고 있다.

McKinsey Global Institute는 공공 분야의 데이터와 데이터분석의 가치를 평가한 결과, EU 국가 중 대부분의 세금 및 사회복지 기능을 온라인으로 전환한 네덜란드가 부처 간 데이터 공유 개선으로 운영경비를 절감해 가장 효율적으로 데이터 가치를 향상시켰다고 평가했다(윤건, 2019).

한편, 네덜란드는 이러한 노하우를 바탕으로 다양한 조직이 부문 간 데이터 공유의 가치를 실현하기 위해 협력하는 개방적이고 성장하는 국제 이니셔티브 데이터 공유 연합(Data Sharing Coalition)을 이끌어가고 있다.

데이터 공유 연합(Data Sharing Coalition)은 기존 데이터 공유 이니셔티브를 기반으로 해당 영역에서 데이터 공유의 가치를 실현하도록 강화하는 것을 목표로 하며, 연합의 일환으로 기존 데이터 공유 이니셔티브, 산업 협회, 지식 기관, 기업, 과학자 및 전문가가 부문 간에 데이터를 공유하기 위한 일반적인 계약 시스템을 만들기 위해 노력하고 있다. 연합은 네덜란드 경제기후정책부(EZK)의 지원으로 2020년 1월에 시작되었으며 38개 기관 및 조직이 참여하고 있다.

## 제4절 소결

---

제2절 국내사례는 정부부처 및 지방자치단체가 실시한 데이터 공동활용 사례를 살펴 보았다. 그 결과, 국내사례의 경우 국정과제 또는 주요 사회 이슈들을 중심으로 관련된 부처가 공동의 문제해결을 위해서 필요한 데이터를 통합이나 연계 과정 없이 수집하고 상관분석 등 통계분석을 이용하여 주요한 영향을 미치는 요인을 발견하고 이에 따른 정책적인 후속조치를 취하여 문제를 해결하고나 개선하는 방식으로 이루어졌다는 것을 알 수 있었다.

제3절 해외사례는 데이터기반행정의 모범적인 사례로 평가받고 있는 영국, 스웨덴, 네덜란드의 데이터기반행정 및 데이터 공동활용 주요 사례를 중심으로 추진배경, 추진내용, 정책환경 등을 살펴보았다.

그 결과, 영국, 스웨덴, 네덜란드의 경우 데이터 개방성이 높은 나라로 오래전부터 각국의 복수의 정부부처가 협력하여 각자 보유한 행정 데이터를 통합 또는 연계한 데이터 세트를 만드는 노력을 기울였다는 것을 알 수 있었다. 그리고 그 과정에서 데이터 연계 과정에서 발생할 수 있는 개인정보보호 문제를 해결하기 위해 가명정보를 생성하고 복잡한 알고리즘을 활용하여 결합하는 노력을 기울였다. 그리고 네덜란드의 플랫폼 또는 영국 행정데이터 연구(ARD UK), 스웨덴의 국가통계청(ONS), 그리고 네덜란드의 통계청(CBS)과 같이 데이터 관리 주체기관을 설립 또는 지정하여 공공정책 개발 및 정책 개선을 위한 목적으로 연구자에게 적극적으로 개방하는 노력을 하고 있다는 것을 확인하였다. 뿐만 아니라 정책을 개선하고 사회 문제를 해결하는 공공의 목적으로 해당 자료를 활용하는 연구자에게 적절하고 책임 있는 데이터 이용방법을 제시하기도 하였다.

정리하면, 우리나라의 경우 사회문제를 해결하기 위해 필요한 경우에 한하여 꼭 필요한 정보만을 부처간 공유하는 양상이라면, 해외의 경우 적극적인 데이터 개방 및 부처간 협력을 바탕으로 공공정책 개발을 위해 필수적으로 활용되는 공공의 행정데이터를

모두 모아 연계하고 데이터 세트를 구성하며, 정책연구 및 공공의 목적을 위해 적극적으로 개방하고 있는 것이다.

이와 같은 국내사례와 비교했을 때 해외사례가 주는 시사점은 다음과 같다. 우리나라도 부처 간 데이터 공유 및 융합이 원활하게 추진할 수 있도록 데이터 관련 제도가 재정비되어야 한다.

영국의 경우, 개인의 정보인권을 보호하는 가운데 행정 데이터를 연계하고 이를 정책 연구에 적극 활용할 수 있도록 하는 안전한 데이터 이용 기반을 만들기 위해 범부처적인 노력을 기울여 왔다. 특히, 2013년 행정 데이터 연구센터(ADRC)를 설립하고, 행정 데이터 연구 네트워크(ADRN)가 발족되면서 안전한 환경에서 행정 데이터의 연계가 이루어지고, 연구자는 효과적으로 데이터를 활용할 수 있는 물리적 기반이 마련되었다.

스웨덴의 경우, 오랜 기간 동안 국가 주도로 행정데이터(registered data)를 체계적으로 수집, 관리하고 또 통합적으로 활용하였으며, 스웨덴 통계청이 중심이 되어 개인, 가족 및 사업체의 등록행정자료를 개인식별번호를 바탕으로 통합자료를 구축하는 동시에 서베이 데이터의 연계, 통합 서비스도 제공하였다. 그 결과 통합자료를 활용한 다양한 정책연구를 통해 스웨덴의 다양한 복지정책의 구체적인 효과를 확인할 수 있었다.

네덜란드의 경우, 경제기후정책부(EZK)와 네덜란드 지방정부연합이 주축이 되어 공공분야의 행정데이터에 대한 규제와 합의를 마련하였다. 이에 네덜란드 중앙과 지방정부를 포함한 모든 공공분야의 조직은 합의된 행정 의제에 제시된 규제와 합의를 준수해야 한다. 특히, 2018년 NL DIGibeter를 토대로 프로그램 VIVET Programme을 실시함으로써 부처 간 데이터 공유 개선으로 운영경비를 절감해 가장 효율적으로 데이터 가치를 향상시켰다는 평가를 받고 있다.

반면에 우리나라는 지난 해 시행된 「데이터기반행정법」 제5조에 명시된 데이터기반 행정 활성화 위원회, 제20조 데이터분석센터 등에 대해 명시하고는 있지만 구체적인 역할과 권한에 대한 부분이 구체적으로 명시되지 않아 데이터 관련 사업을 추진하려는 정부부처 및 기관의 입장에서는 모호한 부분들이 많다. 특히 데이터 공동활용을 추진할 수 있는 총괄기관의 역할과 권한에 대한 부분이 명확하지 않다.

따라서 데이터기반행정의 관리체계를 정립하고 데이터의 활용역량을 강화하기 위해서는 영국의 행정 데이터 연구 네트워크(ADRN)와 같은 역할을 하는 문화체육콘텐츠 관련 공공행정데이터를 전담하는 기관 또는 조직이 필요하다. 이를 위해서는 데이터에 대

한 전문 인력과 예산 또한 확충되어야 한다. 그렇게 만들어진 기관에서는 문화체육콘텐츠 관련 주요 행정데이터에 대한 현황조사를 주기적으로 실시하고, 정책개발을 위해 자주 활용되고 필수적으로 활용되는 행정데이터를 모두 모아 관리할 필요가 있다. 그리고 영국의 'ECHILD 데이터 세트'와 '노동시장 장기 통합 데이터(LISA)'와 같은 행정데이터 세트를 만들고 데이터 시스템 또는 플랫폼에 등록하여 공공의 목적으로 활용하고자 하는 문체부 산하기관의 담당자 및 정책연구자들에게 적극 개방하여야 한다.

문화·관광·콘텐츠 데이터 공동활용 기반 구축 연구

제4장

데이터 현황조사





# 제1절 데이터 현황조사 수행 체계

## 1. 데이터 현황조사의 역할과 의미

문화체육관광부 소속·산하기관들이 보유 및 운영·관리하고 있는 데이터 소재 및 현황에 대한 정확한 파악 및 문체부 차원의 전사적인 데이터 공유체계가 부족 상황에서 2021년 12월10일에 「데이터기반행정 활성화에 관한 법률」이 시행됨에 따라, 데이터기반행정에 맞는 공동활용데이터 현황 파악 및 분석 수요 등을 파악할 수 있는 데이터 현황 조사의 중요성이 부각되었다.

업무관행, 과거사례, 추정치 등 경험과 직관에 의한 일률적인 행정 추진이 아닌 객관성·신뢰성이 담보된 데이터에 기반한 과학적인 행정 추진을 위해서는 데이터 공동활용과 상호 협력을 목표로 하는 데이터 현황조사 수행이 필요하다.

### 《 추진근거 》

- ▶ 문화체육관광부 데이터분석센터 운영기관으로 '한국문화관광연구원' 설치·운영('20.12.10.)
- 「문화체육관광부 데이터 관리 규정」 제12조 제3항에 따라 데이터기반행정의 수행에 필요한 데이터의 분석 등을 통하여 정책 수립 및 의사결정을 지원하기 위해 「데이터기반행정 활성화에 관한 법률」 제20조 제1항에 의거 설치·운영
- 주요 역할: ①문화체육관광부 데이터 정책 총괄 지원, ②공공기관 간 데이터의 연계·제공·공동활용에 관한 업무 총괄 지원, ③데이터분석센터 업무 지시·조정 등
- 중점 업무: ①문체부 공공기관 보유 데이터 전수조사, ②문화체육관광 공동활용 데이터 발굴·가공·등록, ③기획과제 발굴, 기관 분석과제 수요조사 등

본 연구에서 문체부 소속·산하기관 데이터 현황조사를 수행하는 궁극적인 목적은 문화체육관광 분야 공공데이터의 정확한 소재 및 현황을 파악하고 나아가서는 공동활용 데이터 발굴·관리를 통한 문체부 소속 및 산하기관간에 공동활용데이터 이용 활성화 방안을 마련하고자 하는 것이다.

보다 구체적으로는 문체부 소속 및 산하기관들이 행정 업무를 수행함에 있어 필요로 하는 데이터 현황을 파악할 수 있는 기준을 마련하고 데이터기반행정 관점에서 데이터 공동활용 및 제약조건의 실태를 객관적으로 측정할 수 있는 현실적 도구를 마련하는 것이다.

## 2. 주요 원칙

### □ 데이터 현황조사의 수행 목적 및 추진근거를 고려하여 조사 방향 설정

데이터 현황조사의 수행 목적 및 추진근거를 고려하여 조사 수행 방향을 설정하였다. 데이터 현황조사를 설계하기에 앞서 정책의 방향성, 정책목표를 정확히 전제하고 조사를 통해 파악하고자 하는 데이터 주체 및 포괄범위가 어디까지인가를 구체적으로 정의해야 한다.

### □ 데이터 현황을 객관적·정량적으로 측정할 수 있도록 조사표 문항 설정

데이터 보유 및 수요 현황과 데이터 활용 실태를 객관적·정량적으로 측정할 수 있도록 조사표 문항을 설계하였다. 데이터 현황 및 활용 실태를 정확하게 파악할 수도 있도록 조사항목을 설정하되, 객관적·정량적 설정이 어려운 경우라도 최대한 객관성을 담보할 수 있는 형태로 설정하는 것을 원칙으로 하였다. 조사항목의 명확성 및 구체성, 시계열 유지 가능성, 비교 가능성, 활용성 등을 고려하여 향후 지속적으로 작성 가능한 조사항목을 선정하였다.

### □ 국내외 데이터 관련 주요 지표의 유형 및 특징을 분석하여 세부지표 항목 선정

국내외 데이터 공동활용 관련 주요 지표의 유형 및 특징을 분석하여 세부지표 항목을 선정하였다. 데이터 공동활용 관련 선행 연구 및 사례분석을 등을 토대로 조사대상별, 영역별 지표 선정 방안을 검토하였고, 데이터기반행정부에 근간한 과학적 행정 추진의 방향성 모색 및 데이터 공동활용 기반 구축이라는 연구목적에 맞는 조사대상 및 작성대상 데이터의 범위 선정에 대한 명확한 기준을 정립하였다. 이 외에도 공공데이터 이용 활성화 및 데이터기반행정 추진의 기반이 될 수 있도록 하는 지표항목 선정 기준을 마련하고 세부 지표를 선정하였다.

#### □ 검증을 통한 최종 조사표 설계 완료

전문가 자문 및 시험 조사를 통한 조사 항목의 타당성을 검증한 후 데이터 보유 및 수요 조사, 데이터 활용 실태조사 수행이 가능한 항목들을 최종 조사항목으로 선정하였다.

### 3. 데이터 현황조사의 기본 방향

데이터 현황조사를 수행하는 궁극적인 목적인 문화체육관광 분야 공공데이터의 정확한 소재 및 현황을 파악하고 나아가서는 기관 간에 공동활용 가능한 공동활용데이터 이용 활성화 방안을 마련하기 위해서는 무엇보다 문화체육관광 분야 데이터 현황에 대한 총괄적 접근이 필요하다. 본 연구에서는 공공데이터를 공공기관에서 법령 등에서 정하는 고유의 행정업무 수행을 통해 생성, 취득하여 운영하는 모든 데이터베이스(DB)를 대상으로 정의하고 있으며, 이미 외부에 개방된 공공데이터와 미 개방된 공공데이터 현황 모두를 파악하는 개념으로 접근하고 있다. 특히, 문화체육관광부 소속 및 산하기관이 보유하고 있는 데이터 소재 및 현황을 파악하기 위한 전제로서, 데이터 대한 명확한 개념 정의가 사전에 이루어져야 하며, 조사대상 데이터에 대한 정확한 조작적 정의가 제시되어야 할 것이다.

공공기관이 고유 업무 수행을 통해 보유하고 있는 공공데이터는 크게 이미 외부에 개방된 데이터와 외부에 개방하지 않은 데이터로 구분하여 파악할 수 있다. 데이터기반행정법에서는 다른 공공기관에서 정책 수립 및 의사결정에 활용할 수 있도록 제공하는 데이터를 ‘공동활용 데이터’로 정의하고 하고 있는데, 여기서 데이터가 외부에 개방되어 있는지, 아닌지 여부 또는 행정안전부 메타관리시스템에 등록되어 있는지, 아닌지 여부에 따라서 ‘공동활용 데이터’ 신규 등록 데이터 여부가 판단되게 된다. 기 등록된 데이터는 기관간 협조를 통해 현황조사를 통해 파악하는 것이 효율적으로 판단되며, 신규로 등록 또는 외부에 개방되지 않은 데이터는 보유 데이터 전수조사를 통해 데이터 현황을 파악하는 것이 최선의 방안이 될 것으로 판단된다.

공공데이터 현황조사는 행정안전부 메타데이터관리시스템, 공공데이터포털과 한국문화정보원 문화데이터광장 등 문체부 소속 및 산하기관의 데이터를 통합적으로 관리·운영하고 있는 공공데이터는 정보를 통해 파악하고자 하며, 외부에 개방하지 않은 데이터

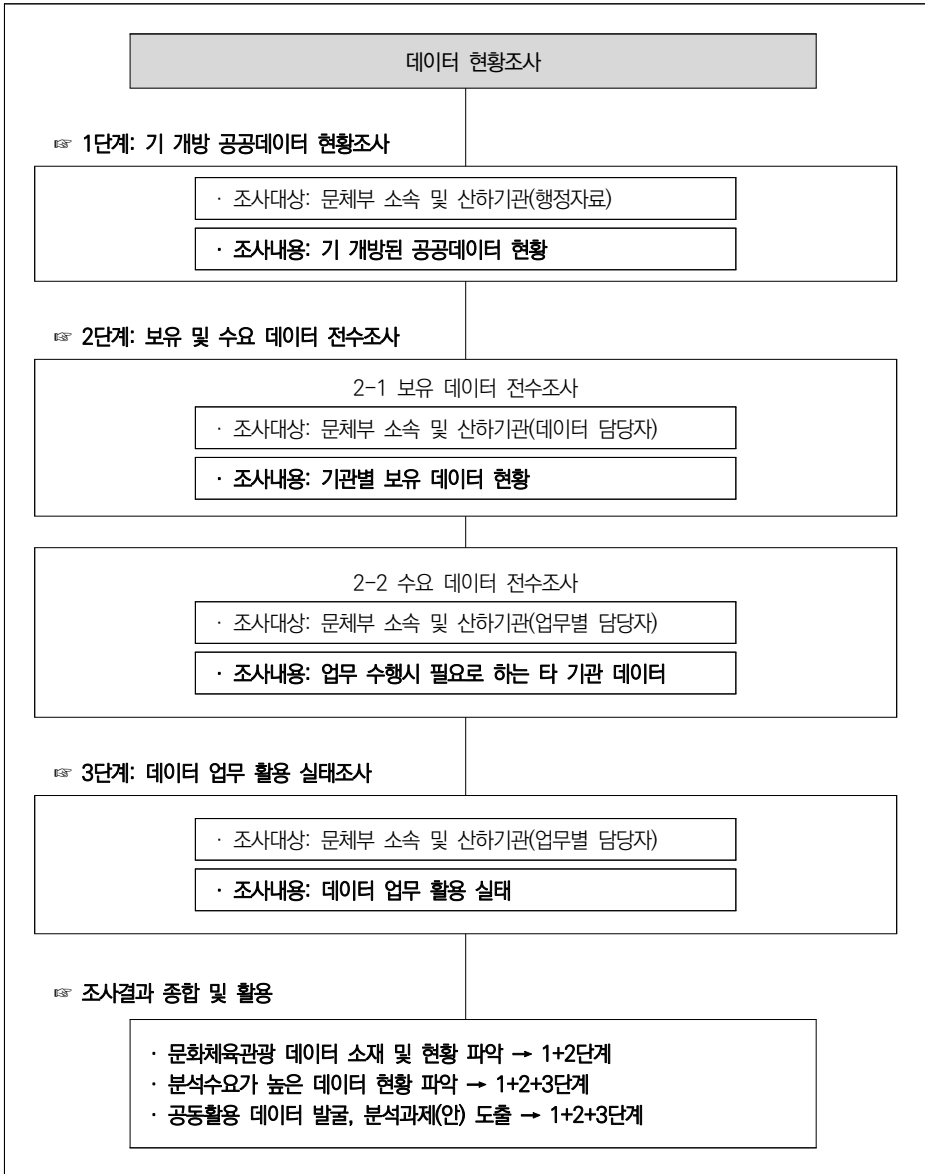
현황은 보유 데이터 전수조사를 통해 파악하고자 한다. 구체적으로는 문화체육관광부 소속 및 산하기관이 생성, 취득하여 보유 또는 운영·관리하고 있는 공공데이터 중 기 개방(공개된) 데이터 정보는 이미 행정안전부에서 수집·정리하여 모니터링하고 있는 관계로 문체부 소속 및 산하기관 보유 데이터 전수조사 대상 데이터에서 제외하고자 한다. 즉, 보유 데이터 전수조사에서는 외부 개방 데이터를 제외한 시스템을 통해 개방(공개)하지 않은 데이터만을 조사대상 범위로 설정할 필요가 있다. 단, 데이터의 누락 없 시 총괄적인 문화체육관광 분야 데이터 소재 및 현황 파악을 위해서는 공공데이터포털 등 데이터 통합플랫폼에 등록되지 않은 자체 기관 홈페이지 및 자체 정보시스템에서만 공개한 데이터와 전체 데이터 중 일부만 공공데이터포털 등에서 등록·개방(공개)된 데이터(ex. 개인정보 포함으로 일부만 개방)는 조사대상에 범위에 포함시켜야 한다.

기 개방된 공공데이터 및 보유 데이터 외에 각 기관 정책수립 및 분석업무 등에 필요한 타 기관 데이터에 대한 수요조사와 실제 데이터를 업무에 활용하고 있는가에 대한 실태를 파악해야 한다. ‘수요 데이터 전수조사’와 ‘데이터 업무 활용 실태조사’ 수행을 통해 분석수요가 높은 데이터 유형을 파악할 필요가 있으며, 데이터기반행정 관점에서 데이터 공동활용 및 제약조건의 실태를 파악하는 것 또한 중요한 조사의 수행 방향이기도 하다.

공공데이터 현황조사, 보유 데이터 전수조사, 수요데이터 전수조사, 데이터 업무활용 실태조사 이상의 4가지 조사의 궁극적인 수행 목적은 문체부 소속 및 산하기관 간에 공동활용데이터 이용 활성화 방안을 모색해보고 나아가서는 데이터기반행정 추진기반 및 역량 강화 방안을 마련하고자 하는 것이다. 이 외에도 조사결과를 기반으로 하여 문화체육관광부 차원의 전사적인 기획과제 및 분석과제 발굴, 기관간 협업 과제 등의 발굴에 필요한 기초자료로 활용 될 수 있을 것이다.

데이터 현황조사는 공공데이터 현황조사, 보유 및 수요 데이터 전수조사, 데이터 업무 활용 실태조사 이상의 3단계 조사로 구성되며, 단계별 수행 체계는 다음 <표 4-1>과 같다.

〈표 4-1〉 데이터 현황조사 수행 체계



## 제2절 공공데이터 현황조사

### 1. 조사개요

공공데이터 현황조사는 행정안전부 메타데이터관리시스템, 공공데이터포털과 한국문화정보원 문화데이터광장 등 문체부 소속 및 산하기관의 데이터를 통합적으로 관리·운영하고 있는 공공데이터 정보 파악을 목적으로 진행하였다.

조사대상은 공공데이터 관련 법 등에 따라 생성·수집하여 운영하는 모든 정보시스템의 데이터베이스(DB) 현황 정보(통계, 파일자료 등 포함)이며, 행정지원DB(메일, 메신저 등), 시스템 운영DB(백업, 성능, 모니터링 등) 등 업무지원용 DB는 조사대상 범위에서 제외하였다.

〈표 4-2〉 조사 제외 대상 예시

- |  |
|--|
| <p>▶ (행정지원업무 DB)</p> <ul style="list-style-type: none"><li>• 행정지원업무(인사·총무·재무 등) 수행을 위한 그룹웨어, 패키지DB(ERP 등), 메일, 팩스, 웹하드, 메신저, 화상회의 등</li></ul> <p>▶ (시스템운영 DB)</p> <ul style="list-style-type: none"><li>• 백업/아카이브, 네트워크관리(NMS), 보안관리(보안USB 등), 솔루션DB(품질관리 SW 등) 등</li></ul> <p>▶ (기타 DB)</p> <ul style="list-style-type: none"><li>• 업무지원을 위한 공통기반 시스템 DB(OTP, GPKI 등), 계측성 DB 등</li></ul> |
|--|

조사내용은 정보시스템 구축·운영 시 관리하고 있는 표준화 산출물을 통해 DB별 세부 항목 수집현황이며, 구체적으로는 문체부 소속 및 산하기관이 제출·등록한 문화체육관광 분야 데이터 정보 중 데이터유형, 데이터분류, 데이터보유기관, 데이터명칭, 컬럼(변수)명, 컬럼별 개방여부 및 개인정보 포함 여부 등을 필수 조사항목으로 설정하고 조사를 실시하였다. 조사결과를 활용하여 기 개방된 문화체육관광 분야 데이터의 정확한 소재 및 현황을 파악할 수 있을 것으로 판단된다. 이 외에도 외부에 개방되지 않은 데이

터를 대상으로 조사가 진행되는 ‘보유 및 수요 데이터 전수조사’와 ‘데이터 업무 활용 실태조사’ 결과와 총괄적인 정리·분석을 통해 분석수요가 높은 데이터 현황 파악하고 데이터 기반의 과학적 정책 분석 실행을 위한 공동활용 데이터 발굴의 기초자료로 활용하고자 한다.

공공데이터포털, 문화데이터광장 등의 통합정보시스템은 문체부 소속 및 산하기관이 보유한 데이터 정보를 수집·저장하여 구축·운영되고 있기 때문에 조사협조가 어려운 개인정보 등을 포함하고 있는 경우가 대부분이다. 따라서 본 연구에서는 문화체육관광부 정책분석팀에 협조를 받아 개인정보가 마스킹 처리된 통합정보시스템 내 필수 메타데이터 정보만을 입수하여 공공데이터 현황조사 진행 및 결과분석을 진행하였다.

〈표 4-3〉 조사 수행 절차

구분	주요내용	비고 (수행 주체)
조사계획고지	<ul style="list-style-type: none"> <li>보유 메타데이터 조사계획 및 작성양식 배포</li> </ul>	한국문화관광연구원 문화체육관광부
↓		
통합정보시스템 구축·운영기관별 자료준비 및 제출	<ul style="list-style-type: none"> <li>기본조사서 작성</li> <li>표준화 산출물 현행화</li> <li>컬럼별 개방 및 개인정보 포함하여 작성</li> <li>메타데이터 제출</li> </ul>	기관 담당자
↓		
현장조사 실시	<ul style="list-style-type: none"> <li>시스템 접속환경 및 담당자 인터뷰</li> <li>미등록 및 누락정보 현장점검</li> </ul>	한국문화관광연구원 기관 담당자
↓		
제출자료 점검·보완	<ul style="list-style-type: none"> <li>제출자료 취합·정리</li> <li>미등록 및 누락정보 확인</li> <li>미등록 및 누락정보 현행화 등 자료 보완</li> </ul>	한국문화관광연구원 기관 담당자
↓		
조사결과 분석	<ul style="list-style-type: none"> <li>보유 데이터 확정 및 결과 분석</li> </ul>	한국문화관광연구원

## 2. 조사결과

〈표 4-4〉 문체부 소속·산하기관 데이터 시스템 구축·운영현황

기관명	시스템명	주요 목적
문체부 및 소속기관	<ul style="list-style-type: none"> <li>• 통합문화이용권정보시스템</li> <li>• 빅데이터 기반 언어 맞춤형시스템</li> <li>• 도서관 빅데이터플랫폼</li> <li>• 사서의사결정시스템</li> <li>• 예술인경력증명시스템</li> <li>• 국제문화교류정보시스템</li> <li>• 지역문화 통합정보시스템</li> <li>• 디지털국악아카이브시스템</li> <li>• 현대사디지털아카이브시스템</li> <li>• 국립중앙극장 공연예술자료관리시스템</li> <li>• 불법오라인도박관리시스템</li> <li>• 국민체력측정시스템</li> <li>• 정책여론수렴시스템</li> <li>• 콘텐츠수출마케팅플랫폼</li> <li>• 공공저작물활용시스템 등 19개 시스템</li> </ul>	문화·예술, 콘텐츠, 관광, 스포츠, 저작권, 소통(홍보) 등을 위한 정형·비정형 빅데이터 등 수집 및 분야별 행정기반 데이터 총괄 관리 등
한국언론진흥재단	• 뉴스빅데이터 분석시스템(BIGKINDS)	뉴스DB(54개 매체, 7천만 건) 구축·제공
한국관광공사	<ul style="list-style-type: none"> <li>• 한국관광공사 관광 빅데이터 플랫폼</li> <li>• 한국관광공사 TourAPI 시스템</li> </ul>	관광 관련 데이터 수집·정제·분석·개방 및 인사이트 발굴 지원 서비스 관광정보 데이터 유통, 개방
한국문화정보원	<ul style="list-style-type: none"> <li>• 문화 공공데이터 통합활용시스템</li> <li>• 문화 빅데이터 플랫폼</li> </ul>	문화 데이터 수집 및 분석, 개방 문화 빅데이터 수집 및 개방, 유통
한국문화관광연구원	• 문화센터, 관광자식정보시스템 등 3개	문화체육관광 분야 데이터 수집 및 관리
한국문화예술위원회	• 문화예술 데이터관리시스템	예술위원회 및 업무협약체결 6개 기관 데이터 수집, 분석, 시각화
예술경영지원센터	• 공연예술통합전산망	공연예술 정보 및 통계서비스 제공
게임물관리위원회	• 자체등급분류 연계 네트워크 시스템	데이터 수집, 게임물 사후관리
한국저작권보호원	• 불법복제물 추적관리시스템 등 5개	저작권침해 데이터 수집, 분석, 시각화
대한체육회	• 체육정보망, 생활체육정보시스템	기관내 데이터 수집 및 분석

자료: 연구자 조사('21.5.31 기준)

〈표 4-5〉 문체부 소속·산하기관 공공데이터 포털 개방 현황

(단위: 개)

구분	파일데이터	오픈API	표준데이터셋	합계
본부 및 소속기관(17개)	112	49	-	161
산하기관(32개)	721	166	-	887
합계	833	215	-	1,048

자료: 행정안전부 공공데이터포털(<https://www.data.go.kr>) 메타데이터 분석



〈표 4-6〉 문체부 소속기관 공공데이터 포털 개방 현황

(단위: 개)

구분	파일데이터	오픈API	표준데이터셋	합계
문화체육관광부	40	9	-	49
국립국악원	1	2	-	3
국립국어원	2	8	-	10
국립민속박물관	4	5	-	9
국립아시아문화전당	8	0	-	8
국립중앙극장	7	0	-	7
국립중앙도서관	4	7	-	11
국립중앙박물관	11	3	-	14
국립한글박물관	2	1	-	3
국립현대미술관	5	1	-	6
대한민국역사박물관	16	2	-	18
대한민국예술원	2	1	-	3
한국예술종합학교	1	1	-	2
한국정책방송원	8	3	-	11
해외문화홍보원	1	6	-	7
합계	112	49	-	161

자료: 행정안전부 공공데이터포털(<https://www.data.go.kr>) 메타데이터 분석

문화체육관광부 소속 및 산하기관에서 보유하고 있는 시스템 구축운영현황을 살펴본 결과, 문화체육관광부 및 소속기관은 통합문화이용권정보시스템, 예술인경력증명시스템, 지역문화 통합정보시스템, 국민체력측정시스템 등 19개의 시스템을 보유하고 있었다. 그리고 문화체육관광 분야 정형·비정형 데이터를 수집 및 분야별 행정기반 데이터 총괄 관리하는데 그 목적이 있었다.

그 밖에 한국저작권보호원은 불법복제물 추적관리시스템 등 5개, 한국문화관광연구원은 문화센터, 관광지식정보시스템 등 3개의 시스템을 구축운영하고 있었다. 한국관광공사와 한국문화정보원은 각각 2개의 시스템을 자료 수집 및 관리의 목적으로 구축운영하고 있었다. 또한, 한국문화예술위원회, 예술경영지원센터, 게임물관리위원회, 대한체육회 등은 각각 1개의 시스템을 구축운영하고 있었다(표 4-4).

문화체육관광부 소속 및 산하기관이 공공데이터 포털에 개방한 데이터를 살펴보면, 본부 및 소속기관 15곳에서 총 161개의 공공데이터를 개방하고 있었는데 그 중 파일데

이터는 112개, 오픈API는 49개를 차지하였으며<표 4-5>, 그 중에서도 문화체육관광부가 파일데이터 40개에 오픈API 9개를 합쳐 총 49개로 가장 많은 공공데이터를 개방하는 것으로 나타났다. 그 다음으로 대한민국역사박물관이 파일데이터 16개에 오픈API 2개를 합쳐 총 18개로 두 번째로 많은 공공데이터를 개방하고 있었으며, 국립중앙박물관은 파일데이터 11개에 오픈API 3개를 합쳐 총 14개로 세 번째로 많은 공공데이터를 개방하고 있었다<표 4-6>.

그 밖에 33개 산하기관의 경우 전체 파일데이터 721개에 오픈API 166개를 합쳐 총 887개를 차지하는 것으로 나타났다<표 4-7>, 그리고 한국문화예술위원회가 파일데이터 145개에 오픈API 14개를 합쳐 총 159개로 가장 많은 공공데이터를 개방하고 있었다. 그 다음으로 한국언론진흥재단이 파일데이터만 117개를 개방하여 두 번째로 많은 공공데이터를 개방하고 있었으며, 한국저작권위원회가 파일데이터 109개에 오픈API 5개를 합쳐 총 114개로 세 번째로 많은 공공데이터를 개방하고 있었다<표 4-7>.

분류별 문화체육관광 분야 공공데이터 포털 개방 현황을 살펴보면, 총 26,979개의 공공데이터 중 문화재 관련 공공데이터가 10,584개로 가장 많았고, 다음으로 관광 관련 공공데이터가 5,880개, 문화예술 관련 공공데이터가 5,099개 등의 순으로 나타났다<표 4-8>. 문체부 소속 및 산하기관이 공공데이터포털에 개방한 데이터와 행정안전부 공공데이터포털에서 개방하고 있는 문화체육관광 분야 전체 데이터를 비교해 보면, 문체부 소속 및 산하기관에서 제공하고 있는 데이터의 비중은 문화체육관광 분야 전체 데이터의 3.2% 수준으로 매우 낮은 비중을 보이는 것으로 나타났다.

<표 4-7> 문체부 산하기관 공공데이터 포털 개방 현황

(단위: 개)

구분	파일데이터	오픈API	표준데이터셋	합계
(재)한국저작권보호원	18	1	-	19
게임물관리위원회	10	1	-	11
국립박물관문화재단	6	0	-	6
국제방송교류재단	7	0	-	7
그랜드코리아레저	53	6	-	59
대한장애인체육회	8	0	-	8
대한체육회	16	0	-	16
국민체육진흥공단	26	57	-	83
세종학당재단	25	1	-	26

구분	파일데이터	오픈API	표준데이터셋	합계
아시아문화원	4	0	-	4
영상물등급위원회	10	4	-	14
영화진흥위원회	0	5	-	5
예술경영지원센터	3	0	-	3
예술의전당	4	0	-	4
재단법인 국악방송	4	1	-	5
태권도진흥재단	2	1	-	3
한국공예디자인문화진흥원	26	0	-	26
한국관광공사	16	17	-	33
한국도박문제관리센터	2	0	-	2
한국문학번역원	0	6	-	6
한국문화관광연구원	15	5	-	20
한국문화예술교육진흥원	4	1	-	5
한국문화예술위원회	145	14	-	159
한국문화예술회관연합회	2	0	-	2
한국문화정보원	44	19	-	63
한국문화진흥주식회사	6	0	-	6
한국언론진흥재단	117	0	-	117
한국영상자료원	4	1	-	5
한국예술인복지재단	4	0	-	4
한국저작권위원회	109	5	-	114
한국체육산업개발주식회사	0	20	-	20
한국출판문화산업진흥원	4	1	-	5
한국콘텐츠진흥원	27		-	27
합계	721	166	-	887

자료: 행정안전부 공공데이터포털(<https://www.data.go.kr>) 메타데이터 분석

〈표 4-8〉 문화체육관광 분야 공공데이터 포털 개방 현황

(단위: 개)

분류명	합계
문화예술	5,099
문화재	10,584
체육	1,750
관광	5,880
문화체육관광 일반	3,666
합계(문화체육관광)	23,313

자료: 행정안전부 공공데이터포털(<https://www.data.go.kr>), 연구자 조사('21.5.31 기준)

다음으로 문체부 산하기관인 한국문화정보원의 문화정보통합시스템 내 데이터 수집·관리 현황을 살펴본 결과, 정형 1,083개, 비정형 74개, 오픈API 487개 등 총 1,651개의 데이터를 수집·관리하고 있는 것으로 나타났다(표 4-9).

그 중 문화예술 관련 공공데이터는 정형 214개, 비정형 6개, 오픈API 140개 등 총 361개로 첫 번째로 많은 것으로 나타났으며, 문화유산 관련 공공데이터는 정형 236개, 비정형 24개, 오픈API 81개 등 총 341개로 두 번째, 문화산업 관련 공공데이터는 정형 210개, 비정형 16개, 오픈API 74개 등 총 300개로 세 번째로 많은 것으로 나타났다.

한편, 한국문화정보원의 문화정보통합시스템 내 데이터 수집·관리되고 있는 공공데이터는 총 1,651개이나 공공데이터포털에 개방된 문체부 소속 및 산하기관 공공데이터는 파일데이터 833개, 오픈API 215개를 합쳐 총 1,048개로 나타나 공공기관에서 보유 데이터 중 일부만을 공공데이터포털을 통해 개방하고 있음을 알 수 있다.

한국문화정보원의 문화정보통합시스템 내 데이터를 세부적으로 살펴본 결과, 총 1,651개의 공공데이터 중 데이터 주제별 구분은 공지 607개, 행사 157개를 제외하고 유물(소장품) 83개, 기타창작물 57개, 교육 49개, 장소 43개, 도서 40개 등의 순으로 나타났다. 그리고 데이터 형태별 구분은 공지/통계 619개, 분류안됨 496개, 행사 157개를 제외하고 창작물 194개, 유형 유산 86개, 장소 43개, 개인/단체 42개, 무형 자산 14개의 순으로 나타났다. 문화정보통합시스템에서 수집·관리되고 있는 주제별 데이터 정보 중 69% 이상이 공지, 행사, 교육 등의 일반 행정데이터로 실제 정책분야에서 활용 가능한 데이터는 매우 낮은 수준에서 개방이 이뤄지고 있는 것으로 나타났다(표 4-10).

〈표 4-9〉 문화정보통합시스템 데이터 수집·관리 현황(1)

(단위: 개)

구분	정형	비정형	오픈API	분류안됨	합계
문화예술	214	6	140	1	361
문화유산	236	24	81	-	341
문화산업	210	16	74	-	300
관광	44	16	44	1	105
체육	57	6	43	2	108
도서	64	3	42	-	109
정책지원	92	1	26	-	119
문화홍보	154	2	2	3	161
맞춤형 API	3	-	35	-	38

구분	정형	비정형	오픈API	분류안됨	합계
기타	9	-	-	-	9
합계	1,083	74	487	7	1,651

자료: 한국문화정보원 문화정보통합시스템 메타데이터 재분류, 연구자 조사('21.5.31 기준)

〈표 4-10〉 문화정보통합시스템 데이터 수집·관리 현황(2)

(단위: 개)

구분	창작물	유형 유산	무형 유산	행사	공지/ 통계	개인/ 단체	장소	분류 안됨	합계
도서	40	-	-	-	-	-	-	-	40
영화/영상	25	-	-	-	-	-	-	-	25
음악	11	-	-	-	-	-	-	-	11
음악녹음	1	-	-	-	-	-	-	-	1
방송	6	-	-	-	-	-	-	-	6
게임	1	-	-	-	-	-	-	-	1
무용	1	-	-	-	-	-	-	-	1
연극뮤지컬	-	-	-	-	-	-	-	-	-
미술	3	-	-	-	-	-	-	-	3
교육	49	-	-	-	-	-	-	-	49
기타창작물	57	-	-	-	-	-	-	-	57
유물(소장품)	-	83	-	-	-	-	-	-	83
유적건조물	-	3	-	-	-	-	-	-	3
음식	-	-	1	-	-	-	-	-	1
무역문화	-	-	10	-	-	-	-	-	10
축제	-	-	3	-	-	-	-	-	3
행사	-	-	-	157	-	-	-	-	157
공지	-	-	-	-	607	-	-	-	607
통계	-	-	-	-	12	-	-	-	12
개인	-	-	-	-	-	31	-	-	31
단체	-	-	-	-	-	11	-	-	11
장소	-	-	-	-	-	-	43	-	43
분류안됨	-	-	-	-	-	-	-	496	9
합계	194	86	14	157	619	42	43	496	1,651

자료: 한국문화정보원 문화정보통합시스템 메타데이터 재분류, 연구자 조사('21.5.31 기준)

## 제3절 보유 및 수요 데이터 전수조사

---

### 1. 조사개요

사회 패러다임의 변화로 데이터가 가치 창출의 핵심 원천으로 부상함에 따라 수많은 데이터의 수집·분석을 통해 새로운 가치를 창출하고 활용하고자 하는 요구 및 수요가 높아지고 있다.

데이터를 정책 수립 및 의사결정에 활용하여 공공기관의 책임성, 대응성 및 신뢰성을 높이고 국민의 삶의 질을 향상시키기 위해 「데이터기반행정 활성화에 관한 법률(이하 ‘데이터기반행정법’)」이 제정(20.6.9.) 및 시행(20.12.10.) 됨에 따라 공공기관은 경험과 직관이 아닌 객관적 증거에 기반한 과학적 행정을 구현하기 위해 데이터 제공, 연계 및 공동활용과 상호 협력이 필요한 상황이다.

이에 본 연구에서는 조사대상 데이터를 크게 두 가지 유형으로 구분하여 기관별 보유 및 수요 데이터 전수조사를 진행하였다. 먼저 문체부 소속·산하기관이 기관 고유의 행정 업무 수행을 통해 생성, 취득하여 보유 또는 운영·관리하고 있되, 외부에 개방하지 않은 데이터 현황 파악을 목적으로 보유 데이터 전수조사를 진행하였다. 동시에 각 기관의 정책수립 및 분석 업무 수행 시 필요한 타 기관 데이터에 대한 수요 파악을 목적으로 수요 데이터 전수조사를 진행하였다.

보유 및 수요 데이터 전수조사의 조사목적은 1차적으로 기관별 보유 및 타 기관 데이터의 수요를 파악하여 문체부 차원에서 소속·산하기관이 보유한 공공데이터의 정확한 소재 및 현황을 파악하기 위함이며, 동시에 조사 결과를 기반으로 공동활용 가능한 데이터 정보를 도출하는 것이다. 2차적으로는 도출된 데이터 정보를 기준으로 정책분야별 데이터 연계·활용 가능성을 진단을 통해 공동활용 데이터 발굴·관리 하고 나아가서는 문체부 소속 및 산하기관간에 공동활용데이터 이용 활성화 방안을 마련하고자 하는 것이다.

## 가. 조사 설계

보유 및 수요 데이터 전수조사는 각 기관에서 생성, 취득하여 보유 또는 운영·관리하고 있으나, 외부에 개방(공개)하지 않은 데이터와 기관별 업무에 필요한 타 기관 데이터를 대상으로 구조화된 데이터 전수조사 양식을 참조하여 엑셀 양식의 조사표를 작성 후 공문으로 제출한다.

〈표 4-11〉 조사 설계 개요

구 분	내 용
조 사 명	• 보유 및 수요 데이터 전수조사
조사근거	• 「문화체육관광부 데이터 관리 규정」제12조(2020.9.25., 문화체육관광부훈령 제430호), 「데이터기반행정 활성화에 관한 법률」제20조
조사목적	• 문체부 소속·산하기관이 보유한 공공데이터 소재 및 현황 파악 • 공동활용 데이터 발굴·관리 및 공동활용데이터 이용 활성화 방안 마련
조사 대상	• 문체부 소속 및 산하기관
조사 내용	• ①기관별 업무 수행을 통해 생성, 취득하여 보유 또는 운영·관리하고 있는 데이터 현황 ②각 기관 정책수립 및 분석업무 등에 필요한 타 기관 데이터에 대한 수요
조사 방법	• 구조화된 보유 데이터 전수조사 양식을 참조하여 엑셀 양식의 조사표를 작성 후 공문으로 제출
조사 일정	• 2021년 7월 ~ 8월

〈표 4-12〉 조사대상 데이터

- ▶(보유 데이터 조사) 기관별로 보유 또는 운영·관리하고 있으나, 외부에 개방(공개)하지 않은 데이터\*\*
- 정책수립, 집행과정 모니터링, 정책평가 등 과정에서 필요할 것으로 예상되는 데이터
  - 기관 간 공동활용 데이터로써 활용성이 높다고 판단되는 보유 데이터
- \* (외부 개방 데이터) 메타데이터관리시스템, 공공데이터포털 등에서 기 개방(공개)된 데이터
- \*\* (외부에 개방하지 않은 데이터) 외부 개방 데이터를 제외한 시스템을 통해 개방(공개)하지 않은 데이터를 연구대상으로 함. 단, ①기관 홈페이지 및 자체 정보시스템에서만 공개한 데이터 ②전체 데이터 중 일부만 외부에 개방(공개)된 데이터(ex. 개인정보 포함으로 일부만 개방)는 연구대상에 포함
- ▶(수요 데이터 조사) 기관별 분석업무에 필요한 타 기관 데이터
- 기존 업무·분석 수행시 이미 활용하고 있는 타 기관 데이터
  - 기존 업무·분석 수행시 필요한 데이터가 있었으나 데이터 보유기관으로부터 제공 되지 않았던 데이터
  - 정책수립, 집행과정 모니터링, 정책평가 등 과정에서 필요할 것으로 예상되는 타 공공기관 데이터

## 나. 조사 내용

설문조사의 내용은 다음과 같다. 보유 데이터 전수조사에서는 데이터기반행정 관점에서 기관 고유의 행정업무를 수행을 통해 생성, 취득하여 보유 또는 운영·관리하되, 외부에 개방하지 않은 데이터의 현황을 파악하기 위한 설문조사의 목적을 고려하여 조사표 양식을 크게 기관정보, 데이터정보, 제공정보의 3가지 부분으로 구성하였다. 3가지 부분별 세부 조사항목을 살펴보면, “기관정보”의 경우 보유 데이터의 실제 존재여부를 확인하고 공동활용 데이터로 활용 협의 등을 추진하기 위해 기관명, 작성자를 필히 작성하도록 조사표를 구성하였다. “데이터정보”의 경우 데이터명칭, 보유 데이터의 공동활용 정책분야, 보유 데이터의 성격을 파악할 수 있는 데이터 설명, 생성·취득 방법, 보유 데이터 개인정보 포함여부 등을 중심으로 구성하였다. “제공정보”의 경우는 데이터 최초 작성시점, 작성주기, 데이터 구축·운영방법, 보유 데이터의 컬럼(변수)명을 입력할 수 있도록 조사표를 설계하였다.

수요 데이터 전수조사에서는 조사 부문의 구분 없이 데이터 활용여부, 데이터 명칭, 데이터 보유기관, 정보시스템명, 공동활용 주제분류, 활용목적, 컬럼(변수)명 이상의 7가지 조사항목으로 조사표를 설계하였다.

### 〈표 4-13〉 주요 용어 정의

#### ▶ 데이터기반행정

- 공공기관이 생성하거나 다른 공공기관 및 법인·단체 등으로부터 취득하여 관리하고 있는 데이터를 수집·저장·가공·분석·표현하는 등의 방법으로 정책 수립 및 의사결정에 활용함으로써 객관적이고 과학적으로 수행하는 행정

#### ▶ 공공기관

- 대통령 소속 기관과 국무총리 소속 기관을 포함한 중앙행정기관과 그 소속기관 및 지방자치단체를 비롯하여 「공공기관의 운영에 관한 법률」 제4조에 따른 공공기관, 「지방공기업법」에 따른 지방공사 및 지방공단

#### ▶ 데이터

- 정보처리 능력을 갖춘 장치를 통하여 생성 또는 처리된 상태의 정보로서 기계에 의한 판독이 가능한 형태로 존재하는 정형 또는 비정형의 모든 것을 ‘데이터’라고 정의

#### ▶ 공동활용 데이터

- 다른 공공기관에서 정책 수립 및 의사결정에 활용할 수 있도록 제공하는 데이터
- 플랫폼에 등록되어 공동활용하는 데이터를 비롯하여 개별 공공기관의 요청에 따라 제공되거나 수집·활용할 수 있도록 하는 데이터도 공동활용 데이터에 포함함

자료: 「데이터기반행정 활성화에 관한 법률」



## 다. 조사 방법

보유 및 수요 데이터 전수조사에 앞서 일부 기관을 대상으로 조사표 보완 및 의견수렴을 목적으로 사전 시범조사를 실시하였다. 문체부 소속 및 산하기관 중 업무 분야 및 보유 데이터 특성을 고려하여 예술인복지재단, 한국장애인문화예술진흥원, 한국도박문제관리센터, 국립현대미술관, 한국문화정보원을 사전 시범조사 대상 기관으로 선정하고 시범조사를 진행하였다. 시범조사를 통해 도출된 조사표 보완사항 및 보완의견을 수렴하여 최종 본 조사 조사표를 확정하였다.

본 조사는 문체부 소속 및 산하기관 전체를 대상으로 구조화된 보유 데이터 전수조사 양식을 참조하여 엑셀 양식의 조사표를 작성 후 공문으로 제출하는 방법으로 조사를 진행하였다. 조사표 중 보유 데이터와 관련하여 기관정보, 데이터정보, 제공정보 등의 대한 사항에 대한 사항을 수집하였으며, 수요 데이터와 관련해서는 데이터 활용여부, 데이터 보유기관, 활용목적, 컬럼(변수)명 등의 대한 사항을 수집하는 것으로 목적으로 조사를 실시하였다.

마지막으로 본 조사를 통해 작성 및 제출된 응답정보에 대한 정확성 점검을 위해 주요 기관을 대상으로 점검조사를 진행하였다.

〈표 4-14〉 조사절차

구분	주요내용	수행주체
조사계획 고지	· 조사계획 및 작성양식 배포	문체부 및 한국문화관광연구원
기관별 자료준비 및 제출	· 보유 및 수요 데이터 조사표 작성·제출	기관 담당자
현장조사 실시	· 제출자료 대비 데이터 현황 확인 · 미작성 및 누락정보 현장점검	한국문화관광연구원
제출자료 점검	· 기관별 제출자료 취합·정리 · 미제출 및 누락정보 확인	한국문화관광연구원
제출자료 보완	· 미제출 및 누락정보 현행화 등 자료 보완	기관 담당자
조사 완료	· 기관별 보유 및 수요 데이터 확정	문체부 및 한국문화관광연구원

〈표 4-15〉 보유 데이터 전수조사 양식

기관 정보	기관명	☞ 응답 기관 명칭				
	작성자 정보	소속(부서명)	성명	연락처		
데이 터 정 보	①데이터명칭(국문)	☞ 보유 데이터의 명칭 EX) 공연예술 올해의 신작 사업성과				
	②공동활용 정책분야	☞ 아래 5가지 분야 중에 1가지 선택 1. 국민의 의견수렴, 2. 특화된 대책 및 맞춤형 서비스, 3. 위험요소 예측 제거 4. 미래수요에 대한 선제적 대응, 5. 행정업무 경제성·효율성 개선 중 선택 엑셀(시트2. 공동활용 정책분야)참고				
	③정부기능분류체계(BRM)	대분류(정책분야)	중분류(정책영역)	소분류(대기능)		
		☞ 엑셀(시트3. 정부기능분류체계)참고				
	④데이터 설명	☞ 보유 데이터의 성격 및 원천, 가공내용, 유의사항 등 설명 EX) 문예진흥기금 공모사업 중 공연예술 분야 "올해의 신작" 지원 사업의 성과(예: 예술작품생산건수, 공연일수, 공연회차 등)				
	⑤키워드	☞ 키워드 3개 입력 EX) 문화예술, 문예진흥기금 공모사업, 올해의 신작지원				
	⑥생성·취득 방법	☞ 직접조사, 2차 가공, 행정자료(기관 내부 자료) EX) 기관 내부 문서				
제공 정보	⑦보유 데이터 개인정보 포함여부	☞ 보유 데이터에 개인정보가 있는지 여부 선택 (Y/N)				
		개인정보 컬럼(변수)명	EX) 이메일, 전화번호, 주소 등			
		암호화 여부	EX) 개인정보 암호화 여부 (Y/N)			
	⑧데이터 최초 작성시점	☞ 데이터의 최초 작성 년 월 입력				
	⑨작성 주기	☞ 작성 주기 선택 (일, 주, 월, 분기, 반기, 년, 일회성, 수시)				
	⑩데이터 구축·운영 방법	☞ 기 개방 데이터를 제외한 보유 데이터 구축·운영 방법 선택 ① 홈페이지 ② 자체 정보시스템 ③ 비공개				
	⑪시스템명	☞ ② 정보시스템 구축 또는 운영·관리되고 있는 경우만 입력				
	⑫제공 예외조건 포함여부	☞ 보유 데이터 제공 예외조건 포함여부 (Y/N) 엑셀(시트4. 제공 예외조건) 참고				
	⑬컬럼(변수)명	☞ 보유 데이터의 컬럼(변수)명 입력 EX) 창작신작품수(건), 기존국내작품수(건), 공연회수(일) 등				
	⑭데이터 확장자	☞ 데이터 확장자 선택 엑셀(시트5. 데이터 확장자)참고				

〈표 4-16〉 수요 데이터 전수조사 양식

기관명	☞ 응답 기관 명칭		
작성자 정보	소속(부서명)	성명	연락처
①활용 여부	☞ 타 기관 데이터 업무 활용 여부 (이미 활용하고 있음, 활용예정, 필요할 것으로 예상)		
②데이터 명칭(국문)	☞ 데이터의 성격을 파악할 수 있는 구체적인 명칭을 기입 EX) 사업체노동력조사		
③데이터 보유기관	☞ 데이터 보유기관 EX) 고용노동부		
④정보시스템명	☞ 정보시스템명 EX) 고노동통계시스템		
⑤공동활용 주제분류	☞ 아래 5가지 분야 중에 1가지 선택 1. 국민의 의견수렴, 2. 특화된 대책 및 맞춤형 서비스, 3. 위험요소 예측 제거 4. 미래수요에 대한 선제적 대응, 5. 행정업무 경제성·효율성 개선 중 선택 <b>엑셀</b> (시트2. 공동활용 정책분야참고)		
⑥활용목적	EX) 관광분야 일자리 정책수립 및 관련 정책을 지원하기 위해 관광분야 세부 산업별 고용현황(종사자수, 임금 등) 파악 필요. 공공데이터 조화결과 한국표준산업분류 기준 중분류 수준 통계표만 제공하여 세부 산업별 정확한 고용 현황파악이 곤란. 산업별 소분류 또는 세분류 수준 자료의 공동활용이 가능하다면 일자리 정책의 효과성을 모니터링할 수 있는 세부 지표로 활용		
⑦컬럼(변수)명	☞ 수요 데이터에 포함되어야할 컬럼(변수)명 입력 EX) 산업분류, 종사자수, 임금 및 근로시간		

〈표 4-17〉 조사표 작성 참고 자료(공동활용 정책분야)

No.	정책분야	내용
1	국민의 의견 수렴	주요 정책을 수립하거나 경제적·사회적 문제 등을 해결하기 위하여 국민의 의견을 신속하게 수렴할 필요가 있는 분야
2	특화된 대책 및 맞춤형 서비스	특정 계층·지역·분야 등에 대한 비교 및 분석 등을 통하여 특화된 대책을 마련하거나 맞춤형 서비스가 필요한 분야
3	위험요소 예측 제거	안전사고, 질병 등 사전에 위험요소와 원인을 예측하고 제거방 법을 제시할 필요가 있는 분야
4	미래수요에 대한 선제적 대응	정치적·경제적·사회적 및 문화적으로 다양한 미래 수요를 충족하기 위하여 선제적으로 대응할 필요가 있는 분야
5	행정업무 경제성·효율성 개선	비용 절감이나 처리 절차의 개선 등을 통하여 행정 업무의 경제성과 효율성을 증가시킬 필요가 있는 분야

자료: 「데이터기반행정 활성화에 관한 법률」 제8조 제 1항

〈표 4-18〉 조사표 작성 참고 자료(제공 예외조건)

공동활용데이터 등록 예외 근거(법 제4조, 법 제11조)

구분	법률	주요내용
데이터 제공 범위	제11조 제1항	공공기관의 장은 제10조제1항에 따라 제공 요청을 받은 데이터가 해당 공공기관이 생성하거나 취득하여 관리하는 데이터인 경우에는 이를 제공 하여야 한다. 다만, 다음 각 호의 어느 하나에 해당하는 경우에는 데이터를 제공하지 아니할 수 있다.
	1호	데이터가 다른 법률 또는 다른 법률에서 위임한 명령 (대통령령 및 조례만 해당한다.)에서 비밀로 규정된 경우
	2호	국가안전보장 또는 국방·통일·외교관계 등에 관한 데이터로서 이를 제공 할 경우 국가의 중대한 이익을 크게 해칠 우려가 있다고 인정되는 경우
	3호	다른 법령에 따라 데이터의 목적 외 이용이 제한되거나 금지되는 경우

자료: 「데이터기반행정 활성화에 관한 법률」 제4조, 제11조

〈표 4-19〉 조사표 작성 참고 자료(정부기능분류체계\*, BRM)

정책분야(15)	정책영역(65)	대기능(484)
공공질서 및 안전	경찰	
	법무 및 검찰	
	안전관리	
	해 경	
과학기술	과학기술진흥	
	과학기술 일반	

정책분야(15)	정책영역(65)	대기능(484)
	기술개발	
교육	고등교육	
	교육 일반	
	유아 및 초·중등교육	
	평생·직업교육	
국방	방위력개선	
	병력운영	
	전력유지	
농림해양수산	농업·농촌	
	임업·산촌	
	해양수산·어촌	
문화체육관광	문화예술	32개 대기능
	체 육	
	관광	
	문화재	
	문화체육관광일반	
보건	건강보험	
	보건의료	
	식품의약품안전	
사회복지	공적연금	
	기초생활보장	
	노동	
	노인·청소년	
	보건의료	
	보육·가족 및 여성	
	보훈	
	사회복지 일반	
	주택	
	취약계층지원	
산업·중소기업	무역 및 투자유치	
	산업·중소기업 일반	
	산업금융지원	
	산업기술지원	
	산업진흥·고도화	
	에너지 및 자원개발	
⋮	⋮	
환경보호	⋮	

\* 정부기능분류체계: 행정기관이 상시적으로 수행하는 업무를 정책분야·정책영역·대기능·중기능·소기능 및 단위과제 등 그 기능의 수준에 따라 분류한 체계를 말함

## 2. 조사결과

### 2.1. 정부기능분류 대기능 재분류

행정안전부에서 운영·관리하고 있는 공공데이터포털 및 온나라시스템 내 업무매뉴얼에서 확인할 수 있는 정부기능분류체계(BRM)를 기준으로 보유 데이터가 정부기능분류 중 어디에 포함되는지를 응답할 수 있도록 조사표를 설계하였다. 조사결과를 취합·정리한 결과, 정부기능분류체계 중 소분류 단위인 32개 대기능 각각의 케이스별로는 실제 응답된 사례가 전혀 없거나, 1~2개 이내로 응답된 사례가 다수 발생하였다.

실제 32개 대기능으로 세분화하여 응답된 사례를 살펴보면 다음과 같다. 먼저, 학교 운영 등 관련 국악·국어 및 예술 교육 등의 기능을 가지고 있는 기관의 경우 한국예술종합학교를 제외하면 국립국악원 운영, 국립국어원 운영, 예술원 운영 등에 포함될 수 있는 데이터는 정확한 2개만이 조사됐으며, 실제 응답된 2개 데이터 모두 대기능 수준에서 문화정책으로 응답하였다. 이 외에도 국립도서관 운영, 국립박물관 운영, 국립미술관 운영 등과 같이 문화기반시설에서 고유의 행정업무 수행을 통해 보유 또는 생성, 취득하여 운영·관리되고 있는 데이터에 대한 응답 사례는 국립중앙도서관 3개, 국립현대미술관 1개, 국립박물관 1개 등 총 5개 이내의 조사결과만이 집계되었다. 마지막으로 정부기능분류 정책분야는 문화체육관광으로 분류되며 정책영역에서 문화재로 분류되어 최종 대기능에서는 궁·능·원 보존관리, 매장문화재 보존 및 보호, 무형문화재보호, 문화재 조사·연구·복원 등 9개 소분류 수준으로 분류되고 있는 문화재 영역 관련 데이터에 대한 응답은 단 1건도 없는 것으로 집계되었다. 이와 같이 대기능 수준에서 응답한 사례가 전혀 없거나, 1~2개 이내로 응답된 사례가 다수 발생한 원인은 5개영역으로 분류되고 있는 정책영역과 달리, 대기능의 경우 32개의 업무 기능 단위로 세분화되어 있기 때문으로 판단된다. 특히, 문화예술 영역에 하위분류에 포함되는 대기능은 18개 이상으로 전체 대기능 분류단위의 56% 이상을 차지하여 다른 정책영역(체육, 관광 등)의 하위분류(체육 대기능 1개, 관광 대기능 3개) 보다 크게 세분화되어 있는 것으로 분석됐다.

이에 본 연구에서는 문체부 소속 및 산하기관이 보유하고 있는 데이터 전수조사 결과를 효과적으로 분석하기 위해 정부기능분류체계 중 대기능 분류를 <표 4-20>과 같이 10개로 재분류한 후 조사결과 분석을 진행하였다.

〈표 4-20〉 정부기능분류체계 대기능 재분류

정책분야	정책영역(5)	대기능(32)	대기능 재분류(10)
문화체육관광	문화예술	문화산업정책	문화산업정책
		문화미디어정책	문화미디어정책
		문화정책	문화정책 데이터 전수조사 결과 한국예술종합학교를 제외하면 국립국악원/국립국어원/예술원 운영 등에 포함될 수 있는 데이터는 2개가 조사됨(2개 모두 문화정책으로 응답함)
		국립국악원 운영	
		국립국어원 운영	
		문화중심도시 조성	
		예술원 운영	
		예술종합학교 운영	
		예술정책	예술정책
		국립대한민국관건립추진단 운영	문화기반시설 운영·관리 *도서관, 박물관, 미술관 등과 같은 문화기반시설에서 고유의 행정업무 수행을 통해 보유 또는 생성, 취득하여 운영·관리하는 데이터를 통합
		국립민속박물관 운영	
		국립박물관 운영	
		국립중앙극장 운영	
		국립중앙도서관 운영	
		국립현대미술관 운영	
		대한민국역사박물관 운영	
		도서관박물관정책기획단 운영	
		종교문화지원	
		종교문화지원	종교문화지원
	체 육	체육정책	체육정책
	관광	관광레저도시 육성	관광정책
		관광정책	
		사행산업통합감독위원회 운영	사행산업통합 감독위원회운영
	문화재	궁·능·원 보존관리	대기능 분류 제외 문화재 관련 업무 소관은 문화재청이며, 데이터전수조사 결과 문화재에 포함되는 데이터를 제출한 기관은 없음
		동산·건조물문화재 보존관리	
		매장문화재 보존 및 보호	
		무형문화재보호	
		문화재 조사·연구·복원	
		문화재 행정지원	
		문화재정책기반조성	
		사적·자연문화재 보존관리	
		전통문화 전문인력 양성	
	문화체육관광일반	문화관광 행정지원	문화관광 행정지원

## 2.2. 보유 데이터 조사

〈표 4-21〉 보유 및 수요 데이터 전수조사 응답(참여) 기관

정책영역	기관명(중복*)	합계
문화예술	국악방송, 게임물관리위원회, 국립국어원, 국립장애인도서관, 국립장애인도서관, 세종학당재단, 한국문화관광연구원, 한국문화예술교육진흥원, 한국문화예술교육진흥원, 한국언론진흥재단, 한국예술종합학교, 한국저작권보호원, 한국저작권위원회, 한국출판문화산업진흥원, 한국콘텐츠진흥원	16개
체육	국민체육진흥공단, 한국스포츠정책과학원, 한국체육산업개발	3개
관광	한국관광공사, 한국도박문제관리센터, 한국문화관광연구원	3개
문화체육관광 일반	게임물관리위원회, 한국문화관광연구원, 한국체육산업개발	3개
문화체육관광부를 제외한 소속 및 산하기관 48개 기관 중 총 21개 기관이 전수조사에 응답(참여)		

\* 중복: 하나의 기관에서 정부기능분류(정책영역) 기준 정책영역별 복수의 데이터를 보유 또는 운영·관리하는 경우가 있음  
(ex. 한국문화관광연구원은 문화예술영역 데이터 00종, 관광영역 데이터 00종을 보유하고 있음)

문화체육관광부 소속 및 산하기관별로 보유 또는 운영·관리하고 있으나, 외부에 개방(공개)하지 않은 보유 데이터 조사 결과를 살펴보면, 소속 및 산하기관 21곳에서 총 452개의 데이터를 보유 또는 운영·관리하고 있었으며, 데이터 생성·취득 방법별로는 2차가공 208개(46.0%), 직접조사 147개(32.5%), 행정자료 97개(21.5%) 순으로 나타났다.

〈표 4-22〉 문체부 소속·산하기관 생성·취득 방법별 보유 데이터 현황

(단위: 개, %)				
구분	직접조사	2차가공	행정자료	합계
보유 데이터(개)*	147	208	97	452
비율(%)	32.5	46.0	21.5	100.0

\* 보유 데이터: 외부 개방 데이터 시스템(공공데이터포털)을 통해 개방(공개)하지 않은 데이터를 대상으로 조사된 수치임.  
단, ① 기관 홈페이지 및 자체 정보시스템에서만 공개한 데이터, ② 전체 데이터 중 일부만 외부에 개방(공개)된 데이터  
(ex. 개인정보 포함으로 일부만 개방)는 조사된 보유 데이터에 포함

정부기능분류 기준 정책분야(문화체육관광)와 정책영역(4개영역)으로 구분한 보유 데이터 조사 결과를 살펴보면 총 452개의 보유 데이터 중 문화예술 관련 데이터가 345개(76.3%)로 가장 많았고, 다음으로 관광 관련 데이터가 45개(10.0%), 체육 40개(8.8%) 순으로 조사됐다.



〈표 4-23〉 정부기능분류\*별(정책영역) 생성·취득 방법별 보유 데이터 현황

(단위: 개, %)

구분	직접조사	2차가공	행정자료	합계(%)
문화예술	112	156	77	345(76.3)
체육	22	9	9	40(8.8)
관광	10	24	11	45(10.0)
문화체육관광 일반	3	19		22(4.9)
합계	147	208	97	452(100.0)

\* 정부기능분류: 중앙행정기관(대통령 및 국무총리 소속기관을 포함하며, 이하 "행정기관"이라 한다)의 업무 및 그와 관련된 정보를 체계적으로 관리·활용하기 위하여 행정기관의 업무를 기능 중심으로 분류하는 것을 말함(행정안전부 공공데이터포털)

정부기능분류를 기준으로 업무 기능별(대기능)로 세분화하여 집계한 보유 데이터 현황을 살펴보았다. 먼저 각 정책영역별 보유 데이터 현황을 살펴보면, 문화예술 정책영역에서는 문화산업정책(154개, 44.6%)과 문화정책(103개, 29.9%) 데이터가 전체 문화예술 정책영역 데이터의 약 75% 비중을 차지하는 것으로 나타났으며, 관광영역에서는 관광정책(42개) 데이터가 93.3%로 높은 비중을 보이는 것으로 나타났다.

〈표 4-24〉 정부기능분류별(대기능) 보유 데이터 현황

(단위: 개, %)

정책영역	데이터 수	대기능	데이터 수
문화예술	345(100.0)	문화정책	103(29.9)
		예술정책	44(12.8)
		문화기반시설 운영관리	12(3.5)
		종교문화지원	
		문화산업정책	154(44.6)
		문화미디어정책	32(9.3)
체육	40(100.0)	체육정책	40(100.0)
관광	45(100.0)	관광정책	42(93.3)
		사행산업통합 감독위원회운영	3(6.7)
문화체육관광 일반	22(100.0)	문화관광 행정지원	22(100.0)

이어서 업무 기능별(대기능) 보유 데이터 현황을 살펴보면, 전체 문화체육관광 분야 데이터 452개 중 문화산업정책 데이터가 154개(34.1%)로 가장 많았고, 다음으로 문화정책 데이터가 103개(22.8%), 예술정책 44개(9.7%), 관광정책 42개(9.3%), 등의 순으로 데이터를 보유 또는 운영·관리하고 있는 것으로 나타났다. 세부 업무 기능별 보유 데

이터 중 종교문화지원 관련 데이터의 경우는 기 시스템(행정안전부 공공데이터포털 등)에 공개된 데이터 외에 외부에 개방(공개)되지 않은 신규 데이터는 없는 것으로 조사됐다.

데이터 생성·취득방법에 따른 업무 기능별 보유 데이터 현황을 살펴보면, 직접조사를 통해 생성·취득 및 보유하고 있는 데이터는 문화예술(14.6%), 2차 가공된 데이터는 문화산업정책(28.5%), 행정자료는 예술정책(5.8%) 관련 업무에서 데이터 비중이 높은 것으로 나타났다.

〈표 4-25〉 정부기능분류별(대기능) 생성·취득 방법별 보유 데이터 현황

(단위: 개, %)

구분	직접조사		2차가공		행정자료		합계	
	데이터 (개)	비율 (%)	데이터 (개)	비율 (%)	데이터 (개)	비율 (%)	데이터 (개)	비율 (%)
문화정책	66	14.6	11	2.4	26	5.8	103	22.8
예술정책	7	1.5	8	1.8	29	6.4	44	9.7
문화기반시설 운영관리	6	1.3	3	0.7	3	0.7	12	2.7
종교문화지원								
문화산업정책	24	5.3	129	28.5	1	0.2	154	34.1
문화미디어정책	9	2.0	5	1.1	18	4.0	32	7.1
체육정책	22	4.9	9	2.0	9	2.0	40	8.8
관광정책	10	2.2	21	4.6	11	2.4	42	9.3
사행산업통합 감독위원회운영			3	0.7			3	0.7
문화관광 행정지원	3	0.7	19	4.2			22	4.9
합계	147	32.5	208	46.0	97	21.5	452	100.0

공동활용 정책분야별 보유 데이터 현황을 정부기능분류별 정책영역과 교차해서 살펴 보았다. 5가지의 공동활용 정책분야 중 “특화된 대책 및 맞춤형 서비스” 관련 데이터가 342개(75.7%)로 가장 많았고, 다음으로 “행정업무 경제성·효율성 개선” 49개(10.8%), “미래수요에 대한 선제적 대응” 37개(8.2%) 순으로 공동활용 정책분야 데이터 비중이 높은 것으로 조사됐다. “국민의 의견 수렴” 관련 보유 데이터는 6개로 전체 데이터의 1.3% 수준으로 매우 미미한 수준을 나타낸 가운데, 5가지 공동활용 정책분야 모두에서 문화예술 영역으로 분류되는 데이터의 비중이 가장 높은 것으로 나타났다.

〈표 4-26〉 공동활용 정책분야별 정부기능분류별(정책영역) 보유 데이터 현황

(단위: 개, %)

구분	문화예술	체육	관광	문화체육관광 일반	합계(%)
국민의 의견 수렴	5	1			6(1.3)
특화된 대책 및 맞춤형 서비스	238	38	45	21	342(75.7)
위험요소 예측 제거	18				18(4.0)
미래수요에 대한 선제적 대응	35	1		1	37(8.2)
행정업무 경제성·효율성 개선	49				49(10.8)
합계	345	40	45	22	452(100.0)

정부기능별(대기능) 보유 데이터 현황을 공동활용 정책분야별로 구분하여 조사결과를 살펴보았다. 전체 문화체육관광 분야 데이터 452개 중 문화산업정책 데이터가 154개로 가장 많은 것으로 나타난 가운데, 문화산업정책 데이터의 49% 이상이 “특화된 대책 및 맞춤형 서비스” 관련 정책분야에서 공동활용이 가능한 데이터로 조사되었다. 문화산업정책을 다른 업무 기능분류와 비교해서 살펴볼 경우에도 5가지 공동활용 정책분야 모두에서 문화산업정책 관련 업무 데이터의 비중이 가장 높은 것으로 분석됐다.

〈표 4-27〉를 살펴보면, 분류된 데이터 수가 “0”인 셀이 상당수 존재하고 있다. 이는 대부분의 데이터가 상대적으로 분류기준이 명확하고 구체적인 “특화된 대책 및 맞춤형 서비스” 활용 목적으로 분류되어 응답했기 때문으로 판단되지만, 행정 업무 기능별로 다른 정책영역에서 활용될 수 있는 데이터에 대한 고려 및 생산·관리가 적절히 이뤄지지 않고 있기 때문으로도 생각해 볼 수 있다. 그 밖에도 “국민의 의견 수렴” 분야에 해당하는 데이터는 전체 452개 데이터 중 6개 밖에 보유 데이터가 없는 것으로 나타났으며, “위험요소 예측제거” 데이터는 문화산업정책을 제외하고는 다른 업무 기능에는 단 1개의 데이터도 없는 것으로 조사됐다

〈표 4-27〉 정부기능분류별(대기능) 공동활용 정책분야별 보유 데이터 현황

(단위: 개, %)

구분	국민의 의견 수렴	특화된 대책 및 맞춤형 서비스	위험요소 예측 제거	미래수요에 대한 선제적 대응	행정업무 경제성· 효율성 개선	합계(%)
문화정책		97		6		103
예술정책	4	29		2	9	44

구분	국민의 의견 수렴	특화된 대책 및 맞춤형 서비스	위험요소 예측 제거	미래수요에 대한 선제적 대응	행정업무 경제성· 효율성 개선	합계(%)
문화기반시설 운영관리		12				12
종교문화지원						
<b>문화산업정책</b>		<b>76(49.4)</b>	<b>18(11.7)</b>	<b>20(13.0)</b>	<b>40(26.0)</b>	<b>154(100.0)</b>
문화미디어정책	1	24		7		32
체육정책	1	38		1		40
관광정책		42				42
사행산업통합 감독위원회운영		3				3
문화관광 행정지원		21		1		22
합계	6	342	18	37	49	452

### 2.3. 수요 데이터 조사

문화체육관광부 소속 및 산하기관별로 업무 수행 시 필요한 타 기관 데이터에 대한 수요조사를 보유 데이터 조사와 함께 실시하였다. 8개 기관에서 총 49개 데이터를 업무 수행 시 필요한 타 기관 데이터로 응답하였으며, 공동활용 정책분야별로는 “특화된 대책 및 맞춤형 서비스(32개)” 분야에 활용할 수 있는 데이터에 대한 수요가 가장 많은 것으로 나타났다.

〈표 4-28〉 정책분야별 공동활용 가능 수요 데이터 현황

(단위: 개)

구분	국민의 의견 수렴	특화된 대책 및 맞춤형 서비스	위험요소 예측 제거	미래수요에 대한 선제적 대응	행정업무 경제성· 효율성 개선	합계
수요 데이터	2	32	2	7	6	49

〈표 4-29〉 정책분야별 공동활용 가능 수요 데이터 세부 현황

구분	공동활용 가능 데이터	데이터 보유기관	활용 목적
국민의 의견 수렴	국민여가활동조사	한국문화관광연구원	여가활동 제약 및 참여 요구, 여가활동으로서의 스포츠참여 현황 등의 자료를 국민체육진흥정책 개발에 활용
	민원·청원 등 의견을 함의한 데이터	각 기관별 홈페이지, 청와대	예술인 민원 데이터에 대한 키워드 분석 및 시간 대별, 지역별 민원 데이터 분석을 통한 민원유형과 빈도수 현황 파악/타 직업군과 비교분석
특화된 대책 및 맞춤형 서비스	예술 분야 고용보험DB	고용노동부 (한국고용정보원)	예술인 복지정책 수립을 위해 고용보험에 가입된 예술인들의 인원수, 지역별 분포, 소득, 실업급여 수혜 현황 자료 등 파악 필요
	사업체노동력조사		관광분야 일자리 정책수립 및 관련 정책을 지원하기 위해 관광분야 세부 산업별 고용현황(종사자수, 임금 등) 파악 필요. 공공데이터 조회결과 한국표준산업분류 기준 중분류 수준 통계표만 제공하여 세부 산업별 정확한 고용 현황파악이 곤란, 산업별 소분류 또는 세분류 수준 자료의 공동활용이 가능하다면 일자리 정책의 효과성을 모니터링할 수 있는 세부 지표로 활용
	예술 분야 직종 소득정보DB	국세청	예술인 복지정책 수립을 위해 건강보험에 가입된 예술인의 가입자 정보 파악 필요
	국세통계		관광사업체의 월별 변화를 분석하기 위해 국세통계의 월별 사업체수, 신규사업체수, 폐업사업체수 통계 활용 필요(현재 대분류 업종 자료가 공개, 관광사업체의 특성을 분석하기 위해서는 세세분류 업종에 해당하는 정보가 필요함, 지방행정개방데이터 보완자료로 활용)
	예술활동증명DB	한국예술인복지재단	지역별 예술인 복지정책 수립을 위해서는 지역 예술인들의 현황 파악이 중요한데, 현재는 지역단위 예술활동증명자료 예술인에 대한 정보가 공개되지 않고 있음
	예술인의 저작권 수입DB	한국저작권위원회	예술인 복지정책 수립을 위해 예술인들의 주요 수입원 중 하나인 저작권 수입에 대한 현황 파악이 필요
	고용보험DB	고용노동부 (한국고용정보원)	문화예술분야 고용현황 파악과 정책 수립을 위해 문화예술분야 데이터 필요 분야정보가 제공되고 있지 않고, 총량만 제공되고 있어 문화예술분야 데이터 공동활용 필요
	노인실태조사	보건복지부	노인의 여가문화활동 참여 행태와 운동참여 장소(노인여가복지시설 등) 선호도 등을 파악하여 노인체육진흥 개발 자료로 활용
	아동종합실태조사		조사대상이 0~18세로, 국민생활체육조사의 조사대상(10세이상)이 아닌 0~9세까지의 운동참

구분	공동활용 가능 데이터	데이터 보유기관	활용 목적
			여 현황, 방과후 희망 활동 등의 자료를 파악, 정책개발의 기초 자료로 활용
	장애인실태조사		장애유형별 장애인 현황, 발생률 등 실태를 파악하여 장애인체육정책수립시반영
	건강보험통계	국민건강보험공단	의료비 지출 관련 정보와 생활체육 참여 자료를 결합하여 체육활동의 경제적 효과를 측정하고자 함
	가계동향조사	통계청	가계의 스포츠 관련 지출(용품, 서비스 등)을 분석하여 생활체육 및 스포츠산업 정책 수립의 기초 자료로 이용 맞춤형 정책 설계를 위해 가계지출동향의 세부 현황파악 필요
	생활시간조사		국민의 시간 활용 현황을 파악하여 생활체육 참여 활성화 정책의 근거 자료로 이용
	기업생멸행정통계		시간 흐름에 따른 관광사업체의 변화를 분석하기 위해 기업생멸행정통계의 신생률, 소멸률, 신생기업 생존율, 고성장기업 비율 등 통계 활용 필요(현재 대분류 업종 자료가 공개, 관광사업체의 특성을 분석하기 위해서는 세세분류 업종에 해당하는 정보가 필요함)
	지식재산권수지	한국은행	맞춤형 정책 설계를 위해 지식재산의 무역의 세부 현황파악 필요
	이동통신데이터	통신 3사 (SK, KT, LG)	통신, 내비게이션 등의 데이터를 활용. 방문 예정일의 관광지별 관광객 수 예측 분석을 기반으로 혼잡도 지표 등 분석·도출
	공공스포츠클럽 운영 현황	대한체육회	기존 시스템에서는 지역별 클럽 배치 현황 및 운영을 클럽별로 제공하여 전반적인 공공스포츠 클럽 운영 및 성과를 종합적으로 판단하는 데 한계가 있음. 기금 지원이 종료된 클럽을 포함하여 모든 클럽의 인력 및 예산(수입, 지출), 연령별 참여자 수, 운영프로그램 등을 통합한 자료의 공동 활용이 가능하다면 공공스포츠클럽 사업의 정책연구 및 지원에 도움을 줄 것이라 판단됨
	장애인생활체육 지원사업 만족도 조사	대한장애인체육회	대한장애인체육회 장애인생활체육지원사업인 교실프로그램, 지도자배치, 대회지원 등 장애인 참여자들의 만족도 조사를 통해 장애인생활체육 지원사업의 운영성과와 개선 방향을 도출하기 위해 활용
	장애인생활체육 실태조사		장애인생활체육 참여 실태를 파악하여 장애인 생활체육활성화 정책 수립 시 활용
	근로자휴가조사	한국문화관광연구원	여가활동 참여 행태와 스포츠 참여 행태 등의 자료를 이용하여 생활체육 정책 수립의 기초자료로 활용

구분	공동활용 가능 데이터	데이터 보유기관	활용 목적
	국민여가활동조사		스포츠 이외 여가 활동 유형 및 참여 현황 비교
	문화체육관광 분야 서비스업생산지수		스포츠산업 분야 서비스업 지원 방안 확보를 위 한 근거자료 활용
	콘텐츠산업통계조사 원자료	문화체육관광부	<ul style="list-style-type: none"> <li>- 기관에서는 매년 '출판산업 실태조사'(비승인 조사통계)를 통해, 연간 출판업계 주요 매출 규모/종사자 수 등을 조사 및 추정하여 발표· 활용하고 있음</li> <li>- 해당 조사의 상위분류이자 국가승인통계인 '콘 텐츠산업통계조사'와 자체조사('출판산업 실태 조사')의 결과 일관성 확보 및 상호 보완 활 용을 위해, 서로 다른 두 조사의 전반적 설계 와 구체적 방법(표본설계 및 선정대상, 추정방 식)을 비교·검토하여 자체 조사통계의 보완점 파악 및 개선, 통계품질의 제고가 시급함</li> <li>- 현재 공개되는 자료(콘텐츠산업통계조사 결과 보고서)에서는 기 분석된 통계표만 제공·조회 가능하여, 그 활용이 다소 제한적.</li> <li>- 콘텐츠산업조사 내 출판 분야에 대한 각 조사 항목별(매출액, 종사자 수, 부가가치액 등) 표 본구성(안) 및 조사대상 사업체 목록 또는 관 련 원자료의 공동활용이 가능하다면, 출판산 업과 관련된 기초통계자료를 다각적으로 확보 하여 관련 정책개발 및 수립에 높은 활용성을 갖게됨</li> </ul>
	영화 등급분류 목록	영상물등급위원회	저작물의 정보가 담긴 공공데이터를 이용하여 온라인상의 저작권 침해에 대응하기 위한 기초 자료로 활용
	비디오물 등급분류 목록		
	게임 API	게임물관리위원회	
위험요소 예측 제거	신용평가사 신용 평점 데이터	나이스 신용평가	개인의 여가 활동 기반의 개인 파산 고위험 고객 군 사전 예측
	금융 (DTI 등) 활동 상세 정보	금융감독원	개인정보 비식별화를 통해 카지노를 포함한 겜 블성 여가 (경마, 경륜 등) 의 이용 고객에 대한 정보를 민간 신용 평가사의 데이터와 결합하여 개인의 신용도 혹은 개인 파산 위험 고객군을 사 전에 예측할 수 있는 지수 개발
미래수요에 대한 선제적 대응	영화정보	영화진흥위원회	KMDb 영화정보DB 구축 및 수집자료(납본자 료, 기증자료 등) 목록DB 구축에 참고, 활용.
	영화인정보		KMDb 영화인DB 구축에 참고, 활용.
	상영현황정보		KMDb 영화정보DB와 융합 서비스
	영화사		KMDb 영화정보DB와 융합 서비스
	영화제		KMDb 영화/영화인정보DB와 융합 서비스

구분	공동활용 가능 데이터	데이터 보유기관	활용 목적
	서지정보	국립중앙도서관	수집자료(납본자료, 기증자료 등) 목록DB 구축에 참고, 활용
	교통정보, 고속도로 통행량	교통연구원, 한국도로공사	관광지 정보와 교통정보를 연계하여 축제 수요 분석 및 지역 경제 활성화 전략 수립
행정업무 경제성·효율성 개선	장애인증명 정보	보건복지부	국립장애인도서관 및 장애인 대상 도서관 서비스를 제공하는 공·사립 도서관에서 회원의 장애 사실여부를 온라인으로 확인 가능하도록 하여 서류 제출에 따른 이용자의 불편을 감소시키는 데 목적이 있음
	학교 교사 정보	한국교육학술정보원	학교 문화예술교육 정책을 지원하기 위해 학교 분야 학교 교사 정보(교원번호, 성명, 근무학교명 등) 파악 필요. 학교 예술강사 지원사업 학교 신청 및 예술강사 출강 승인 관련 업무를 위하여 학교담당자 권한 부여 필요, 현행 학교담당자 확인을 위하여 지역센터에서 학교 홈페이지 직원검색을 통하여 확인 후 권한 부여. 학교 교사 자료의 공동활용이 가능하다면 학교예술강사지원사업, 교원 행정 연수 관련 행정업무 경제성·효율성이 개선될 것으로 사료됨.
	국고 보조금 현황	한국재정정보원	보조금의 효율적 운영을 위해 보조금의 세부 현황 파악 필요
	국가자료종합목록 소장자료	국립중앙도서관	국가자료종합목록의 전국 공공도서관 소장자료 정보를 국립장애인도서관 책나래 서비스 제공에 활용 (도서 검색, 대출 신청)
	국립중앙도서관 소장자료		국립중앙도서관의 소장자료 정보를 국립장애인도서관 소장자료 서지정보 정리 및 이용자 대체 자료 제작 신청에 활용
	지자체별 관광지 이용 현황 정보	지자체별 행정자료	교통정보와 관광지정보를 연계하여 문화재 및 지역 관광지 관람을 위한 순환버스 도입을 위한 현황 및 이용률 파악



## 제4절 데이터 업무 활용 실태조사

---

### 1. 조사개요

데이터 공동활용의 제약조건 및 관계 분석을 위해 데이터 업무 활용 관련 경험 가능성이 높은 문체부 및 소속·산하기관 데이터 및 정책업무 담당자와 연구원을 대상으로 설문 조사를 실시하였다. 표본크기는 연구의 목적과 예산을 고려하여 총200명을 목표로 하여 조사를 진행하였다.

표본추출틀은 문체부 및 소속기관의 경우 공공데이터 담당 공무원 리스트를 활용하였고, 산하기관의 리스트는 각 기관 홈페이지 담당자 및 연구원 리스트를 활용하였다.

조사는 웹서베이 방식이 적용되었다. 본 연구에서 구축한 소속 및 산하기관 조사대상 리스트를 기준으로 웹조사 링크를 전달하여 조사참여를 요청하였다.

자료수집 도구는 구조화된 질문지를 사용하였으며, 구체적인 설문지 내용은 부록으로 첨부하였다. 설문내용의 구성은 심층면접 자료에 기초하였으며, 설문지의 타당성 확보를 위해 5인의 전문가 검토를 거쳤다. 조사기간은 2021년 8월 9일부터 8월 27일까지 19일 동안 진행되었다.

데이터 업무 활용 실태조사의 궁극적인 조사목적은 1차적으로 데이터기반행정 공동 활용 기반 구축을 위해 데이터기반행정 관점에서 데이터 공동활용 및 제약조건의 실태를 파악하는 것이다. 2차적으로는 4장 제1절의 “보유 및 수요 데이터 전수조사”의 조사 목적과 같이 문체부 소속 및 산하기관 간에 공동활용 데이터 이용 활성화 방안을 모색해 보고 나아가서는 데이터기반행정 추진기반 및 역량 강화 방안을 마련하고자 하는 것이다.

## 가. 조사 설계

데이터 업무 활용 실태조사는 문체부 소속 및 산하기관별 데이터 및 정책업무 담당자, 연구원을 대상으로 구조화된 설문지를 이용하여 웹서베이 방식으로 조사를 진행하였다.

〈표 4-30〉 조사 설계 개요

구 분	내용
조사 근거	<ul style="list-style-type: none"> <li>「문화체육관광부 데이터 관리 규정」제12조(문화체육관광부훈령 제430호),</li> <li>「데이터기반행정 활성화에 관한 법률」제20조</li> </ul>
조사 대상	문화체육관광부 소속 및 산하기관 데이터 및 정책업무 담당자와 연구원
조사 규모	192명 (유효표본 기준)
표본 오차	95% 신뢰수준에서 3.27%p
조사 방법	구조화된 설문지를 이용하여 웹서베이 방식으로 조사
조사 일정	2021년 7월15일 ~ 7월31일(19일간)

## 나. 조사 내용

설문조사의 내용은 다음과 같다. 데이터기반행정 관점에서 데이터 공동활용 및 제약 조건의 실태 등을 파악하기 위한 목적에 맞게 설문지는 총 5개 주제의 데이터기반행정 실태, 데이터기반행정 인식, 데이터 활용(연계, 결합, 공동활용 등) 실태, 데이터 공동활용 인식, 응답자 특성 등으로 구분하여 구성하였다. 데이터기반행정실태는 데이터 관련 업무 수행 경험, 타 기관 데이터 이용 경험, 데이터 제공 요청 받은 경험, 데이터 공동활용 사업 참여 경험 등의 질문으로 구성하였다.

데이터기반행정 인식은 데이터 활용의 중요도와 데이터 활용도, 데이터기반행정 활성화를 위한 주요 과제의 중요도와 실현 정도 등으로 구성하였다. 데이터 활용 실태의 경우 데이터 공동활용 사업 참여 경험, 사업의 목적과 현황, 참여 데이터 활용 사업의 데이터 종류(민간/공공), 보유주체, 정책분야, 확보 채널 등을 질문하였다. 데이터 공동활용 인식의 경우 데이터 공동활용의 중요성 인식, 데이터 공동활용 제약조건의 심각성 등을 질문하였다.

〈표 4-31〉 조사 항목

구 분	조사 항목
데이터기반행정 실태	<ul style="list-style-type: none"> <li>• 주 활용 데이터(기본 데이터 유형, 정형/비정형 데이터, 공공/민간 데이터, 단일/결합 데이터, 원시/가공 데이터, 결합 데이터 유형, 공공데이터 제공 기관, 공공데이터 정책분야)</li> <li>• 타 기관 데이터 이용 경험 및 경로</li> <li>• 타 기관 데이터 이용 시 데이터 제공 원활정도</li> <li>• 타 기관 데이터 이용 시 데이터 제공 원활하지 않은 이유</li> <li>• 타 기관 데이터 이용 시 데이터 수령 소요기간</li> <li>• 내부/외부 기관으로의 데이터 제공 요청을 받은 경험</li> <li>• 내부/외부 기관으로의 데이터 제공 시 원활정도</li> <li>• 내부/외부 기관으로의 데이터 제공 시 원활하지 않은 이유</li> <li>• 내부/외부 기관으로의 데이터 제공 시 소요기간</li> <li>• 데이터 교육 경험</li> <li>• 데이터 교육 주관기관</li> <li>• 데이터 교육 방식(환경)(온/오프라인)</li> <li>• 데이터 교육 방법(강의/실습/토론)</li> <li>• 데이터 교육 내용</li> <li>• 데이터 자격증 보유</li> <li>• 데이터 자격증 개수</li> <li>• 데이터 자격증 분야</li> </ul>
데이터기반행정 인식	<ul style="list-style-type: none"> <li>• 데이터 활용의 중요도</li> <li>• 업무별 데이터 활용 중요도(정책기획/정책집행/정책평가)</li> <li>• 데이터 활용 정도</li> <li>• 업무별 데이터 활용 정도(정책기획/정책집행/정책평가)</li> <li>• 데이터기반행정 활성화를 위한 주요 과제 중요도</li> <li>• 데이터기반행정 활성화를 위한 주요 과제 실현 정도</li> </ul>
데이터 활용 실태	<ul style="list-style-type: none"> <li>• 데이터 공동활용 사업 참여 경험</li> <li>• 참여 데이터 공동활용 사업 목적 및 현황(사업기간, 예산, 투입인력)</li> <li>• 참여 데이터 활용 사업의 데이터 종류(민간/공공)</li> <li>• 참여 데이터 활용 사업의 데이터 종류 보유주체</li> <li>• 참여 데이터 활용 사업의 활용 데이터 정책분야</li> <li>• 참여 데이터 활용 사업의 활용 데이터 확보 채널</li> </ul>
데이터 공동활용 인식	<ul style="list-style-type: none"> <li>• 데이터 공동활용의 중요성</li> <li>• 데이터 공동활용이 실현 정도</li> <li>• 주요 과제별 데이터 공동활용 제약조건의 심각성</li> <li>• 데이터 공동활용 활성화를 위한 과제</li> </ul>
응답자 특성	<ul style="list-style-type: none"> <li>• 전공</li> <li>• 근무처</li> <li>• 직무</li> <li>• 근무경력</li> </ul>

## 다. 응답자 특성

설문조사의 응답자 특성은 다음과 같다. 전체 응답자 192명의 직무별로는 ‘행정’이 32.8%로 가장 많았으며, 다음으로 ‘연구’ 27.1%, ‘전산/정보’ 22.9%, ‘통계’ 8.9%, ‘기타’ 8.3%를 차지하는 것으로 나타났다.

전공의 경우 ‘전산/컴퓨터공학’이 22.9%로 가장 많았으며, ‘경영학(MIS 포함)’ 10.9%, ‘사회학’ 10.9%, ‘행정학’ 7.8%, ‘경제학’ 5.2%, ‘통계학’ 3.6%, ‘기타’ 38.5%를 차지하는 것으로 나타났다.

근무경력의 경우 ‘3~5년’이 23.4%로 가장 많았으며, 다음으로 ‘1년 미만’이 21.4%, ‘10년 이상’이 21.4%, ‘1~2년’이 18.8%, ‘6~9년’이 15.1%를 차지하는 것으로 나타났다.

〈표 4-32〉 응답자 특성

(단위: %)

구 분		사례수	비율
전체		(192)	100.0
직무	행정	(63)	32.8
	통계	(17)	8.9
	전산/정보	(44)	22.9
	연구	(52)	27.1
	기타	(16)	8.3
전공	행정학	(15)	7.8
	통계학	(7)	3.6
	경제학	(10)	5.2
	경영학(MIS 포함)	(21)	10.9
	전산/컴퓨터공학	(44)	22.9
	사회학	(21)	10.9
	기타	(74)	38.5
근무경력	1년 미만	(41)	21.4
	1~2년	(36)	18.8
	3~5년	(45)	23.4
	6~9년	(29)	15.1
	10년 이상	(41)	21.4

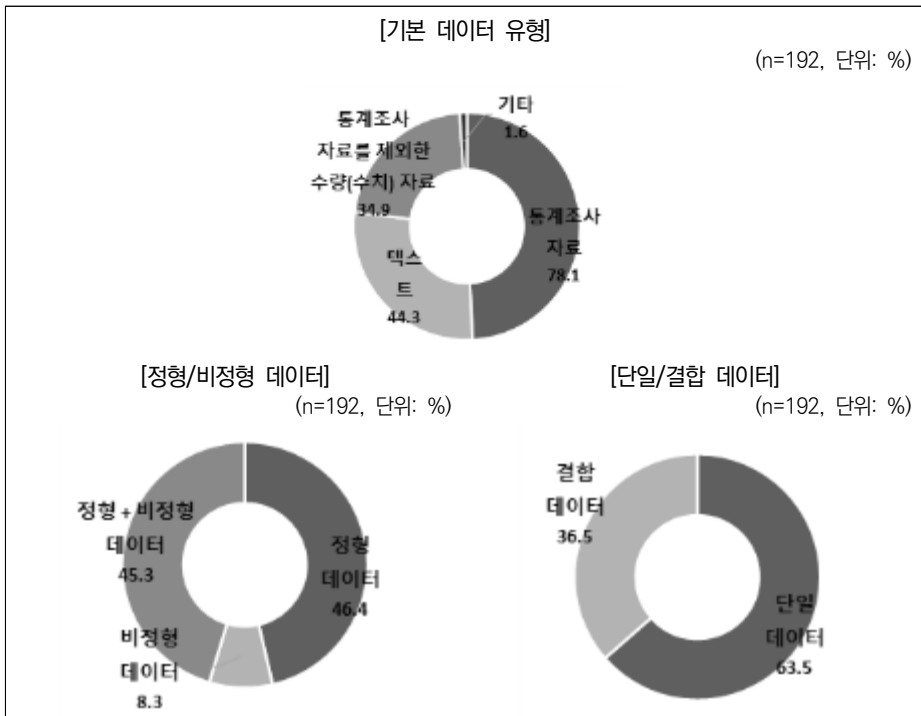
## 2. 조사결과

### 2.1. 데이터 유형

#### 가. 데이터 유형

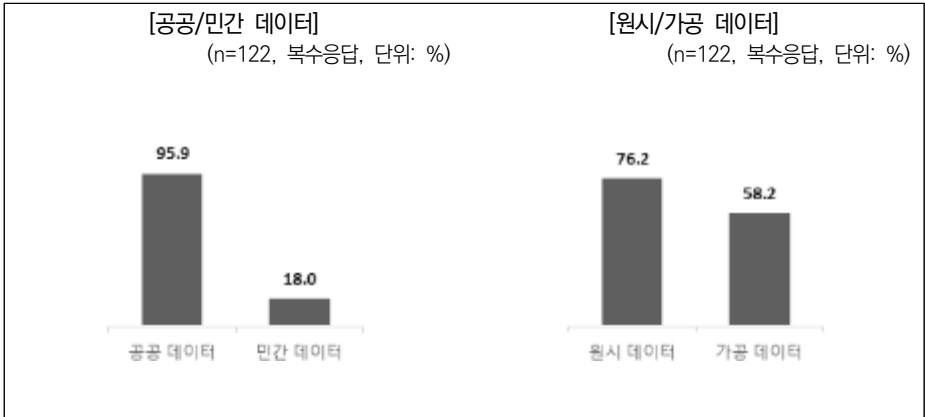
문화체육관광부 소속 및 산하기관 데이터 및 정책업무 담당자와 연구원이 주로 다루는 [데이터 기본 유형](복수응답)은 ‘통계조사 자료’(78.1%)의 비율이 높게 나타났으며, 다음으로 ‘텍스트’(44.3%)와 ‘통계조사 자료를 제외한 수량(수치) 자료’(34.9%)가 높게 나타났다. [정형/비정형 데이터] 중에서는 ‘정형 데이터’(46.4%), ‘정형+비정형 데이터’(45.3%), ‘비정형 데이터’(8.3%)의 순으로 나타났으며, [단일/결합 데이터] 중에서는 ‘단일 데이터’(63.5%), ‘결합 데이터’(36.5%)의 비율로 나타났다.

[그림 4-1] 데이터 기본 유형



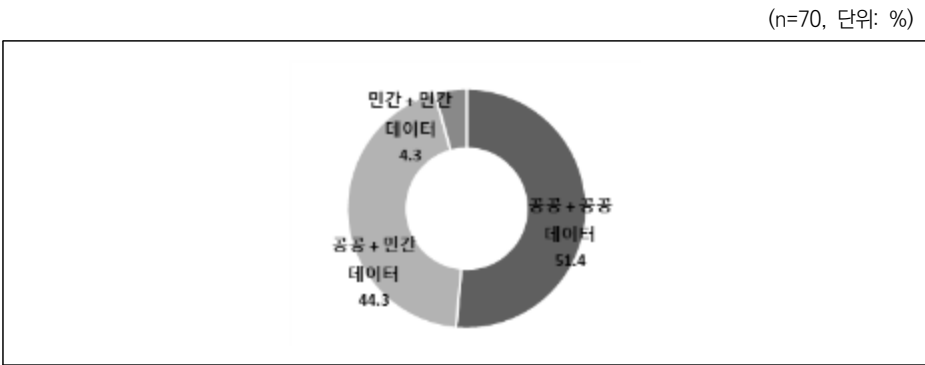
단일 데이터 이용 경험이 있는 응답자를 대상으로 단일 데이터의 유형을 묻은 결과, [공공/민간 데이터]는 ‘공공 데이터’(95.9%), ‘민간 데이터’(18.0%)의 비율로 나타났으며, [원시/가공 데이터]는 ‘원시 데이터’(76.2%), ‘가공 데이터’(58.2%)의 비율로 나타났다.

[그림 4-2] 단일 데이터 유형



결합 데이터 이용 경험이 있는 응답자(n=70, 복수응답)를 대상으로 [데이터의 유형]을 묻은 결과, ‘공공 + 공공 데이터’(51.4%), ‘공공 + 민간 데이터’(44.3%), ‘민간 + 민간 데이터’(4.3%)의 비율로 나타났다.

[그림 4-3] 공공/민간 데이터 결합 유형



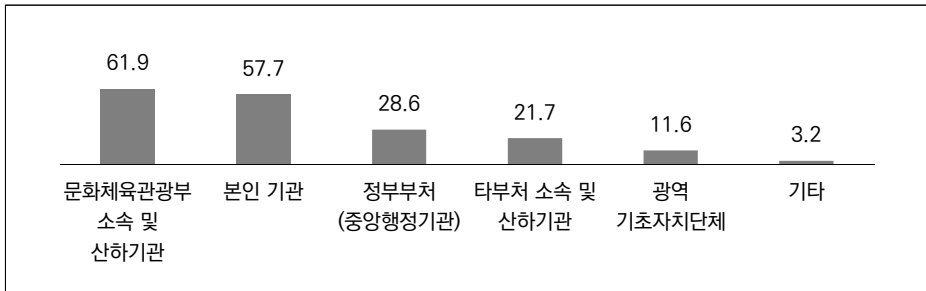
## 나. 공공 데이터

### 1) 공공 데이터 제공 기관

공공 데이터를 이용한 경험이 있는 응답자(n=189, 복수응답)를 대상으로 [공공 데이터 제공 기관]을 묻은 결과, '문화체육관광부 소속 및 산하 기관'(61.9%)의 비율이 가장 높게 나타났으며, 다음으로 '본인 기관'(57.7%), '정부부처(중앙행정기관)'(28.6%), '타부처 소속 및 산하기관'(21.7%), '광역 기초자치단체'(11.6%) 등의 순으로 나타났다.

[그림 4-4] 공공 데이터 제공 기관

(n=189, 복수응답, 단위: %)



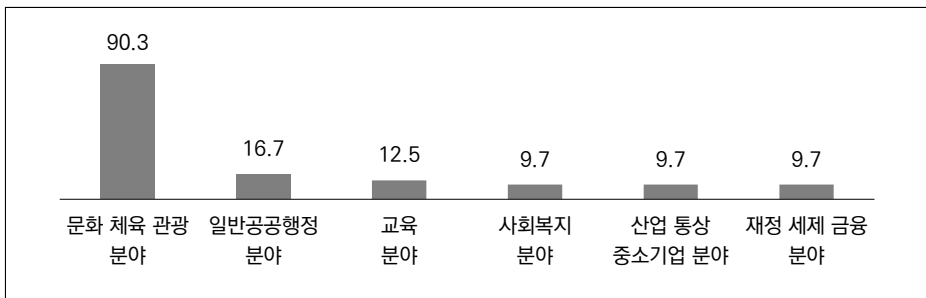
### 2) 공공 데이터 정책분야

타 기관의 공공 데이터를 이용한 경험이 있는 응답자(n=144, 복수응답)를 대상으로 공공 데이터의 [정책분야]를 묻은 결과, '문화 체육 관광'(90.3%)의 비율이 가장 높게 나타났으며, 다음으로 '일반공공행정'(16.7%), '교육'(12.5%) 등의 순으로 나타났다.

[그림 4-5] 공공 데이터 정책분야

(n=144, 복수응답, 단위: %)

\*상위 6개 항목



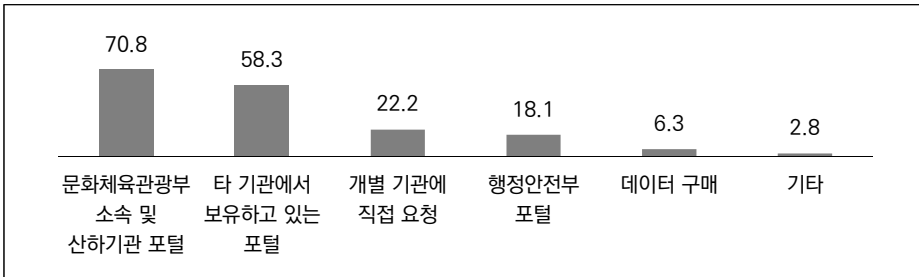
## 다. 타 기관의 데이터 활용 경험

### 1) 외부 기관 공공 데이터 이용 경로

타 기관의 공공 데이터를 이용한 경험이 있는 응답자(n=144, 복수응답)의 [이용 경로]는 ‘문화체육관광부 소속 및 산하기관 포털’(70.8%)의 비율이 가장 높게 나타났으며, 다음으로 ‘타 기관에서 보유하고 있는 포털(홈페이지 포함)’(58.3%), ‘개별 기관에 직접 요청’(22.2%) 등의 순으로 나타났다.

[그림 4-6] 타 기관 공공 데이터 이용 경로

(n=144, 복수응답, 단위: %)

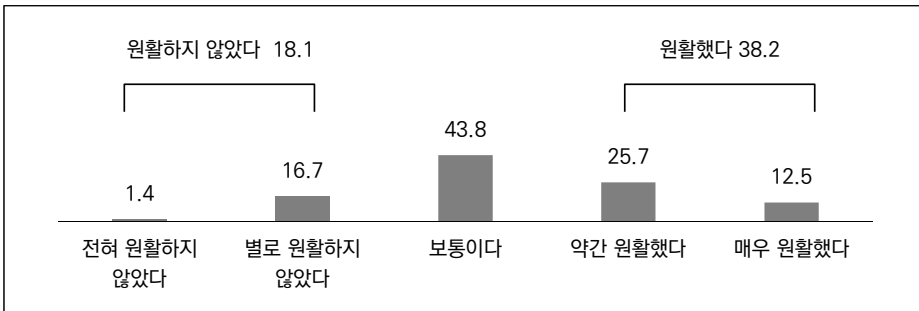


### 2) 외부 기관의 데이터 제공 원활 정도

타 기관의 데이터를 이용한 경험이 있는 응답자(n=144)의 10명 중 3명 정도(38.2%)가 데이터 제공이 ‘원활했다’(약간 원활했다 + 매우 원활했다)고 응답하였고, 타 기관의 데이터 제공 원활 정도 점수(5점 기준)는 평균 3.4점으로 나타났다.

[그림 4-7] 타 기관의 데이터 제공 원활 정도

(n=144, 단위: %)



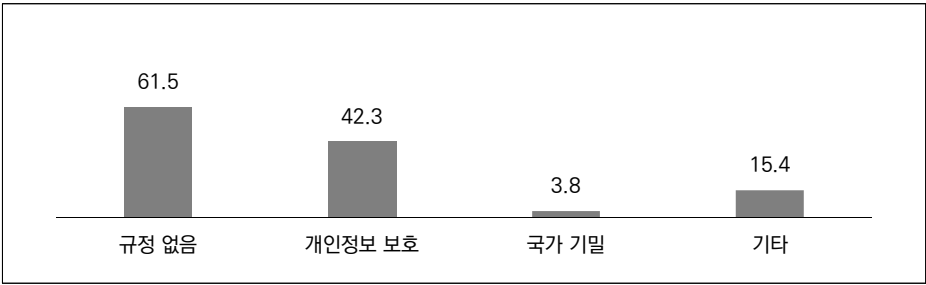


3) 외부 기관의 데이터 제공이 원활하지 않았던 이유

타 기관의 데이터 제공이 원활하지 않았다는 응답자(n=26, 복수응답)에게 그 [원활하지 않은 이유]를 묻은 결과, ‘규정 없음’(61.5%)의 비율이 가장 높게 나타났으며, 다음으로 ‘개인정보 보호’(42.3%), ‘국가 기밀’(3.8%) 등의 순으로 나타났다.

[그림 4-8] 타 기관의 데이터 제공이 원활하지 않았던 이유

(n= 26, 복수응답, 단위: %)

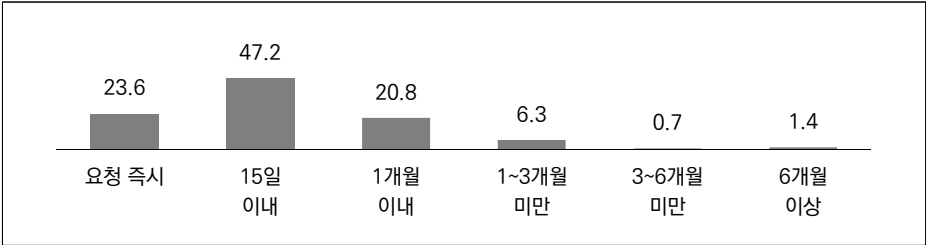


4) 외부 기관의 데이터 제공 시 소요 기간

타 기관의 데이터를 이용 해본 응답자(n=144)가 그 데이터를 제공 받았을 때, 평균적인 [소요 기간]은 ‘15일 이내’(47.2%)의 비율이 가장 높게 나타났으며, 다음으로 ‘요청 즉시’(23.6%), ‘1개월 이내’(20.8%) 등의 순으로 나타났다.

[그림 4-9] 타 기관의 데이터 제공 시 소요 기간

(n=144, 단위: %)



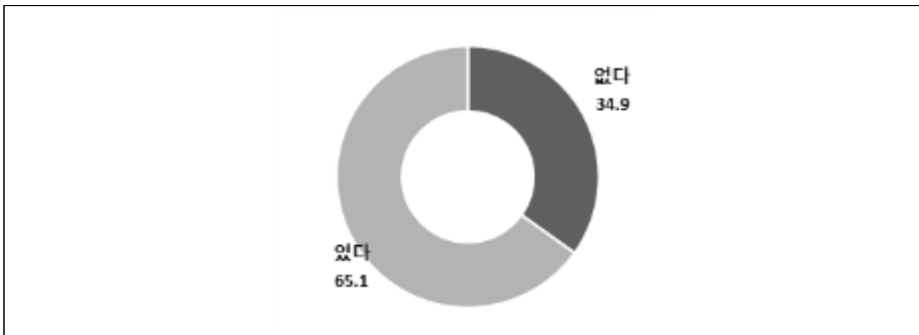
## 라. 내부/외부 기관에 데이터 제공

### 1) 내부/외부 기관에 대한 데이터 제공 경험

[내부/외부 기관에 데이터 제공 경험]을 묻은 결과, 전체 응답자 10명 중 6명 이상 (65.1%)이 데이터 제공 요청을 받은 경험이 '있다'고 응답하였다.

[그림 4-10] 내부/외부 기관에 데이터 제공 경험

(단위: %)

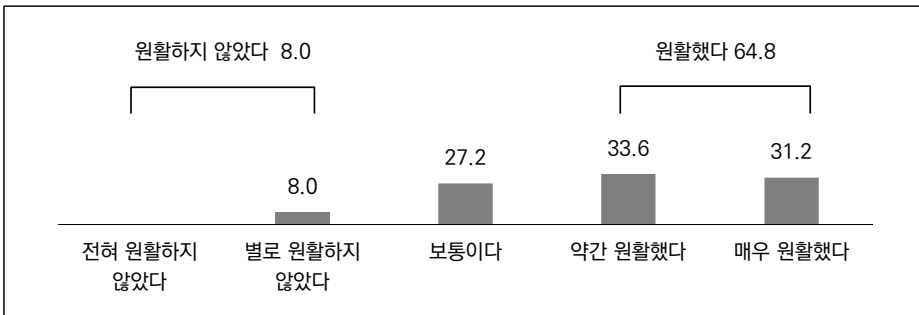


### 2) 내부/외부 기관에 대한 데이터 제공 원할 정도

내부/외부 기관에 데이터를 제공 해본 응답자(n=125)의 10명 중 6명 정도(64.8%)가 데이터 제공이 '원활했다'(약간 원활했다 + 매우 원활했다)고 응답하였고, 내부/외부 기관에 대한 데이터 제공 원할 정도 점수(5점 기준)는 평균 3.9점으로 나타났다.

[그림 4-11] 외부 기관에 대한 데이터 제공 원할 정도

(n=125, 단위: %)



3) 내부/외부 기관에 대한 데이터 제공이 원활하지 않은 이유

내부/외부 기관에 대한 데이터 제공이 원활하지 않았다는 응답자(n=10, 복수응답)에게 그 [원활하지 않은 이유]를 묻은 결과, ‘규정 없음’(70.0%)이 가장 많았으며, 다음으로 ‘개인정보 보호’(50.0%)로 나타났다.

[그림 4-12] 외부 기관에 대한 데이터 제공이 원활하지 않은 이유

(n= 10, 복수응답, 단위: %)

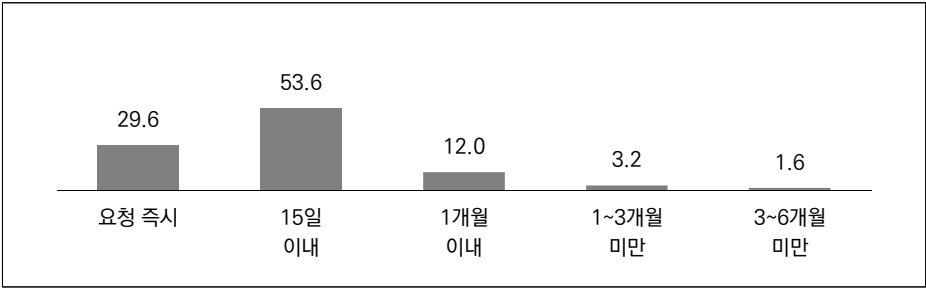


4) 내부/외부 기관에 데이터 제공 시 소요 기간

내부/외부 기관에 데이터를 제공 해본 응답자(n=125)가 그 데이터를 제공 받았을 때, [평균적으로 소요되는 기간]은 ‘15일 이내’(53.6%)가 가장 많았으며, 다음으로 ‘요청 즉시’(29.6%), ‘1개월 이내’(12.0%) 등의 순으로 나타났다.

[그림 4-13] 외부 기관에 데이터 제공 시 소요 기간

(n=125, 단위: %)

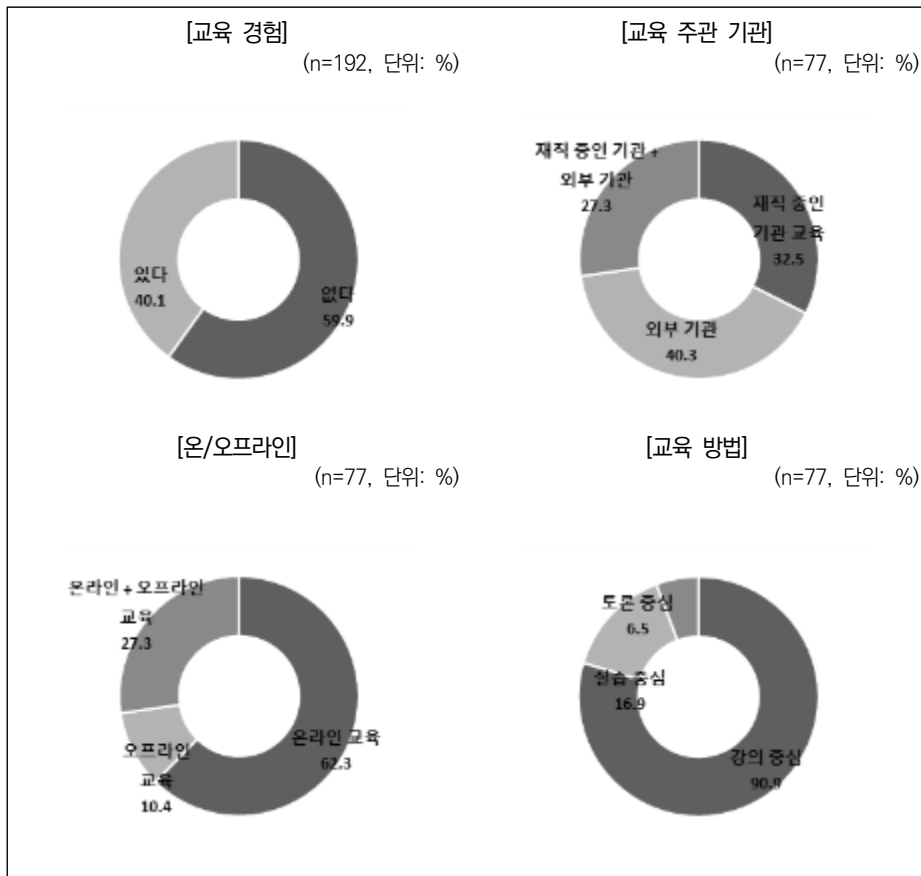


## 마. 데이터 교육 경험

### 1) 교육 경험 및 방법

최근 1년 간 [데이터 교육을 받은 경험] 여부를 묻은 결과, ‘있다’(40.1%), ‘없다’(59.9%)로 나타났다. 이 중 데이터 교육을 받은 경험이 있는 응답자(n=77)를 대상으로 [주관 기관]을 묻은 결과, ‘외부 기관’(40.3%), ‘재직 중인 기관 교육’(32.5%), ‘재직 중인 기관 + 외부 기관’(27.3%)의 비율로 나타났으며, [교육 방식 및 환경]을 묻은 결과, ‘온라인 교육’(62.3%), ‘온라인 + 오프라인 교육’(27.3%), ‘오프라인 교육’(10.4%)의 비율로 나타났다. 또한 [교육 방법]을 묻은 결과, ‘강의 중심’(90.9%), ‘실습 중심’(16.9%), ‘토론 중심’(6.5%)의 비율로 나타났다.

[그림 4-14] 데이터 교육 경험

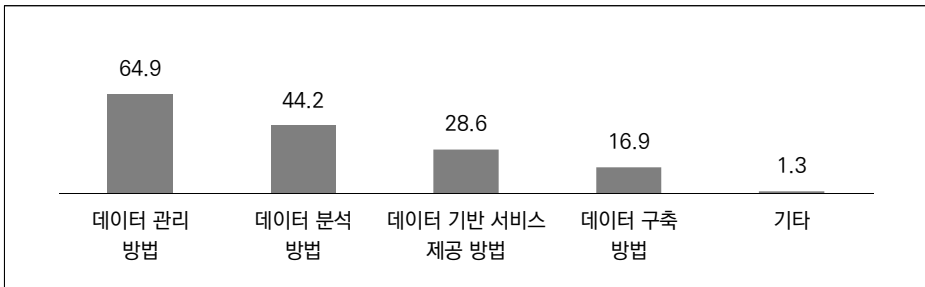


## 2) 교육 내용

최근 1년 간 [데이터 교육을 받은 경험]이 있는 응답자(n=77)를 대상으로 교육 내용을 묻은 결과, '데이터 관리 방법'(64.9%)이 가장 많았으며, 다음으로 '데이터 분석 방법'(44.2%), '데이터 기반 서비스 제공 방법'(28.6%) 등의 순으로 나타났다.

[그림 4-15] 데이터 교육 내용

(n=77, 복수응답, 단위: %)

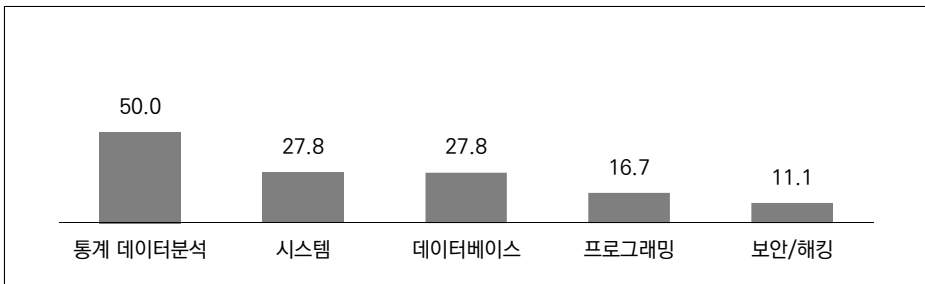


## 바. 데이터 관련 자격증

데이터 관련 [자격증 보유 여부]를 묻은 결과, '없다'(90.6), '있다'(9.4%)의 비율로 나타났다. 데이터 관련 자격증 보유한 응답자(n=18)는 평균 1.5개의 자격증을 가진 것으로 나타났으며, [자격증 항목]은 '통계 데이터분석'(50.0%), '시스템'(27.8%), '데이터베이스'(27.8%), '프로그래밍'(16.7%), '보안/해킹'(11.1%)의 비율로 나타났다.

[그림 4-16] 보유 데이터 자격증 개수 및 항목

(n=18, 복수응답, 단위: %)



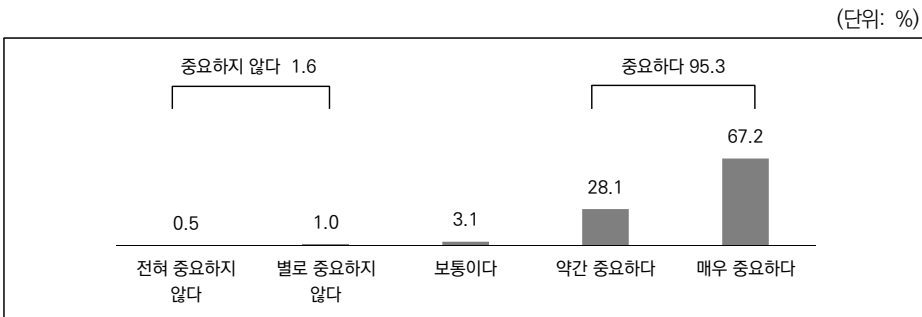
## 2.2. 데이터기반행정 인식

### 가. 데이터 활용 중요도

#### 1) 담당 업무 수행 시 데이터 활용 중요도

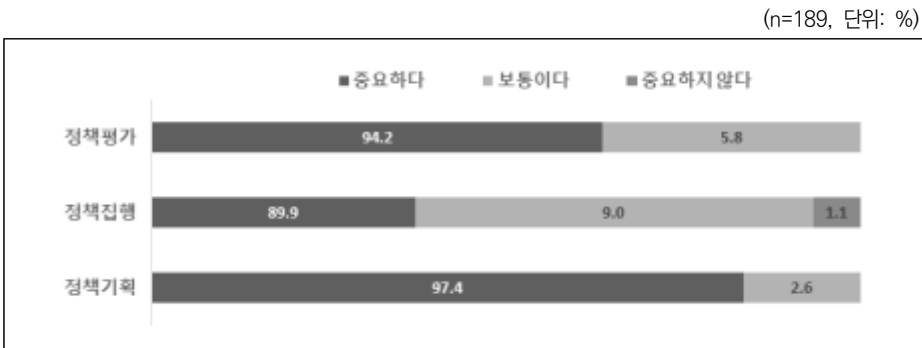
[담당 업무 수행 시 데이터 활용 중요도]를 묻은 결과, 데이터 활용이 ‘중요하다’(약간 중요하다 + 매우 중요하다)(95.3%), ‘보통이다’(3.1%), ‘중요하지 않다’(전혀 중요하지 않다 + 별로 중요하지 않다)(1.6%)의 비율로 나타났으며, 평균 4.6점으로 나타났다.

[그림 4-17] 담당 업무 수행 시 데이터 활용 중요도



[담당 업무 수행 시 데이터 활용 중요도]에 ‘중요하다’는 응답자(n=189)에게 업무 유형별 데이터 활용 중요도를 묻은 결과, ‘중요하다’(약간 중요하다 + 매우 중요하다)는 응답이 [정책기획](97.4%), [정책집행](89.9%), 그리고 [정책평가](94.2%)에서 가장 높은 비율로 나타났다.

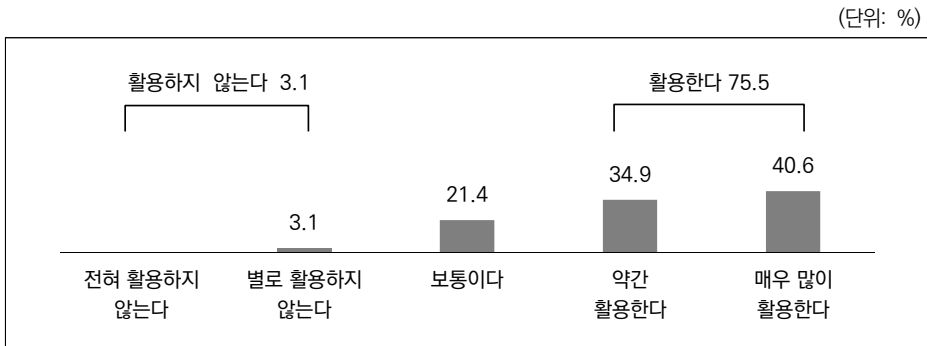
[그림 4-18] 업무 유형별 데이터 활용 중요도\_정책기획



## 2) 담당 업무 수행 시 데이터 활용 정도

[담당 업무 수행 시 데이터 활용 여부를 묻은 결과, 전체 응답자 10명 중 7명 정도 (75.5%)가 담당 업무를 수행하면서 데이터를 ‘활용한다’(약간 활용한다 + 매우 많이 활용한다)고 응답하였으며, 데이터 활용 정도 점수(5점 기준)는 평균 4.2점으로 나타났다.

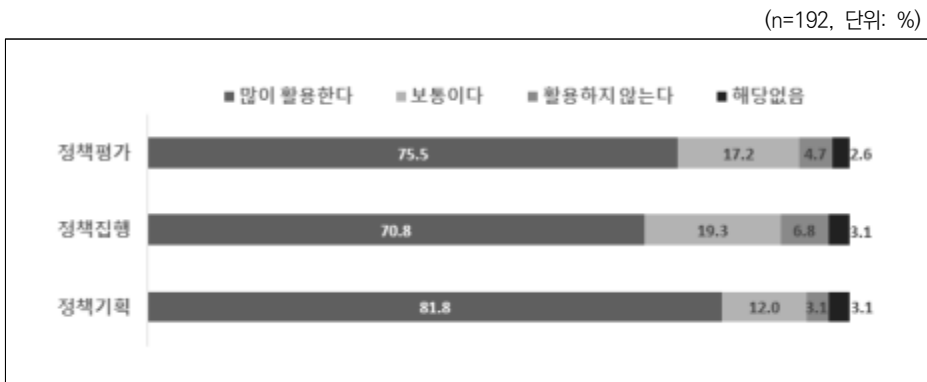
[그림 4-19] 담당업무 수행 시 데이터 활용 정도



담당 업무 수행 시 데이터 활용도에 [정책기획]에서는 ‘활용한다’(약간 활용한다 + 매우 많이 활용한다)(81.8%), [정책집행]에서는 ‘활용한다’(약간 활용한다 + 매우 많이 활용한다)(70.8%), 그리고 [정책평가]에서는 ‘활용한다’(약간 활용한다 + 매우 많이 활용한다)(75.5%)가 가장 높은 비율로 나타났다.

본 설문에 응답한 문화체육관광부 소속 및 산하기관 데이터 및 정책업무 담당자와 연구원은 담당 업무 수행 시 데이터 활용도가 높은 것을 확인할 수 있다.

[그림 4-20] 업무 유형별 데이터 활용 정도



## 나. 데이터기반행정 활성화를 위해 필요한 것

문화체육관광부 소속 및 산하기관 데이터 및 정책업무 담당자와 연구원을 대상으로 데이터기반행정 활성화를 위한 다양한 과제에 대한 중요도를 묻은 결과, ‘중요하다’(중요하다 + 매우 중요하다)는 응답이 [정책에 활용할 데이터 관련 기획]과 [데이터 관련 예산 확보]는 89.1%, [데이터 관련 전문인력 확보]는 86.5%, [데이터 관련 교육 강화]는 82.3%, [데이터 관련 장비·시설 확충]은 81.3%, [데이터 관련 부서 간 협력]은 87.0%, [조직 리더의 데이터 역량 강화]는 78.1%, [일반 조직 구성원의 데이터 역량 강화]는 81.3%, [데이터를 중시하는 조직 문화]는 85.4%로 전반적으로 높은 비율로 나타났다.

한편, 데이터기반행정 활성화를 위해 필요한 것의 중요도 평균 점수(5점 기준)는 [데이터 관련 전문인력 확보] 4.4점, [데이터 관련 부서 간 협력] 4.4점, [데이터 관련 예산 확보] 4.3점, [정책에 활용할 데이터 관련 기획] 4.3점, [데이터를 중시하는 조직 문화] 4.3점 등의 순으로 나타났다. 전반적으로 큰 차이를 보이지 않은 가운데에서도 인력, 예산, 교육에 대한 부분에서 중요도가 더 높게 나타난 것을 확인할 수 있다.

[그림 4-21] 데이터기반행정을 위한 과제들의 중요도

(n=192, 단위: %)





#### 다. 데이터기반행정 활성화 실현 정도

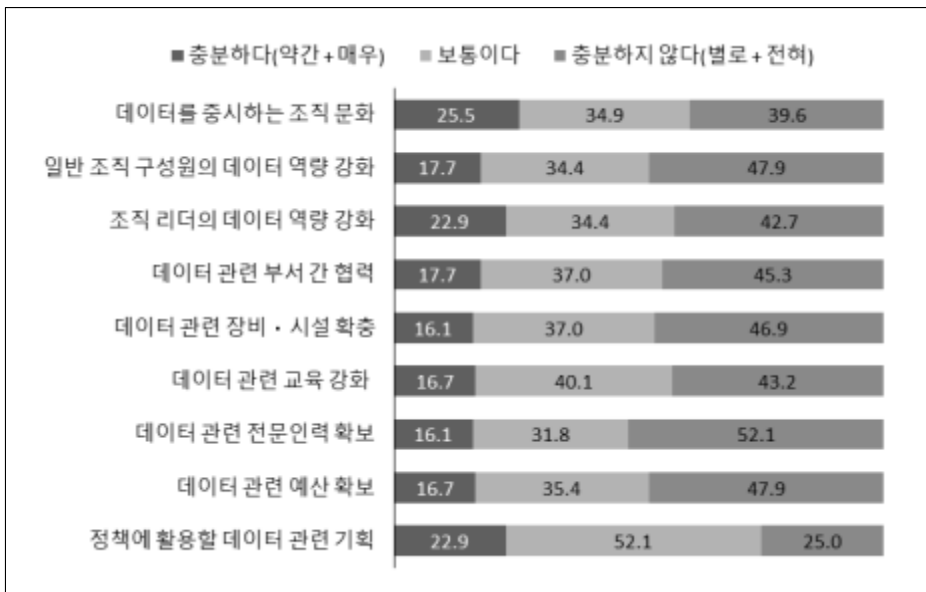
데이터기반행정 활성화를 위한 다양한 과제가 어느 정도 이뤄지고 있다고 생각하는지 묻은 결과, ‘충분하지 않다’(별로 충분하지 않다 + 전혀 충분하지 않다)는 응답이 [정책에 활용할 데이터 관련 기획]은 25.0%, [데이터 관련 예산 확보] 47.9%, [데이터 관련 전문인력 확보]는 52.1%, [데이터 관련 교육 강화] 43.2%, [데이터 관련 장비·시설 확충]은 46.9%, [데이터 관련 부서 간 협력]은 45.3%, [조직 리더의 데이터 역량 강화]는 42.7%, [일반 조직 구성원의 데이터 역량 강화]는 47.9%, [데이터를 중시하는 조직 문화]는 39.6%로 ‘충분하다’(충분하다 + 매우 충분하다)는 응답보다 높은 비율로 나타났다.

한편, 평균 점수(5점 기준)는 [정책에 활용할 데이터 관련 기획] 3.0점, [데이터를 중시하는 조직 문화] 2.8점, [조직 리더의 데이터 역량 강화] 2.7점, [데이터 관련 교육 강화] 2.7점 등의 순으로 나타났다.

전반적으로 큰 차이를 보이지 않은 가운데에서도 인력, 예산, 장비·시설에 대한 부분에서 실현 정도가 더 낮게 나타났으며, 중요도가 높게 나타났던 과제들이 실현 정도에 대한 점수가 낮은 것을 확인 할 수 있다.

[그림 4-22] 데이터기반행정을 위한 과제들의 실현 정도

(n=192, 단위: %)



## 2.3. 데이터 공동활용 실태

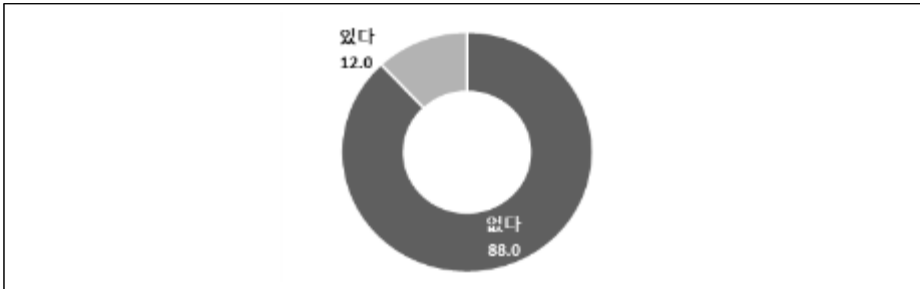
### 가. 데이터 공동활용 경험

#### 1) 데이터 공동활용 정책/사업 참여 경험

문화체육관광부 소속 및 산하기관 데이터 및 정책업무 담당자와 연구원(n=192)을 대상으로 [데이터 공동활용 정책 및 사업 참여 경험을 묻은 결과, 전체 응답자 10명 중 8명 이상(88.0%)이 ‘없다’고 응답하였다.

[그림 4-23] 데이터 공동활용 정책 및 사업 참여 경험

(단위: %)



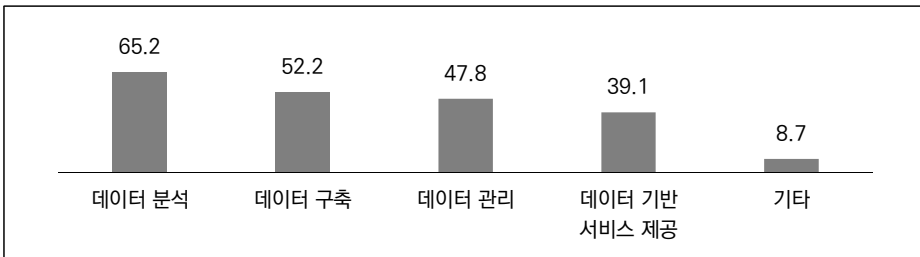
#### 2) 데이터 공동활용 사업 관련 정보

##### ① 사업목적

데이터 공동활용 정책 및 사업 참여 경험이 있는 응답자(n=23, 복수응답)의 [사업목적]은 ‘데이터 분석’(65.2%)이 가장 많았으며, 다음으로 ‘데이터 구축’(52.2%), ‘데이터 관리’(47.8%), ‘데이터 기반 서비스 제공’(39.1%) 등의 순으로 나타났다.

[그림 4-24] 사업목적

(n=23, 복수응답, 단위: %)



② 사업 기간, 사업 예산, 투입 인력

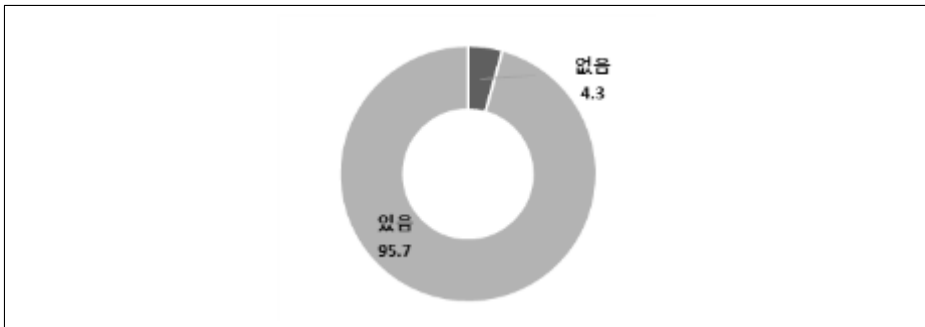
데이터 공동활용 정책 및 사업 참여 경험이 있는 응답자(n=23)의 사업 기간은 평균 12.5개월이며, 예산은 평균 59,501.7만 원, 평균 투입 인력은 12.5명으로 나타났다.

③ 공공/민간 데이터 공동활용 경험

데이터 공동활용 정책 및 사업 참여 경험이 있는 응답자(n=23, 복수응답)의 [공공 데이터 공동활용 경험]은 '있다' 95.7%, '없다' 4.3%의 비율로 나타났다.

[그림 4-25] 공공 데이터 공동활용 경험

(n=23, 단위: %)

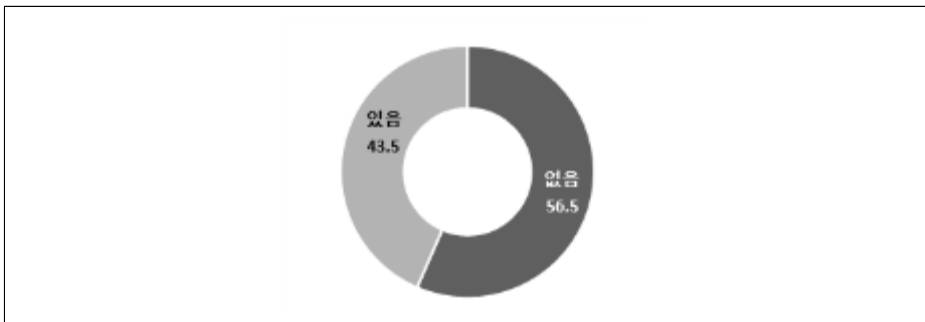


④ 민간 데이터 공동활용 경험

데이터 공동활용 정책 및 사업 참여 경험이 있는 응답자(n=23)의 절반 이상(56.5%)이 [민간 데이터 공동활용 경험]이 '없다'고 응답했으며, '있다'는 43.5%로 나타났다.

[그림 4-26] 민간 데이터 공동활용 경험

(n=23, 단위: %)



⑤ 공동활용 데이터(공공)

데이터 공동활용 정책 및 사업 참여 경험이 있는 응답자(n=23, 복수응답)를 대상으로 공공데이터 공동활용 경험 및 데이터 보유 주체명을 묻은 결과, ‘통계청’(6건)과 ‘문화체육관광부’(5건) 등으로 나타났다.

〈표 4-33〉 공공 데이터 보유 주체

(n=22, 복수응답, 단위: 건)

사업명	사례수
전 체	(37)
통계청	(6)
문화체육관광부	(5)
한국문화예술위원회	(3)
국립중앙도서관	(2)
한국문학번역원	(2)
고용노동부	(1)
공공데이터포털	(1)
관련 협회	(1)
국립중앙박물관	(1)
국립현대미술관	(1)
국세청	(1)
근로복지공단	(1)
대한장애인체육회	(1)
문화예술위원회	(1)
서울특별시	(1)
예술경영지원센터	(1)
재단법인 국악방송	(1)
한국관광공사	(1)
한국문화정보원	(1)
한국산업단지공단	(1)
한국연구재단	(1)
한국은행	(1)
한국청소년진흥원	(1)
한국출판문화산업진흥원	(1)

#### ⑥ 공동활용 데이터(민간)

민간 데이터 공동활용 경험이 있는 응답자(n=10, 복수응답)를 대상으로 데이터 보유 주체명을 물은 결과, ‘ 신한카드’(5건), ‘SKT’(3건), ‘KT’(2건) 등이 조사되었다.

〈표 4-34〉 데이터 공동활용 사업명

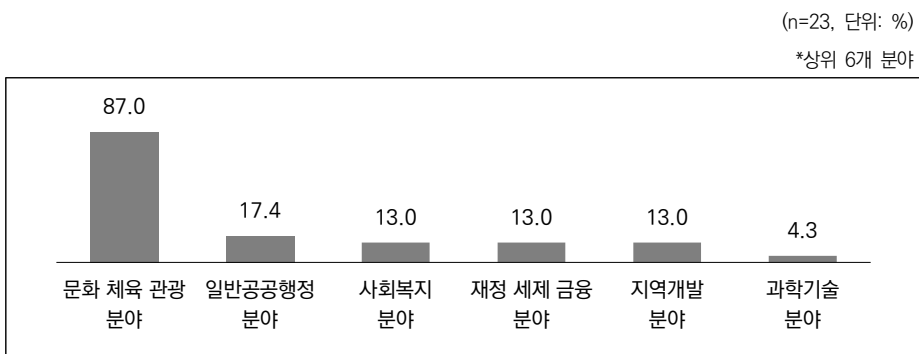
(n=10, 복수응답, 단위: 건)

사업명	사례수
전 체	(18)
신한카드	(5)
SKT	(3)
KT	(2)
SK	(1)
VISA	(1)
하나카드	(1)
레드테이블	(1)
공연단체	(1)
대한장애인체육회	(1)
대한체육회	(1)
서울대학교	(1)

#### ⑦ 공동활용 정책분야

데이터 공동활용 정책 및 사업 참여 경험이 있는 응답자(n=23, 복수응답)의 [공동활용 정책분야]는 ‘문화 체육 관광’(87.0%)의 비율이 가장 높게 나타났으며, 다음으로 ‘일반공공행정’(17.4%), ‘교육’(13.0%) 등의 순으로 나타났다.

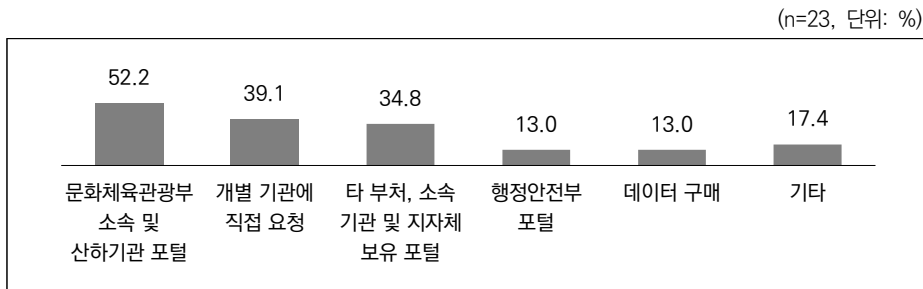
[그림 4-27] 공동활용 정책분야



#### ⑧ 데이터 확보 채널

데이터 공동활용 정책 및 사업 참여 경험이 있는 응답자(n=23, 복수응답)의 [데이터 확보 채널]은 '문화체육관광부 소속 및 산하기관 포털'(52.2%)의 비율이 가장 높게 나타났으며, 다음으로 '개별 기관에 직접 요청'(39.1%), '타 기관에서 보유하고 있는 포털(홈페이지 포함)'(34.8%) 등의 순으로 나타났다.

[그림 4-28] 데이터 확보 채널

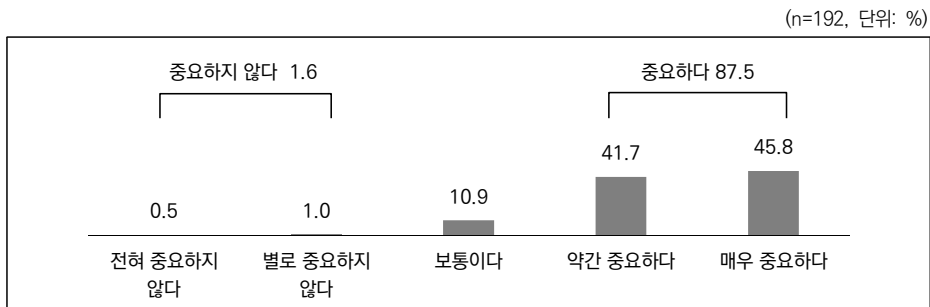


### 1.4. 데이터 공동활용 인식

#### 가. 데이터 공동활용 중요도

데이터기반행정 강화를 위한 데이터 공동활용 중요도를 묻은 결과, 전체 응답자 10명 중 8명 이상(87.5%), 데이터기반행정 강화를 위해 데이터 공동활용이 '중요하다'(약간 중요하다 + 매우 중요하다)고 응답하였고, 데이터 공동활용 중요도 점수(5점 기준)는 평균 4.4점으로 나타났다.

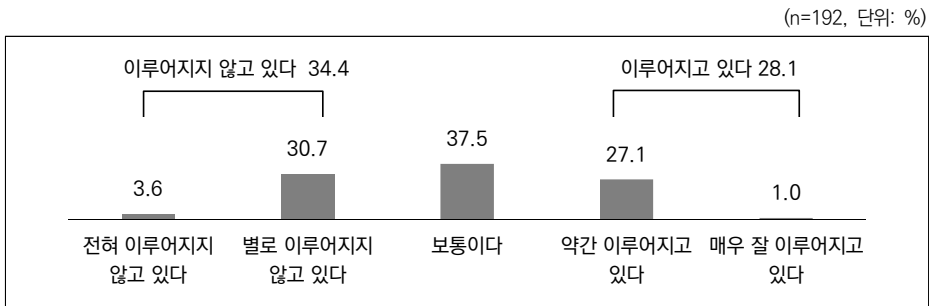
[그림 4-29] 데이터기반행정 강화를 위한 데이터 공동활용 중요도



## 나. 데이터 공동활용 실현 정도

데이터기반행정 강화를 위한 기관 간 데이터 공동활용 실현 정도를 묻은 결과, 전체 응답자 10명 중 3명 정도(34.4%)가 ‘이루어지지 않고 있다’(전혀 이루어지지 않고 있다 + 별로 이루어지지 않고 있다)고 응답하였고, 데이터 공동활용 실현 정도 점수(5점 기준)는 평균 2.9점으로 나타났다.

[그림 4-30] 데이터기반행정 강화를 위한 데이터 공동활용 실현 정도



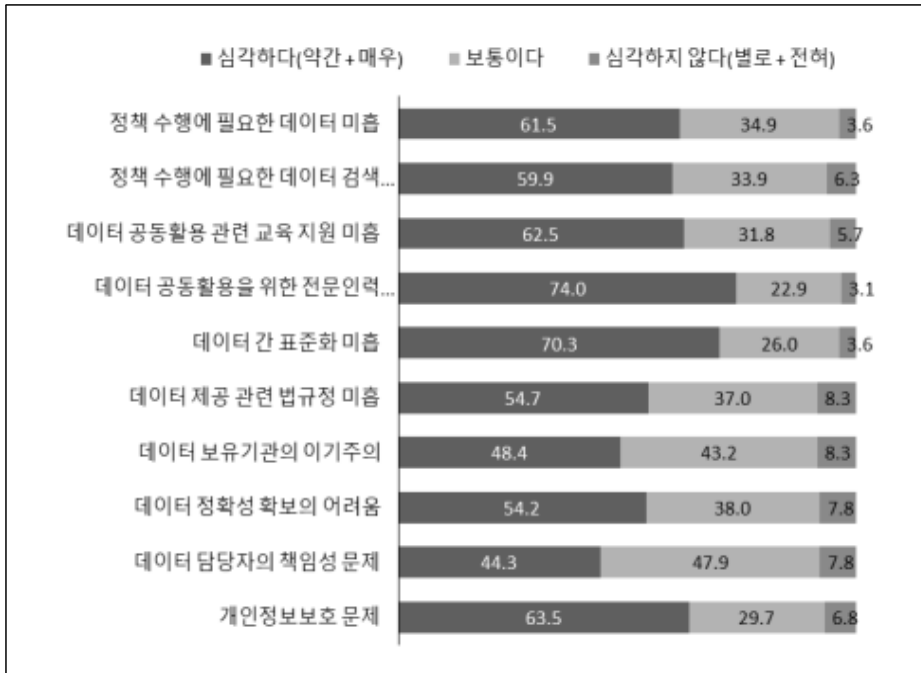
## 다. 데이터 공동활용 제약조건

문화체육관광부 소속 및 산하기관 데이터 및 정책업무 담당자와 연구원(n=192)을 대상으로 데이터기반행정 강화를 위한 데이터 공동활용에 있어 제약조건을 묻은 결과, ‘심각하다’(약간 심각하다 + 매우 심각하다)는 응답이 [개인정보보호 문제]는 63.5%, [데이터 담당자의 책임성 문제]는 44.3%, [데이터 정확성 확보의 어려움]은 54.2%, [데이터 보유기관의 이기주의]는 48.4%, [데이터 제공 관련 법규정 미흡]은 54.7%, [데이터 간 표준화 미흡]은 70.3%, [데이터 공동활용을 위한 전문인력 부족]은 74.0%, [데이터 공동활용 관련 교육 지원 미흡]은 62.5%, [정책 수행에 필요한 데이터 검색 어려움]은 59.9%, [정책 수행에 필요한 데이터 미흡]은 61.5%의 비율로 나타났다.

한편, 데이터 공동활용 제약조건 심각성 점수는 [데이터 공동활용을 위한 전문인력 부족]은 4.0점, [데이터 간 표준화 미흡]은 4.0점, [데이터 공동활용 관련 교육 지원 미흡]은 3.8점, [정책 수행에 필요한 데이터 미흡]은 3.8점 등의 순으로 나타났다.

[그림 4-31] 데이터 공동활용 제약조건의 심각성

(n=192, 단위: %)



## 라. 데이터 공동활용 활성화를 위해 필요한 부분

문화체육관광부 소속 및 산하기관 데이터 및 정책업무 담당자에게 데이터 공동활용 활성화를 위해 필요한 부분을 물은 결과, ‘데이터 전문인력 확보 및 교육’(36건)이 필요하다는 의견이 가장 많았으며, 다음으로 ‘법이나 제도마련 등 체계 강화’(22건), ‘기관 간 협력 및 협조를 통한 협업’(16건), ‘데이터 접근성을 높이고 공개범위 확장’(14건) 등의 순으로 나타났다.

〈표 4-35〉 데이터 공동활용 활성화를 위해 필요한 부분

(n=192, 단위: 건)

사업명	사례수
전 체	(133)
데이터 전문인력 확보 및 교육	(36)
법/제도 등 기반 마련	(22)
기관 간 협력 및 협조	(16)



사업명	사례수
데이터 개방 노력	(14)
데이터 플랫폼 및 통합 시스템	(11)
데이터 공동활용에 대한 홍보 및 인식제고	(10)
데이터 표준화 또는 가이드라인	(10)
양질의 데이터 생산	(7)
개인정보관리에 대한 고민	(3)
전반적인 데이터 내실화 및 수요 반영	(2)
데이터 관련 기초연구 및 분석방법연구	(2)
의견없음	(59)

## 제5절 조사결과 종합 및 활용

### 1. 조사결과 종합

#### 1.1. 공공데이터 현황조사 요약

먼저, 문화체육관광부 소속 및 산하기관이 공공데이터 포털에 개방한 데이터를 살펴 보면, 본부 및 소속기관 47곳에서 총 1,048개의 공공데이터를 개방하고 있었는데 그 중 파일데이터는 833개(79.5%), 오픈API는 215개(20.5%)를 차지하는 것으로 나타났다. 문화체육관광부를 포함한 소속기관의 경우는 15곳에서 총 161개의 공공데이터를 개방하고 있었는데 그 중 파일데이터는 112개(69.6%), 오픈API는 49개(30.4%)를 차지했다. 산하기관은 33곳에서 총 887개의 공공데이터를 개방하고 있었는데 그 중 파일데이터는 721개(81.3%), 오픈API는 166개(18.7%)를 차지한 것으로 조사됐으며, 한국문화예술위원회(159개), 한국언론진흥재단(117개), 한국저작권위원회(109개) 순으로 공공데이터를 많이 개방하고 있는 것으로 나타났다.

문화체육관광 분야 공공데이터 포털 개방 현황을 살펴보면, 총 26,979개의 공공데이터 중 문화재 관련 공공데이터가 10,584개(39%)로 가장 많았고, 다음으로 관광 관련 공공데이터가 5,880개(22%), 문화예술 관련 공공데이터가 5,099개(19%) 등의 순으로 나타났다. 문체부 소속 및 산하기관이 공공데이터포털에 개방한 데이터와 행정안전부 공공데이터포털에서 개방하고 있는 문화체육관광 분야 전체 데이터를 비교해 보면, 문체부 소속 및 산하기관에서 제공하고 있는 데이터의 비중은 문화체육관광 분야 전체 데이터의 3.2% 수준으로 매우 낮은 비중을 보이는 것으로 조사됐다.

다음으로 문체부 산하기관인 한국문화정보원의 문화정보통합시스템 내 데이터 수집·관리 현황을 살펴본 결과, 정형 1,083개, 비정형 74개, 오픈API 487개 등 총 1,651개의 데이터를 수집·관리하고 있는 것으로 나타났다.

문화정보통합시스템 내 데이터를 세부적으로 살펴본 결과, 총 1,651개의 공공데이터 중 데이터 주제별 구분은 공지 607개, 행사 157개를 제외하고 유물(소장품) 83개, 기타창작물 57개, 교육 49개, 장소 43개, 도서 40개 등의 순으로 나타났다. 그리고 데이터 형태별 구분은 공지/통계 619개, 분류안됨 496개, 행사 157개를 제외하고 창작물 194개, 유형 유산 86개, 장소 43개, 개인/단체 42개, 무형 자산 14개의 순으로 나타났다. 문화정보통합시스템에서 수집·관리되고 있는 주제별 데이터 정보 중 69% 이상이 공지, 행사, 교육 등의 일반 행정데이터로 실제 정책분야에서 활용 가능한 데이터는 매우 낮은 수준에서 개방이 이뤄지고 있는 것으로 나타났다.

## 1.2. 보유 및 수요 데이터 전수조사 결과 요약

문화체육관광부 소속 및 산하기관별로 보유 또는 운영·관리하고 있으나, 외부에 개방(공개)하지 않은 보유 데이터 조사 결과를 살펴보면, 소속 및 산하기관 21곳에서 총 452개의 데이터를 보유 또는 운영·관리하고 있었으며, 데이터 생성·취득 방법별로는 2차가공 208개(46.0%), 직접조사 147개(32.5%), 행정자료 97개(21.5%) 순으로 나타났다.

정부기능분류 기준 정책분야(문화체육관광)와 정책영역(4개영역)으로 구분한 보유 데이터 조사 결과를 살펴보면 총 452개의 보유 데이터 중 문화예술 관련 데이터가 345개(76.3%)로 가장 많았고, 다음으로 관광 관련 데이터가 45개(10.0%), 체육 40개(8.8%) 순으로 조사됐다. 이어서 앞서 분석된 정책영역을 행정업무 단위로 세분화한 정부기능별(대기능) 보유 데이터 현황을 공동활용 정책분야별로 구분하여 조사결과를 살펴보았다. 전체 문화체육관광 분야 데이터 452개 중 문화산업정책 데이터가 154개로 가장 많은 것으로 나타난 가운데, 문화산업정책 데이터의 49% 이상이 “특화된 대책 및 맞춤형 서비스” 관련 정책분야에서 공동활용이 가능한 데이터로 조사되었다. 문화산업정책을 다른 업무 기능분류와 비교해서 살펴볼 경우에도 5가지 공동활용 정책분야 모두에서 문화산업정책 관련 업무 데이터의 비중이 가장 높은 것으로 분석됐다.

문화체육관광부 소속 및 산하기관별로 업무 수행 시 필요한 타 기관 데이터에 대한 수요조사를 보유 데이터 조사와 함께 실시하였다. 8개 기관에서 총 49개 데이터를 업무 수행 시 필요한 타 기관 데이터로 응답하였으며, 공동활용 정책분야별로는 “특화된 대책 및 맞춤형 서비스(32개)” 분야에 활용할 수 있는 데이터에 대한 수요가 가장 많은 것으로 나타났다.

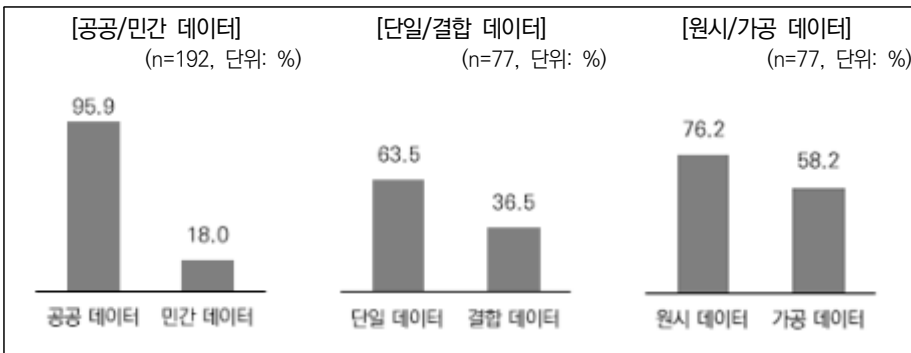
### 1.3. 데이터 업무 활용 실태조사 결과 요약

문화체육관광부 소속 및 산하기관 데이터 및 정책업무 담당자 대상 데이터 업무 활용 실태조사를 실시한 결과, 주로 활용하는 데이터 유형은 ‘통계조사 자료’(78.1%), 데이터 형식은 ‘정형 데이터’(46.4%), 데이터 구성은 ‘단일 데이터’(63.5%)가 높게 나타났다.

단일 데이터 이용 경험이 있는 경우는 ‘공공 데이터’(95.9%), 결합 데이터 유형은 ‘공공 데이터 + 공공 데이터’(51.4%), 제공기관은 ‘문화체육관광부 소속 및 산하 기관’(61.9%), 정책분야는 ‘문화 체육 관광 분야’(90.3%)의 데이터를 주로 활용하였다.

타 기관의 공공 데이터 활용 경험이 있는 경우 ‘문화체육관광부 소속 및 산하기관 포털’(70.8%)의 데이터를 주로 이용하였는데, 데이터 제공의 원활도는 ‘보통이다’(43.8%)라고 하였으며, 그 이유로는 ‘규정 없음’(61.5%)이 가장 높게 나타났다.

[그림 4-32] 데이터 교육 경험



반면에 내부/외부 기관에 데이터 제공한 경험(65.1%)이 있는 경우 데이터 제공이 원활도는 ‘원활했다’(64.8%), 원활하지 않은 경우 그 이유는 ‘규정 없음’(70.0%)이 가장 많았다.

최근 1년 간 데이터 교육을 받은 경험은 5명 중 2명이 ‘있음’(40.1%)이며, 데이터 관련 자격증 보유한 비율은 9.4%이며, 자격증은 평균 1.5개로 나타났다.

담당 업무 수행 시 데이터 활용 중요도는 ‘중요하다’(중요하다 + 매우 중요하다)(95.3%)는 응답이 높게 나타났으며, [정책기획](97.4%), [정책집행](89.9%), [정책평가](94.2%) 업무에도 역시 ‘중요하다’(약간 중요하다 + 매우 중요하다)가 높게 나타났다.

담당 업무 수행 시 데이터 활용도는 ‘활용한다’(약간 활용한다 + 매우 많이 활용한다)(75.5%)[정책기획](81.8%), [정책집행](70.8%), [정책평가](75.5%) 업무에서도 역시 ‘활용한다’(약간 활용한다 + 매우 많이 활용한다)가 높게 나타났다.

데이터 공동활용 정책 및 사업 참여 경험자는 12.0%로 주 업무는 ‘데이터 분석’ (65.2%) 이 가장 많았으며, 사업 기간은 평균 12.5개월이며, 예산은 평균 59,501.7만 원, 평균 투입 인력은 12.5명으로 나타났다.

〈표 4-36〉 담당 업무 수행 시 데이터 활용 중요도 및 활용도

(단위: %)

구분	중요도				활용도				
	중요하다 (약간+매우)	보통 이다	중요하지 않다 (별로+전혀)	평균 점수 (5점 기준)	활용한다 (약간+매우)	보통 이다	활용하지 않는다 (별로+전혀)	해당 없음	평균 점수 (5점 기준)
[전체]	95.3	3.1	1.6	4.6	75.5	21.4	3.1	-	4.1
정책기획	97.4	2.6	-	4.8	81.8	12.0	3.1	3.1	4.3
정책집행	89.9	9.0	1.1	4.5	70.8	19.3	6.8	3.1	4.0
정책평가	94.2	5.8	-	4.7	75.5	17.2	4.7	2.6	4.1

데이터기반행정 활성화를 위해 필요한 과제를 묻은 결과, ‘중요하다’(중요하다 + 매우 중요하다)는 응답이 [데이터 관련 예산 확보](89.1%), [데이터 관련 부서 간 협력](87.0%), [데이터 관련 전문인력 확보](86.5%) 등의 순으로 높은 비율로 나타났다.

〈표 4-37〉 데이터기반행정 활성화를 위해 필요한 과제

(단위: %)

구분	중요도				실현 정도			
	중요하다 (약간+매우)	보통 이다	중요하지 않다 (별로+전혀)	평균 점수 (5점 기준)	충분하다 (약간+매우)	보통 이다	충분하지 않다 (별로+전혀)	평균 점수 (5점 기준)
정책에 활용할 데이터 관련 기획	89.1	8.9	2.1	4.3	22.9	52.1	25.0	3.0
데이터 관련 예산 확보	89.1	9.9	1.0	4.3	16.7	35.4	47.9	2.6
데이터 관련 전문인력 확보	86.5	12.5	1.0	4.4	16.1	31.8	52.1	2.5
데이터 관련 교육 강화	82.3	17.2	0.5	4.2	16.7	40.1	43.2	2.7

구분	중요도				실현 정도			
	중요하다 (약간+매우)	보통 이다	중요하지 않다 (별로+전혀)	평균 점수 (5점 기준)	충분하다 (약간+매우)	보통 이다	충분하지 않다 (별로+전혀)	평균 점수 (5점 기준)
데이터 관련 장비·시설 확충	81.3	16.7	2.1	4.2	16.1	37.0	46.9	2.6
데이터 관련 부서 간 협력	87.0	12.0	1.0	4.4	17.7	37.0	45.3	2.6
조직 리더의 데이터 역량 강화	78.1	19.3	2.6	4.1	22.9	34.4	42.7	2.7
일반 조직 구성원의 데이터 역량 강화	81.3	17.7	1.0	4.1	17.7	34.4	47.9	2.6
데이터를 중시하는 조직 문화	85.4	13.5	1.0	4.3	25.5	34.9	39.6	2.8

데이터기반행정 활성화 실현 정도는 ‘충분하지 않다’(별로 충분하지 않다 + 전혀 충분하지 않다)는 응답이 [데이터 관련 전문인력 확보](52.1%), [데이터 관련 예산 확보](47.9%), [데이터 관련 장비·시설 확충](46.9%) 등의 순으로 높게 나타났다.

데이터 공동활용 중요도는 데이터 공동활용이 ‘중요하다’(약간 중요하다 + 매우 중요하다)(87.5%)는 응답이 높게 나타난 가운데, 데이터 공동활용 실현도는 ‘이루어지지 않고 있다’(전혀 이루어지지 않고 있다 + 별로 이루어지지 않고 있다)(34.4%)가 높게 나타났다. 또한 데이터기반행정 강화를 위한 데이터 공동활용에 있어 제약조건을 묻은 결과, ‘심각하다’(약간 심각하다 + 매우 심각하다)는 응답이 [데이터 공동활용을 위한 전문인력 부족](74.0%), [데이터 간 표준화 미흡](70.3%), [개인정보보호 문제](63.5%) 등의 순으로 나타났다.

〈표 4-38〉 데이터기반행정 활성화를 위해 필요한 과제

(단위: %)

구분	중요도			
	중요하다 (약간+매우)	보통이다	중요하지 않다 (별로+전혀)	평균점수 (5점기준)
개인정보보호 문제	89.1	8.9	2.1	4.3
데이터 담당자의 책임성 문제	89.1	9.9	1.0	4.3
데이터 정확성 확보의 어려움	86.5	12.5	1.0	4.4
데이터 보유기관의 이기주의	82.3	17.2	0.5	4.2
데이터 제공 관련 법규정 미흡	81.3	16.7	2.1	4.2

구분	중요도			
	중요하다 (약간+매우)	보통이다	중요하지 않다 (별로+전혀)	평균점수 (5점기준)
데이터 간 표준화 미흡	87.0	12.0	1.0	4.4
데이터 공동활용을 위한 전문인력 부족	78.1	19.3	2.6	4.1
데이터 공동활용 관련 교육 지원 미흡	81.3	17.7	1.0	4.1
정책 수행에 필요한 데이터 검색 어려움	85.4	13.5	1.0	4.3
정책 수행에 필요한 데이터 미흡	85.4	13.5	1.0	4.3

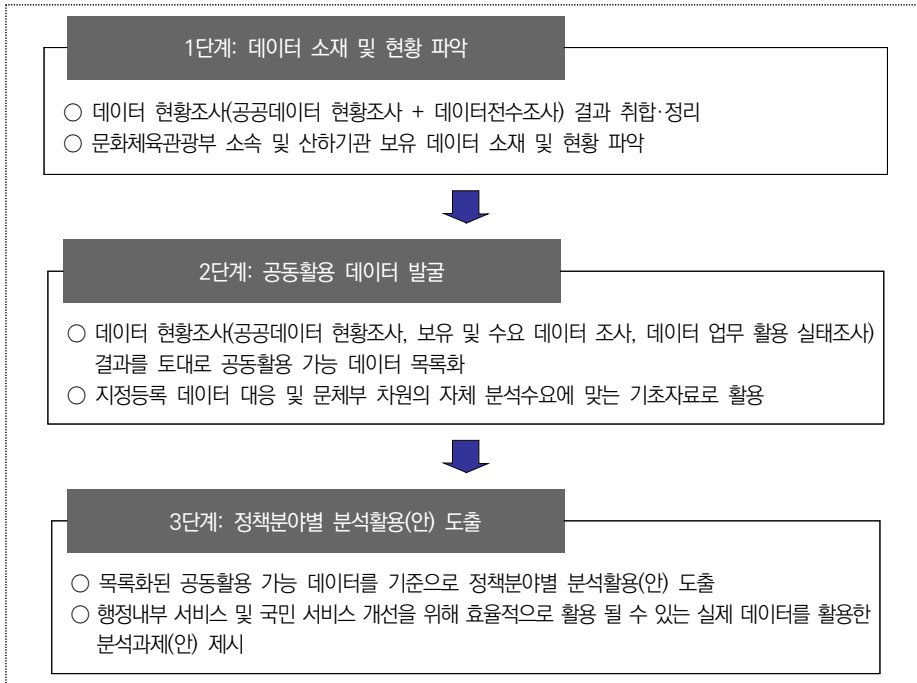
전반적으로 문화체육관광부 소속 및 산하기관 데이터 및 정책업무 담당자들은 담당 업무를 수행할 때 데이터를 많이 활용하고 있으며, 데이터기반행정 활성화를 위해서는 인력, 예산, 장비시설의 확충이 중요하다고 생각하고 있다. 그러나 데이터기반행정 실현 정도는 중요도가 높았던 항목들을 충분하지 않다고 생각하고 있음을 알 수 있다.

또한 데이터기반행정 강화를 위한 데이터 공동활용의 중요성은 높게 생각했으나 실현 정도는 잘 이뤄지지 않고 있다고 보았다. 그리고 공동활용의 제약조건으로는 ‘전문인력이 부족’하고, ‘데이터의 표준화가 미흡’하며, ‘개인정보보호의 문제’가 있음을 알 수 있다. 한편, 데이터 공동활용 활성화를 위해 필요한 부분으로는 ‘데이터 전문인력 확보 및 교육’(36건)이 필요하다는 의견이 가장 많았으며, 다음으로 ‘법이나 제도마련 등 체계 강화’(22건), ‘기관 간 협력 및 협조를 통한 협업’(16건), ‘데이터 접근성을 높이고 공개범위 확장’(14건) 등의 순으로 나타났다.

## 2. 조사결과 활용

본 연구에서 수행한 데이터 현황조사(공공데이터 현황조사, 보유 데이터 전수조사, 수요데이터 전수조사, 데이터 업무활용 실태조사)의 궁극적인 수행 목적은 문체부 소속 및 산하기관이 보유 또는 필요로 하는 데이터의 소재 및 현황을 파악해 보고 나아가서는 조사결과를 토대로 문화체육관광부 차원의 공동활용 데이터 발굴, 기획과제 및 분석과제 발굴 등에 필요한 기초자료로 활용하기 위함이다. 실제 데이터 현황조사 조사결과는 데이터 소재 및 현황 파악, 공동활용 데이터 발굴, 정책분야별 분석활용(안) 도출 이상의 3가지 분야에서 활용될 수 있으며, 단계별 조사활용 체계는 아래 <표 4-39>와 같다

〈표 4-39〉 조사결과 활용 체계



#### 가. 문화체육관광부 소속 및 산하기관 보유 데이터 소재 및 현황 파악

4장에서 수행한 공공데이터 현황조사 결과와 보유 데이터 전수조사 결과를 동일한 수준의 영역으로 취합·정리하면, 문화체육관광부 소속 및 산하기관이 유 또는 생성·취득하여 운영·관리하고 있는 데이터 총괄 현황을 파악할 수 있는 결과로 활용 할 수 있다. 데이터 총괄 현황을 크게 ‘정부기능분류 기준’과 ‘문화체육관광부 조직분류 기준’으로 구분하여 살펴볼 수 있으며, 데이터기반행정법에서 제시하고 있는 5가지의 정책영역을 각각의 분류기준별 교차하여 문제부 소속 및 산하기관이 보유하고 있는 데이터 소재 및 현황을 파악할 수 있다.

#### 나. 공동활용 데이터 발굴

본 연구에서 실시한 데이터 현황조사(공공데이터 현황조사, 보유 및 수요 데이터 조사, 데이터 업무 활용 실태조사) 결과를 토대로 데이터기반행정법에 명시된 정책분야



에 부합하는 공동활용 데이터 목록화 작업을 진행하였다. 일반적인 행정(보도자료, 취업현황 등) 데이터를 제외한 고유 사업 추진(운영) 및 정책수립, 집행과정 모니터링, 정책평가 등 과정에서 관리 중인 데이터, 또는 행정안전부가 권고하고 있는 공동활용 정책 분야(데이터기반행정부 제8조제1항)에 포함되는 데이터로서 기관 간 공동 활용성이 높다고 판단되는 보유 데이터를 목록화 작업 대상으로 설정하였다. 본 연구의 데이터 전수조사를 통해 수집·정리됐으나, 외부 개방 시스템(포털) 등에서 단 한 번도 공개된 적이 없는 비공개 데이터(103개)에 대해서는 기관별 상황(데이터 신뢰성 및 원데이터 요청에 대한 우려, 기관 간 비교 등)을 고려하여 공동활용 가능 데이터 목록화 대상에서 제외하였다. 본 연구에서 목록화한 공동활용 가능 데이터를 토대로 범정부 차원의 데이터기반 행정 활성화를 위한 지정등록 데이터 대응 및 문체부 차원의 자체 분석수요에 맞는 기초 자료로 활용할 수 있을 것으로 보인다.

#### 다. 정책분야별 분석활용(안) 도출

앞서 목록화된 공동활용 가능 데이터를 기준으로 정책분야별 분석활용 과제(안)을 도출하였다. 행정안전부가 권고하고 있는 공동활용 정책분야별로 활용될 수 있는 문화체육관광 분야 데이터를 기준으로 타 부처 및 민간 데이터와 공동활용할 수 있는 대상 데이터를 설정하였고 이를 토대로 행정내부 서비스 및 국민 서비스 개선을 위해 효율적으로 활용 될 수 있는 정책분야별 분석활용(안)을 도출하고 실제 데이터를 활용한 분석과제(안)을 제시하였다.

### 2.1. 데이터 소재 및 현황 파악

#### 가. 정부기능분류 기준 데이터 현황 파악

문화체육관광부 소속 및 산하기관이 보유 또는 생성·취득하여 운영·관리하고 있는 데이터 총괄 현황(공공데이터 현황조사+데이터 전수조사)을 살펴보면, 총 1,364개의 데이터를 보유 또는 운영·관리하고 있었는데 그 중 이미 공공데이터포털에서 개방되어 있는 공공데이터(기 개방된 데이터)는 1,048개로 나타났으며, 보유 데이터 전수조사를 통해 파악된 전수데이터(신규 데이터)는 452개로 집계되었다.

공공데이터 및 전수데이터를 정부기능분류(정책영역)에 따라 조사된 현황을 살펴보면, 두 데이터 모두 문화예술영역의 데이터가 전체 데이터 중에 차지하는 비중이 70% 이상으로 가장 많은 것으로 나타난 가운데, 관광영역 데이터는 공공데이터(9.7%), 전수데이터(10.0%) 수준으로 집계되어 전체 데이터 대비 관광영역의 데이터 분포가 거의 유사한 것으로 분석됐다.

본 연구를 통해 파악된 문체부 소속 및 산하기관이 보유하고 있는 데이터 총 1,364개 중 문화예술영역 데이터가 954개(69.9%)로 가장 많았고, 다음으로 체육 170개(12.5%), 관광 147개(10.8%), 문화체육관광일반 93개(6.8%) 순으로 나타났다.

〈표 4-40〉 문체부 소속·산하기관 보유 데이터 현황(전수+공공)

(단위: 개, %)

구분	문화예술	체육	관광	문화체육관광 일반	합계
공공데이터*	742(70.8)	130(12.4)	102(9.7)	74(7.1)	1,048(100.0)
전수데이터**	345(76.3)	40(8.8)	45(10.0)	22(4.9)	452(100.0)
합계***	954(69.9)	170(12.5)	147(10.8)	93(6.8)	1,364(100.0)

\* 공공데이터: 행정안전부 메타데이터관리시스템, 공공데이터포털 등에서 이미 개방(공개)되어 서비스되고 있는 데이터

\*\* 전수데이터: 외부 개방 데이터 시스템(공공데이터포털)을 통해 개방(공개)하지 않은 데이터를 대상으로 조사된 수치임.

단, ① 기관 홈페이지 및 자체 정보시스템에서만 공개한 데이터, ② 전체 데이터 중 일부만 외부에 개방(공개)된 데이터(ex. 개인정보 포함으로 일부만 개방)는 조사된 보유 데이터에 포함

\*\*\* 합계: 공공데이터 현황조사(1,048개)와 보유데이터 전수조사(452개) 결과 합산 시 중복데이터로 선별된 136개 데이터를 제외한 후 합계를 제시함에 따라 개별 데이터 값과 합계 값이 일치하지 않음

문화체육관광부 소속 및 산하기관이 보유 및 운영·관리하고 있는 총괄 데이터 현황을 정부기능분류를 기준으로 업무 기능별(대기능)로 세분화하여 집계한 결과를 살펴보았다. 각 정책영역별 보유 데이터 현황의 경우 문화예술 정책영역에서는 문화산업정책(284개, 29.8%)과 예술정책(215개, 22.5%), 문화정책(163개, 17.1%) 데이터가 전체 문화예술 정책영역 데이터의 약 70% 비중을 차지하는 것으로 나타났으며, 관광영역에서는 관광정책(142개) 데이터가 96.6%로 높은 비중을 보이는 것으로 나타났다.

〈표 4-41〉 정부기능분류별(대기능) 보유 데이터 현황(전수+공공)

(단위: 개, %)

정책영역	데이터 수	대기능	데이터 수
문화예술	954(100.0)	문화정책	163(17.1)
		예술정책	284(29.8)

정책영역	데이터 수	대기능	데이터 수
		문화기반시설 운영관리	111(11.6)
		종교문화지원	2(0.2)
		문화산업정책	215(22.5)
		문화미디어정책	179(18.8)
체육	170(100.0)	체육정책	170(100.0)
관광	147(100.0)	관광정책	142(96.6)
		사행산업통합 감독위원회운영	5(3.4)
문화체육관광 일반	93(100.0)	문화관광 행정지원	93(100.0)

데이터 총괄 현황을 공동활용 정책분야별 보유 데이터 현황을 정부기능분류별 정책영역과 교차해서 살펴보았다. 앞서 분석한 전수데이터 조사결과와 유사한 형태의 분포를 가지는 것으로 분석됐다. 5가지의 공동활용 정책분야 중 “특화된 대책 및 맞춤형 서비스” 관련 데이터가 1,166개(85.5%)로 가장 많았고, 다음으로 “행정업무 경제성·효율성 개선” 112개(8.2%), “미래수요에 대한 선제적 대응” 40개(2.9%) 순으로 공동활용 정책분야 데이터 비중이 높은 것으로 조사됐다. “위험요소 예측 제거” 및 “국민의 의견 수렴” 관련 보유 데이터는 각각 18개, 28개로 전체 데이터의 1.3%, 2.1% 수준으로 보유 데이터의 수가 매우 작은 것으로 나타난 가운데, 5가지 공동활용 정책분야 모두에서 문화예술 영역으로 분류되는 데이터의 비중이 가장 높은 것으로 나타났다.

〈표 4-42〉 공동활용 정책분야별 정부기능분류(정책영역)별 보유 데이터 현황(전수+공공)

(단위: 개, %)

구분	문화예술	체육	관광	문화체육관광 일반	합계(%)
국민의 의견 수렴	27	1			28(2.1)
특화된 대책 및 맞춤형 서비스	825	168	147	26	1,166(85.5)
위험요소 예측 제거	18				18(1.3)
미래수요에 대한 선제적 대응	35	1		4	40(2.9)
행정업무 경제성·효율성 개선	49			63	112(8.2)
합계	954	170	147	93	1,364(100.0)

업무 기능별(대기능) 보유 데이터 현황을 살펴보면, 전체 문화체육관광 분야 데이터 1,364개 중 예술정책 데이터가 284개로 가장 많았고, 다음으로 문화산업정책 데이터가 215개, 문화미디어정책 179개, 체육정책 170개, 문화정책 163개, 관광정책 142개 등

의 순으로 데이터를 보유 또는 운영·관리하고 있는 것으로 나타났다.

보유 또는 운영·관리 데이터가 가장 많은 예술정책의 경우 “특화된 대책 및 맞춤형 서비스” 264개(94.7%)를 제외하면 다른 공동활용 정책분야별 관련 데이터는 모두 10개 미만인 것으로 나타났다. 다음으로 데이터수가 많았던 문화산업정책 관련 데이터는 다른 업무 기능별 분류와 달리, 공동활용 정책분야별로 고르게 분포되어 있는 것으로 분석됐다.

〈표 4-43〉 공동활용 정책분야별 정부기능분류(대기능)별 보유 데이터 현황(전수+공공)

(단위: 개, %)

구분	국민의 의견 수렴	특화된 대책 및 맞춤형 서비스	위험요소 예측 제거	미래수요에 대한 선제적 대응	행정업무 경제성·효율성 개선	합계(%)
문화정책		157		6		163
<b>예술정책</b>	<b>4(1.4)</b>	<b>269(94.7)</b>		<b>2(0.7)</b>	<b>9(3.2)</b>	<b>284(100.0)</b>
문화기반시설 운영관리		111				111
종교문화지원		2				2
<b>문화산업정책</b>		<b>137(63.7)</b>	<b>18(8.4)</b>	<b>20(9.3)</b>	<b>40(18.6)</b>	<b>215(100.0)</b>
문화미디어정책	23	149		7		179
체육정책	1	168		1		170
관광정책		142				142
사행산업통합 감독위원회운영		5				5
문화관광 행정지원		26		4	63	93
합계	28	1,166	18	40	112	1,364

## 나. 문화체육관광부 조직 분류 기준 데이터 현황 파악

문화체육관광부 조직체계는 문화체육관광부가 상시적으로 수행하는 업무를 기능수준에 따라 정부기능분류체계(BRM)를 기준으로 각 부처 대기능 분류체계와의 연계성을 고려하여 체계화(정부기능의 분류 및 관리에 관한 규정 제2조, 행정안전부 예규 제19호)되어 있다. 이에 본 연구에서는 문화체육관광부 행정 업무 기능과의 연계 및 행정 업무 기능별 데이터 현황 파악을 위해 데이터 기능수준에 따라 정책분야·정책영역·대기능으로 업무수행 기능을 구분 후 데이터전수조사를 진행하였다. 정부기능분류체계를 기준으로 연계한 문화체육관광 조직 분류와 정부기능분류 연계표는 〈표 4-44〉와 같다.

〈표 4-44〉 문화체육관광부 조직 분류와 정부기능분류 연계표

문화체육관광부 조직 분류*			분류(BRM)		
분야(6개)	대분류 (11개)	중분류(44개)	정책 분야	정책영역	대기능
문화예술	문화정책	문화정책, 국어정책, 전통문화, 국제문화, 예술정책, 공연전통예술, 시각예술디자인, 문화예술교육, 지역문화정책, 문화기반, 도서관정책기획, 중무	문화 체육 관광	문화예술	문화정책
	예술정책				예술정책
					문화기반시설 운영관리
					종무
콘텐츠	콘텐츠정책	문화산업정책, 영상콘텐츠산업, 게임콘텐츠산업, 대중문화사업, 한류지원협력, 미디어정책, 방송영상관고, 출판인쇄독서진흥		문화예술	문화산업정책
	미디어정책				문화미디어정책
저작권	저작권	저작권정책, 저작권산업, 저작권보호, 문화통상협력		문화예술	문화산업정책
체육	체육정책	체육정책, 체육진흥, 스포츠산업, 국제체육, 장애인체육, 스포츠유산		체육	체육정책
	체육협력				
관광	관광정책	관광정책, 관광진흥, 국제관광, 관광기반, 관광산업, 융합관광산업, 관광개발		관광	관광정책
	관광산업				사행산업통합 감독위원회 운영
공공행정	공공행정	정책기획, 소통, 홍보, 재정, 규제/법률, 정보화, 행정	문화체육관광 일반	문화관광 행정지원	

\* 문화체육관광부 조직도 기준(2021년 8월31일 기준)

문화체육관광부 소속 및 산하기관이 보유 및 운영·관리하고 있는 총괄 데이터 현황을 문화체육관광부 조직 분류를 기준으로 정부기능분류체계와 연계한 결과를 살펴보았다. 먼저, 분야별 보유 데이터 현황의 경우 전체 1,364개 데이터 중 문화예술(560개, 41.1%)이 차지하는 비중이 가장 높았고, 다음으로 콘텐츠(232개, 17.0%), 체육(170개, 12.5%), 저작권(162개, 11.9%) 등의 순으로 보유 또는 운영·관리되고 있는 데이터의 비중이 높은 것으로 나타났다.

이어서, 문화체육관광부 조직 분류 대분류별 보유 데이터 현황을 살펴보았다. 조직 분류 대분류는 문화체육관광부의 실국 수준의 업무수행 기능을 고려하여 정부기능분류 대기능과 연계한 분류단위로 볼 수 있다. 연계하여 분류된 조직 분류별 데이터 현황을 살펴보면, 전체 문화체육관광 분야 데이터 1,364개 중 예술정책 데이터가 284개(20.8%)로 가장 많았고, 다음으로 문화정책 데이터가 274개(20.1%), 미디어정책 179개

(13.1%), 저작권 162개(11.9%), 체육정책 156개(11.4%) 등의 순으로 데이터를 보유 또는 운영·관리하고 있는 것으로 나타났다. 종교문화지원과 관련된 종무 데이터의 경우는 단 2개(0.1%)의 데이터만이 생산·관리 되고 있는 것으로 분석됐다.

〈표 4-45〉 문화체육관광부 조직 분류 기준 보유 데이터 현황(전수+공공)

(단위: 개, %)

분야	데이터 수	대분류	데이터 수
문화예술	560(41.1)	문화정책	274(20.1)
		예술정책	284(20.8)
		종무	2(0.1)
콘텐츠	232(17.0)	콘텐츠 정책	53(3.9)
		미디어 정책	179(13.1)
저작권	162(11.9)	저작권	162(11.9)
체육	170(12.5)	체육정책	156(11.4)
		체육협력	14(1.0)
관광	147(10.8)	관광정책	58(4.3)
		관광산업	89(6.5)
공공행정	93(6.8)	공공행정	93(6.8)
합계		1,364(100.0)	

문화체육관광부 조직 분류를 기준으로 보유 데이터 현황을 공동활용 정책분야별로 구분하여 조사결과를 살펴보았다. 전체 문화체육관광 분야 데이터 1,364개 중 문화예술 데이터가 560개로 가장 많은 것으로 나타난 가운데, 문화예술을 다른 조직 분류(분야)와 비교해서 살펴볼 경우에도 5가지 공동활용 정책분야 모두에서 문화예술 관련 업무 데이터의 비중이 가장 높은 것으로 나타났다.

공동활용 정책분야별 보유 또는 운영·관리 데이터 현황을 조직분류와 교차해서 살펴 보았다. 〈표 4-45〉, 〈표 4-46〉에서 제시하고 있는 5가지의 공동활용 정책분야별 데이터 비중을 살펴보면, “특화된 대책 및 맞춤형 서비스” 관련 데이터가 1,164개(85.5%)로 가장 많았고, 다음으로 “행정업무 경제성·효율성 개선” 112개(8.2%), “미래수요에 대한 선제적 대응” 40개(2.9%) 순으로 공동활용 정책분야 데이터 비중이 높은 것으로 조사됐다.

〈표 4-45〉, 〈표 4-46〉를 살펴보면, 앞서 분석한 정부조직분류별 결과와 마찬가지로 조직 분류별로 분류된 데이터 수가 “0”인 셀이 상당수 존재하고 있었다. 특히, 11개 업무 단위로 구분된 문화체육관광부 조직 분류 중 종무, 콘텐츠정책, 체육협력, 관광정책,

관광산업 이상 5가지 분류는 “특화된 대책 및 맞춤형 서비스”를 제외한 나머지 4개 정책 분야에 포함되는 데이터가 전혀 없는 것으로 나타났다. 이와 같은 결과는 정부조직분류에 따른 결과와 동일하게 행정 업무 기능별로 다른 정책영역에서 활용될 수 있는 데이터에 대한 고려 및 생산·관리가 적절히 이뤄지지 않고 있기 때문으로도 생각해 볼 수 있다.

저작권 관련 데이터의 경우는 다른 10개 업무 단위 조직 분리와 달리, “국민의 의견 수렴”을 제외한 나머지 4개 공동활용 정책분야에 고르게 분포되어 있 것으로 나타났다. 구체적으로 저작권 관련 데이터 총 162개는 “특화된 대책 및 맞춤형 서비스(84개)”, “위험요소 예측제거(18개)”, “미래수요에 대한 선제적 대응(20개)”, “행정업무 경제성·효율성 개선(40개)로 데이터의 활용 분야를 분류할 수 있었다. 물론, “특화된 대책 및 맞춤형 서비스”에 활용될 수 있는 데이터의 비중이 크게 높았지만, 다른 조직분류 기준 업무 단위와 비교 시 데이터의 활용 유형이 고르게 분포되어 있음을 <표 4-45>, <표 4-46>를 통해 확인 할 수 있다. 이와 같은 결과는 저작권을 기반으로 생성·취득되고 있는 데이터들의 특성이 저작권산업 분쟁·침해·조정·심의 등과 관련된 고유의 행정 업무 수행을 통해 생산·관리되고 있기 때문으로 볼 수 있다. 실제 데이터전수조사에 참여한 저작권 관련 기관들이 응답한 결과에서도 동일한 이유로 복수의 공동활용 정책분야에서 활용 가능한 데이터인 보유 또는 생산·관리하고 있는 것으로 나타났다.

<표 4-46> 공동활용 정책분야별 문화체육관광부 조직 분류(분야)별 보유 데이터 현황(전수+공공)

(단위: 개, %)

구분	국민의 의견 수렴	특화된 대책 및 맞춤형 서비스	위험요소 예측 제거	미래수요에 대한 선제적 대응	행정업무 경제성·효율성 개선	합계(%)
문화예술	4	539		8	9	560(41.1)
콘텐츠	23	202		7		232(17.0)
저작권		84	18	20	40	162(11.9)
체육	1	168		1		170(12.5)
관광		147				147(10.8)
공공행정		26		4	63	93(6.8)
합계	28(2.1)	1,166(85.5)	18(1.3)	40(2.9)	112(8.2)	1,364(100.0)

〈표 4-47〉 공동활용 정책분야별 문화체육관광부 조직 분류(대분류)별 보유 데이터 현황(전수+공공)

(단위: 개, %)

구분	국민의 의견 수렴	특화된 대책 및 맞춤형 서비스	위험요소 예측 제거	미래수요에 대한 선제적 대응	행정업무 경제성·효율성 개선	합계(%)
문화정책		268		6		274(20.1)
예술정책	4	269		2	9	284(20.8)
종무		2				2(0.1)
콘텐츠 정책		53				53(3.9)
미디어 정책	23	149		7		179(13.1)
저작권		84	18	20	40	162(11.9)
체육정책	1	154		1		156(11.4)
체육협력		14				14(1.0)
관광정책		58				58(4.3)
관광산업		89				89(6.5)
공공행정		26		4	63	93(6.8)
합계	28(2.1)	1,166(85.5)	18(1.3)	40(2.9)	112(8.2)	1,364(100.0)

## 2.2. 공동활용 데이터 발굴

본 연구에서 실시한 데이터 현황조사(공공데이터 현황조사, 보유 및 수요 데이터 조사, 데이터 업무 활용 실태조사) 결과를 활용하여 데이터기반행정법에 명시된 정책분야에 부합하는 공동활용 데이터 목록화 작업을 진행하였다. 목록화 작업을 위해 기관별 보유 데이터 외에 기관 담당자별로 업무 수행 시 필요한 공동활용 데이터 수요를 파악하였고 필요시 기관이 보유한 공동활용 데이터 목록 및 시스템 정보 등 부가정보 요청도 병행하여 추진하였다. 단, 각 기관별 자체시스템에서 개방하고 있는 데이터를 제외한 외부에 개방(공개)하지 않은 데이터에 대해서는 공동활용 데이터 목록화 내용을 공개하지는 않았다.

### 가. 공동활용 가능 데이터 목록(문체부 소속 및 산하기관 보유)

문화체육관광부 소속 및 산하기관이 고유의 행정업무 수행을 통해 생성, 취득하여 운영·관리하고 있으나 외부에 개방(공개)하지 않은 데이터를 공동활용 가능 데이터 목록



대상으로 설정하였다. 기본적으로 메타데이터관리시스템, 공공데이터포털 등에서 개방(공개)된 데이터는 이미 각 부처 및 산하기관에서 공동활용 데이터로 활용되고 있으므로 데이터 목록화 작업에서는 제외하였다. 공동활용 데이터 목록에 포함될 수 있는 데이터는 기관 홈페이지 및 자체 정보시스템에서만 공개한 데이터를 기준으로 전체 데이터 중 일부만 외부에 개방(공개)된 데이터(ex. 개인정보 포함으로 일부만 개방)를 목록 대상 데이터에 포함하여 리스트 작업을 진행하였다. 구체적으로는 기관 홈페이지 및 자체 정보시스템에서만 공개한 데이터를 기준으로 일반적인 행정(보도자료, 취업현황 등) 데이터를 제외하고 고유 사업 추진(운영) 및 정책수립, 집행과정 모니터링, 정책평가 등 과정에서 관리 중인 데이터, 또는 행정안전부가 권고하고 있는 공동활용 정책분야(데이터기반 행정법 제8조제1항)에 포함되는 데이터로서 기관 간 공동 활용성이 높다고 판단되는 보유 데이터를 목록화 작업 대상으로 설정하였다.

본 연구의 데이터 전수조사를 통해 수집·정리됐으나, 외부 개방 시스템(포털) 등에서 단 한 번도 공개된 적이 없는 비공개 데이터(103개)에 대해서는 기관별 상황(데이터 신뢰성 및 원데이터 요청에 대한 우려, 기관 간 비교 등)을 고려하여 공동활용 가능 데이터 목록화 대상에서 제외하였다.

〈표 4-48〉 공동활용 가능 데이터 활용 분야

구분	공동활용 가능 데이터	활용 분야(안)
국민의 의견 수렴	뉴스빅데이터	• 국민의 삶과 밀접한 코로나19, 저출산, 고령화, 미세먼지 등의 주제에 대한 뉴스분석 및 공공데이터로 제공
특화된 대책 및 맞춤형 서비스	문화비소득공제DB	• 문화비 소득공제 제도의 '22년 이후 문화 분야 확대를 위한 기반 자료 활용
	문화누리카드DB	• 사회보장시스템(복지부) 빅데이터 연계를 통해 '문화누리카드' 미발급자 권리구제 및 비대면 신청·개선
	예술인활동증명DB	• 자체 예술인경력증명 데이터와 고용보험(고용부) 신고데이터 연계, 대상자 심사 등에 활용
	한국민속대백과사전DB	• 한국민속대백과사전의 표제어에 따른 사진목록, 사진상세정보, 음원목록 등을 가공하여 서비스
	말뭉치 (북한어 말뭉치 포함)DB	• 국어말뭉치, 한국어-외국어 병렬 말뭉치, 한국어점자 병렬 말뭉치 등 빅데이터 구축 및 SI학습 등에 제공
	국민체력100센터 측정DB	• 국민·지역 맞춤형 생활체육 정책 개발에 활용
	국가문화예술지원DB	• 예술위원회 보유데이터 현황 분석 및 문화예술과 유관 산업의 융합 도모 방안 연구 • 예술인 보조금 신청·지원·심의 분석
	게임물등급분류 정보	• 민간 사업자 간 등급 분류 정보 공유 및 분석을 통한 등급 분류 적정성 및 신뢰성 제고

구분	공동활용 가능 데이터	활용 분야(안)
	영상물등급분류 정보	• KMDb 영화정보DB 구축 및 수집자료(납본자료, 기증자료 등) 목록DB 구축에 참고, 활용
	문화체육관광산업 표본틀	• 문체부 사업체 조사 필수 표본틀 데이터로 활용
	출입국통계	• 국민 및 외국인의 출입국 데이터를 활용하여 연구 및 관광 정책 수립의 기초자료 활용을 위한
	업종별 신용카드 지출액 분석	• 코로나19 전후 국내외 신용카드 지출액 분석, 구간단위 속보성 동향 분석으로 코로나19 정책대응 내부자료로 활용
	문화체육관광 서비스업생산지수	• 콘텐츠, 스포츠, 관광산업 분야 서비스업 지원 방안 확보를 위한 근거자료 활용
	올림픽공원 운영·관리 정보	• 공단 RunningMate 앱으로 측정된 올림픽공원 이용객 데이터 분석을 통한 공원 운영 환경 개선
위험요소 예측 제거	관광불편신고 종합분석 정보	• 관광불편신고센터에서 접수된 관광불편신고 사항 종합 분석
	국가별 침해대응 정보	• 저작권 침해·분쟁·심의 기초자료로 활용
	카지노 객장 운영 정보	• 7Luck 카지노 지점별 Google Maps Review 분석
미래수요에 대한 선제적 대응	저작권 침해·분쟁·심의·조정 정보	• 온라인저작권 침해 등 선제 대응을 위해 저작권침해 정보, 침해사이트 정보 등 통합 구축, 심의·수사용 정보 제공
	공연예술통합전산망DB	• 코로나19 이후의 공연·영화 등 주요 문화 관련 업종의 일일 피해 현황(예매취소, 관객변동, 피해액추정 등) 분석
	영화관입장권 통합전산망DB	
행정업무 경제성·효율성 개선	국가자료종합목록 소장자료	• 국가자료종합목록의 전국 공공도서관 소장자료 정보를 국립장애인도서관 책나래 서비스에 활용
	국립중앙도서관 소장자료	• 국립중앙도서관의 소장자료 정보를 국립장애인도서관 소장자료 서지정보 정리 및 이용자 대체자료 제작 신청에 활용

〈표 4-49〉 정책분야별 공동활용 가능 데이터 목록(문체부 소속 및 산하기관 보유)

구분	공동활용 가능 데이터	생성·취득 방법	데이터 보유기관
국민의 의견 수렴	뉴스빅데이터	2차 가공	한국언론진흥재단
특화된 대책 및 맞춤형 서비스	문화비소득공제DB	행정자료	한국문화정보원
	문화누리카드DB	행정자료	한국문화예술위원회
	예술인활동증명DB	행정자료	한국예술인복지재단
	한국민속대백과사전DB	2차 가공	국립민속박물관
	말뭉치(북한어 말뭉치 포함)DB	2차 가공	국립국어원
	국민체력100센터 측정DB	행정자료	국민체육진흥공단
	국가문화예술지원DB	직접조사(행정자료)	한국문화예술위원회
	게임물등급분류 정보	2차 가공	게임물관리위원회
	영상물등급분류 정보	2차 가공	영상물등급위원회
	문화체육관광산업 표본틀	2차 가공	한국문화관광연구원
	출입국통계	2차 가공	한국문화관광연구원
	업종별 신용카드 지출액 분석	2차 가공	한국문화관광연구원
	문화체육관광 서비스업생산지수	2차 가공	한국문화관광연구원
	올림픽공원 운영·관리 정보	2차 가공	한국체육산업개발

구분	공동활용 가능 데이터	생성·취득 방법	데이터 보유기관
위험요소 예측 제거	관광불편신고 종합분석 정보	행정정보	한국관광공사
	국가별 침해대응 정보	2차 가공	한국저작권위원회
	카지노 객장 운영 정보	행정자료	그랜드코리아레저
미래수요에 대한 선제적 대응	저작권 침해·분쟁·심의·조정 정보	2차 가공	한국저작권위원회
	공연예술통합전산망DB	행정자료	예술경영지원센터
	영화관입장권 통합전산망DB	행정자료	영화진흥위원회
행정업무 경제성·효율성 개선	국가자료종합목록 소장자료	행정자료	국립중앙도서관
	국립중앙도서관 소장자료	행정자료	국립중앙도서관

## 나. 공동활용 가능 데이터 목록(타 기관 보유)

각 기관의 정책수립 및 분석업무 등에 필요한 타 기관 데이터에 대한 수요조사 결과를 통해 수집·정리된 데이터를 목록화 하였다. 기존 업무 수행 시 이미 활용하고 있는 타 기관 데이터를 포함하여, 기존 업무 수행 시 필요한 데이터가 있었으나, 데이터 보유기관 으로부터 제공이 되지 않았던 데이터, 사업추진의 효율화 및 정책수립, 집행과정 모니터링, 정책평가 등 과정에서 필요한(필요할 것으로 예상되는) 타 기관 데이터를 기준으로 <표 4-50>과 같이 공동활용 가능 데이터를 목록화 하였다.

<표 4-50> 정책분야별 공동활용 가능 데이터 목록(타 기관 보유)

구분	공동활용 가능 데이터	데이터 보유기관
국민의 의견 수렴	민원·청원 등 의견을 함의한 데이터	각 기관별 홈페이지, 청와대
특화된 대책 및 맞춤형 서비스	예술분야 고용보험DB	고용노동부 (한국고용정보원)
	사업체노동력조사	
	예술인 소득정보DB	국세청
	국세통계	
	노인실태조사	보건복지부
	아동종합실태조사	
	장애인실태조사	
	건강보험통계	국민건강보험공단
	가계동향조사	통계청
	생활시간조사	
	기업생멸행정통계	
	지식재산권 수지	한국은행
	이동통신데이터	통신 3사(SK, KT, LG)
위험요소 예측 제거	신용평가사 신용 평점 데이터	나이스 신용평가
	금융 (DTI 등) 활동 상세 정보	금융감독원

구분	공동활용 가능 데이터	데이터 보유기관
미래수요에 대한 선제적 대응	교통정보, 고속도로 통행량	교통연구원, 한국도로공사
행정업무 경제성· 효율성 개선	장애인증명 정보	보건복지부
	학교 교사 정보	한국교육학술정보원
	국고 보조금 현황	한국재정정보원
	지자체별 관광지 이용 현황 정보	지자체별 행정자료

〈표 4-51〉 공동활용 가능 데이터 활용 분야

구분	공동활용 가능 데이터	활용 분야(안)
특화된 대책 및 맞춤형 서비스	예술분야 고용보험DB	예술인 복지정책 수립을 위해 고용보험에 가입된 예술인들의 인원수, 지역별 분포, 소득, 실업급여 수혜 현황 파악
	사업체노동력조사	문화체육관광 분야 산업별 소분류 또는 세분류 수준 자료의 공동활용이 가능하다면 일자리 정책의 효과성을 모니터링할 수 있는 세부 지표로 활용
	예술인 소득정보DB	예술인 복지정책 수립을 위해 건강보험에 가입된 예술인의 가입자 정보 파악
	국세통계	관광사업체의 특성을 분석하기 위해서는 세세분류 업종에 해당하는 정보가 필요함, 지방행정개방데이터 보완자료로 활용
	노인실태조사	노인의 여가문화활동 참여 행태와 운동참여 장소(노인여가복지시설 등) 선호도 등을 파악하여 노인체육진흥 개발 자료로 활용
	아동종합실태조사	조사대상이 0~18세로, 국민생활체육조사의 조사대상(10세이상)이 아닌 0~9세까지의 운동참여 현황, 방과후 희망 활동 등의 자료를 파악, 정책개발의 기초 자료로 활용
	장애인실태조사	장애유형별 장애인 현황, 발생률 등 실태를 파악하여 정책수립 시 반영
	건강보험통계	의료비 지출 관련 정보와 생활체육 참여 자료를 결합하여 체육활동의 경제적 효과를 측정
	가계동향조사	가계의 스포츠 관련 지출을 분석하여 생활체육 및 스포츠산업 정책 수립의 기초 자료로 이용 맞춤형 정책 설계를 위해 가계지출동향의 세부 현황파악
	생활시간조사	국민의 시간 활용 현황을 파악하여 생활체육 참여 활성화 정책의 근거 자료로 이용
	기업생멸행정통계	시간 흐름에 따른 관광사업체의 변화를 분석하기 위해 기업생멸행정통계의 신생률, 소멸률, 신생기업 생존율, 고정장기업 비율 등 통계 활용 필요
	지식재산권 수지	맞춤형 정책 설계를 위해 지식재산의 무역의 세부 현황파악 필요
행정업무 경제성·효율성 개선	장애인증명 정보	국립장애인도서관 및 장애인 대상 도서관 서비스를 제공하는 공·사립 도서관에서 회원의 장애사실여부를 온라인으로 확인 가능하도록 하여 서류 제출에 따른 이용자의 불편을 감소시키는데 활용
	학교 교사 정보	학교 문화예술교육 정책을 지원하기 위해 학교분야 학교 교사 정보(교원번호, 성명, 근무학교명 등) 파악
	국고 보조금 현황	보조금의 효율적 운영을 위해 보조금의 세부 현황 파악

## 2.3. 정책분야별 분석활용(안)

### 가. 정책분야별 분석활용(안) 도출

본 연구에서 실시한 데이터현황조사 결과를 기반으로 수집·정리된 문체부 소속 및 산하기관에서 분산 축적·관리되고 있는 다양한 유형의 데이터 정보 및 공동활용 가능 데이터로 목록화된 데이터를 기준으로 정책분야별 분석활용 과제(안)을 도출하였다. 데이터기반행정법 제8조 1항에서 명시됨과 동시에 행정안전부가 권고하고 있는 공동활용 정책분야별(국민의 의견 수렴, 특화된 대책 및 서비스, 위험요소 예측 제거, 미래수요에 대한 선제적 대응, 행정업무 경제성·효율성 개선)로 활용될 수 있는 문화체육관광 분야 데이터를 기준으로 타 부처 및 민간 데이터와 공동활용할 수 있는 대상 데이터를 설정하였고 이를 토대로 행정내부 서비스 및 국민 서비스 개선을 위해 효율적으로 활용될 수 있는 정책분야별 분석활용(안) 17개를 <표 4-52>와 같이 제시하였다.

<표 4-52> 정책분야별 분석활용(안)

정책분야	분석명(안)	공동활용 데이터(안)	활용(안)
국민의 의견 수렴	국민 삶의 질 (체감 만족도) 분석	<ul style="list-style-type: none"> <li>• (문체부)지역별 문화·예술 제공 정보, 문화·예술 분야 통계(국민문화예술활동조사, 국민여가활동조사)</li> <li>• (관련부처)인구 정보, 자살률, 실업률, 비만율, 스트레스, 국민소득정보, 가계지출통계</li> </ul>	문화·체육·예술 분야의 공공데이터를 기반으로 사회적 병리현상 데이터와 연계·분석하여 국민 삶의 만족도 향상을 위한 정책 분석
	예술인 사회적 포용·이동성 지표 개발·분석	<ul style="list-style-type: none"> <li>• (문체부)예술인등록DB</li> <li>• (관련부처)근로소득, 사업소득, 기타소득, 연금소득, 인구가구통계등록부, 주택통계등록부, 경제활동통계등록부, 고용보험DB, 학정 정보, 주택소유DB</li> </ul>	문화예술인 소득계층별 고용수준/주거수준/건강수준 등의 현황을 파악하고 상대적 격차 파악
특화된 대책 및 맞춤형 서비스	K-컬처 해외 홍보 활성화 전략 수립 분석	<ul style="list-style-type: none"> <li>• 제작 및 배급사 현황, 해외 공연 현황, 한국 문화원 해외반응 정보</li> <li>• 해외 SNS 정보, 해외 차트 순위, 영상관련 순위</li> </ul>	K-컬처(음악, 영상, 음식) 해외 홍보 강화를 위한 현황 및 해외 진출 촉진을 위한 대상자 선별
	문화체육관광 산업 기업생멸현황 분석	<ul style="list-style-type: none"> <li>• (문체부)콘텐츠, 스포츠, 관광 산업 모집단DB</li> <li>• (관련부처)사업자등록자료, 부가가치세, 법인세, 근로소득지급명세서 및 법인합병신고서</li> </ul>	시간 흐름에 따른 기업의 신생, 소멸, 생존율, (고)성장 등 기업의 생애 주기를 파악분석
	문화관광 분야별 프리랜서 /	<ul style="list-style-type: none"> <li>• (공공데이터) - 정부/지자체 코로나 지원금 수급 내역</li> </ul>	문화계 종사 비정규직 프리랜서 종사자들의

정책분야	분석명(안)	공동활용 데이터(안)	활용(안)
	특수 고용자들의 코로나 고용 불안정 현황 분석	<ul style="list-style-type: none"> <li>실업 급여 및 고용 보험 정보</li> <li>•(민간데이터)               <ul style="list-style-type: none"> <li>- 여가 분야별 지출액 규모 추이 (코로나 이후)</li> </ul> </li> <li>각 지역별 문화관광관련 업체 개폐업 정보</li> </ul>	분야별 정부/지자체 지원금 수급 현황 분석
	지역경제 활성화를 위한 맞춤형 타겟 마케팅 분석	<ul style="list-style-type: none"> <li>•(민간데이터-정형)               <ul style="list-style-type: none"> <li>- 최근 2년간 소비 매출 데이터</li> </ul> </li> <li>•(공공데이터-정형)               <ul style="list-style-type: none"> <li>- 지역별 업종별 매장 정보</li> </ul> </li> </ul>	외부 관광객을 대상으로 한 소비 데이터를 분석하여 업종별 주요 고객층 정보 구축
	고용보험 자료를 활용한 문화체육관광 총조사 동향분석	<ul style="list-style-type: none"> <li>•(공공데이터)               <ul style="list-style-type: none"> <li>- 문체부 데이터 연계(예술인활동증명DB, 예술분야 고용보험DB, 사업체노동력조사, 예술인 소득정보DB)</li> <li>- 고용센터 콜 상담자료(문화체육관광 관련)</li> <li>- 워크넷 구인정보, 입사지원, 직업상담자료</li> <li>- 고용보험DB</li> </ul> </li> </ul>	고용부와 문체부 연계 “창·단·다·탈 일자리 사업” 등 고용과 문화의 연계한 사회현상 분석
	33개 전국의 관광특구 산업현황	<ul style="list-style-type: none"> <li>•(공공데이터)               <ul style="list-style-type: none"> <li>- 관광특구 지정현황 자료</li> <li>- 전국사업체조사</li> <li>- 통계지리정보서비스(SGIS)</li> </ul> </li> </ul>	관광특구에서 산업활동을 하는 사업체의 산업현황을 파악하여 지원정책 수립에 활용
	국민체력 빅데이터 기반 선제적 맞춤 건강유지 정책 마련	<ul style="list-style-type: none"> <li>•(공공데이터)               <ul style="list-style-type: none"> <li>- 국민체력100센터 측정DB</li> <li>- 국민건강보험 빅데이터</li> <li>- 건강보험심사평가원 빅데이터</li> <li>- 국립암센터 암환자 빅데이터 통계청</li> </ul> </li> </ul>	국민체력100센터 DB를 바탕으로 체력등급과 국민건강보험등 다양한 의료빅데이터를 연계하여 체력등급 및 유자가 개인의 건강과 어떠한 연관성이 있는지를 분석
위험요소 예측 제거	문화·관광수요 영향요인 분석	<ul style="list-style-type: none"> <li>•(공공데이터)               <ul style="list-style-type: none"> <li>- 경제 및 금융, 질병 및 위생, 외교 갈등, 재난재해 및 테러, 남북관계, 사회문화적 요인 정보 및 데이터</li> <li>- 기상·대기 정보, 위성사진, 위험시설물</li> <li>- 시계열 재난 재해 현황, 기상정보, 위성사진, 재난재해 민원/영상정보</li> </ul> </li> </ul>	경제위기, 질병위기, 자연재해, 인적재해 등 국가적 차원의 대응책 마련 및 피해 최소화를 위한 예측 분석에 활용
	개인의 여가 활동 기반의 개인 파산 고위험 고객군 사전 예측	<ul style="list-style-type: none"> <li>•(공공데이터)               <ul style="list-style-type: none"> <li>- 내국인의 카지노 이용 DB</li> <li>- 개인 파산 신청 이용 고객</li> </ul> </li> <li>•(민간데이터)               <ul style="list-style-type: none"> <li>- 신용평가사 신용 평점 데이터</li> <li>- 금융 (DTI 등) 활동 상세 정보</li> </ul> </li> </ul>	개인정보 비식별화를 통해 카지노를 포함한 gambling 여가 (경마, 경륜 등) 의 이용 고객에 대한 정보를 민간 신용 평가사의 데이터와 결합하여 개인의 신용도 혹은 개인 파산 위험 고객군을 사전에 예측할 수 있는 지수 개발

정책분야	분석명(안)	공공활용 데이터(안)	활용(안)
미래수요에 대한 선제적 대응	관광·축제 분야 흥행 요인 분석	<ul style="list-style-type: none"> <li>• (문체부)관광지 정보, 유동(유입)인구 현황, 축제지역(위치), 교통정보, 축제만족도 정보, 축제 이용현황</li> <li>• (공공 및 민간)소비현황, 지출정보, 교통정보</li> </ul>	축제 수요 및 성공적인 축제를 통한 지역 경제 활성화 전략 수립
	주요 국가별 외국인의 국내 관광 수요 예측	<ul style="list-style-type: none"> <li>• (내부데이터) <ul style="list-style-type: none"> <li>- 국가별 출입국 통계</li> <li>- 외국인 국내 관광 실태 조사</li> </ul> </li> <li>• (외부데이터) <ul style="list-style-type: none"> <li>- 외국인의 국내 소비 데이터</li> <li>- 외국인의 자국 외 관광 관련 소비 데이터</li> </ul> </li> </ul>	국가별 입국자수 자료와 외국인의 국내 관광 관련 업종의 소비 관련 데이터, 국가별 백신접종을 추이, 외국인의 자국 외 기타 국가에서의 관광 관련 소비 데이터를 활용하여 시계열 예측 모델 기반의 외국인 관광객 예측
행정업무 경제성·효율성 개선	관광지 순환형 버스 도입 분석	<ul style="list-style-type: none"> <li>• 지역별관광지정보,</li> <li>• 예상관람시간예측정보,</li> <li>• 지자체내버스이용현황</li> <li>• 관광지선호도조사, 관광홍보지원현황, 지역내 관광버스현황, 준공영버스현황, 시내버스노선 정보, 대중교통요금정보</li> </ul>	문화재 및 지역 관광지 관람을 위한 순환버스 도입을 위한 현황 및 이용률 파악
	체육시설별 이용 현황 분석을 통한 체육장비 보급 및 추가 시설 위치 선정	<ul style="list-style-type: none"> <li>• (민간데이터-비정형) <ul style="list-style-type: none"> <li>- SNS에 언급된 체육시설 현황 및 필요사항 등</li> </ul> </li> <li>• (공공데이터-정형) <ul style="list-style-type: none"> <li>- 지자체별 체육시설 현황</li> <li>- 체육시설별 이용객 현황 및 장비 보유 현황</li> </ul> </li> </ul>	지역별 체육시설 현황과 보유 장비 및 이용객 현황 등을 분석
	주요관광지에 위치한 음식점들의 SNS 화제어를 파악하여 홍보 방안 제시	<ul style="list-style-type: none"> <li>• (민간데이터-비정형) <ul style="list-style-type: none"> <li>- SNS에 노출된 관광지별 음식점 등 화제어 현황</li> </ul> </li> <li>• (공공데이터-정형) <ul style="list-style-type: none"> <li>- 음식점(모범)업 주요 현황</li> <li>- 주요 관광지별 입장객(월별, 주간별) 현황</li> </ul> </li> </ul>	지역별 주요 관광지와 관광지에 위치한 주요 음식점의 현황 등을 분석
	지자체 단위 둘레길 코스 및 활성화 아이디어 발굴을 위한 이용객 분석	<ul style="list-style-type: none"> <li>• (기초자치단체 데이터) <ul style="list-style-type: none"> <li>- 둘레길 코스 현황</li> <li>- 지자체별 완주자 증감 현황</li> <li>- 둘레길 관련 민원 키워드</li> </ul> </li> <li>• (광역, 민간 데이터) <ul style="list-style-type: none"> <li>- 유동 인구 데이터</li> <li>- 코스 인근 지역 대중교통 이용 정보</li> </ul> </li> </ul>	지역별 둘레길 코스의 난이도 * 접근성에 따른 이용객 추이 비교 분석

## 나. 정책분야별 분석과제(안) 발굴

앞서 제시한 공동활용 데이터 목록 및 정책분야별 분석활용(안)을 활용 대상으로 총 3차례의 전문가 자문회의 및 FGI를 실시하여 정책분야별 분석과제(안)을 발굴·제시하였다. 먼저, 2차례의 전문가 자문회의를 통해 문화체육관광 분야 데이터에 대한 데이터품질, 데이터 공동활용, 데이터 연계·결합 방안 등에 대해 논의하였고, 1차례의 FGI 실시를 통해 앞서 도출된 17개 분석활용(안)을 토대로 문화체육관광 분야 데이터로 활용 가능한 정책분야별 기획과제 및 기관 간 협업 과제 등에 대한 의견을 수렴하였다. 최종적으로 전문가 자문회의 및 FGI 실시를 통해 논의된 결과를 취합·정리하여 정책분야별로 분석 가능 주제, 분석 배경 및 목적, 공동활용 대상 데이터, 분석 설명 및 방법, 기대효과 등으로 구체화한 12개 분석과제(안)를 제시하였다.

〈표 4-53〉 전문가 자문회의 및 FGI 의견 요약

분야	논의 내용
데이터 품질	<ul style="list-style-type: none"> <li>공공데이터 수집 및 제공은 행정안전부 국가사업의 일환으로 공공데이터 포털에서 많은 자료가 올라와 있으나, 공공데이터포털에 올라온 데이터를 보면 실적을 채우기 위해 올리는 단순 리스트 형태의 파일이 대부분이라 데이터 질이 낮음</li> <li>공공데이터를 쓰고 싶어도 데이터 품질 및 연계성이 좋지 않아 민간기업 내부통계나 기존 행정통계를 활용하는 경우가 대부분</li> <li>데이터의 정확성을 담보하기 어려움. 데이터 제공 요청 시, 데이터의 품질보다는 건수에 대해 초점이 맞추어지다보니, 때로는 건수가 부풀려지는 경우도 있을 것으로 예상됨. 결과적으로, 데이터 수집에 대해서 데이터 요청 기관의 명확한 정의와 기준이 있어야 함</li> <li>데이터 수집 시 용어에 대한 이해가 기관별, 담당자별 상이함. 즉, 즉 표준용어의 정립이 필요함(ex. 예를 들어 수량 1이 뜻하는 바가 '1페이지', '1권', '1건' 등으로 다양하게 해석 가능함. 따라서 이러한 기본적인 수량값에 대한 명확한 정의가 필요하며, 아울러 조사 항목의 유지를 통해 변동성 없이 유지해야 하는 사항으로 판단됨. 즉, 데이터추적에 대한 방법을 고정화시켜놔야 데이터의 일관성과 활용성이 상승됨)</li> <li>문체부 소속 및 산하기관 데이터 수집·관리 현황을 살펴보면, 대부분의 기관에서 디지털 데이터로 쌓이긴 하나, 표준화가 되어 있지 않고, 데이터 적체를 개인 pc 등에 쌓아두고 있음. 즉, 생산 데이터의 활용 체계가 구비되어 있지 않음 △ 관련 법들이 제정·시행될 때마다 필요에 의해 데이터 제공이 개별 부서 수준에서 이루어짐 △ 데이터 제공이 통합적인 통로를 통해 이루어지지 않고 있기 때문에 결과적으로, 담당자가 변경 시 데이터 관리에 문제가 발생할 수 있음</li> <li>데이터 수집의 비연계성 등에 대한 문제는, 기관마다 해결하는 방법 밖에는 대안이 없음. 보다 체계적으로 데이터 수집을 유지하고자 한다면, 경영평가와 같은 체계에 해당 내용이 명시되어야 구조적 문제 해결이 어느 정도 해소 가능할 것으로 예상됨</li> </ul>
데이터 공동활용 현황	<ul style="list-style-type: none"> <li>데이터를 제공받더라도 데이터 프레임, 칼럼 기준 등이 일정하지 않다보니 관련 내용을 정리하는데 시간이 더 걸리는 상황이 발생</li> <li>데이터 작업의 기본인 데이터 정의부터 문제가 생기고 데이터 구조, 변수하나 추가하는</li> </ul>



분야	논의 내용
	<p>것에 있어 법적, 행정적 문제가 생기자보니 데이터를 제대로 활용하지 못하고 있음</p> <ul style="list-style-type: none"> <li>이전에 조사가 불가능하였던 장애인이 있는 사업체도 복지부의 장애인등록명부, 사업체 대표자 이름 매칭, 연령대 분류, 거주지 등의 세팅 과정을 통해 추정 과정을 진행</li> <li>그럼에도 실제 조사대상 매칭 확률은 낮았는데 근데 여기에 공공데이터 중 보건복지부 생년월일 데이터를 통해 추가 추정과정을 진행하니 매칭 확률이 크게 증가함</li> <li>정보의 비대칭성으로 인해 소외계층의 배제문제에서 최근 코로나로 그나마 DB가 쌓이고 있는 추세</li> <li>재난지원금으로 문화체육관광 종사자에 대한 DB, 특히 취약계층 DB가 새롭게 만들어졌을 것이고 재난지원금을 사용한 카드사 데이터와 매칭시켜서 연구과제를 발굴해 보는 것을 권함</li> <li>재난지원금으로 프리랜서 관련 데이터가 쌓여가고 있으며, 문화·관광·체육은 코로나 피해가 크기 때문에 포스트코로나 관련 종사자 재고용 등의 사업 추진에서 카드사 데이터를 활용할 수 있을 수 있음</li> <li>국민체육진흥공단에서 관리하고 있는 체육시설 안전 정보에 대해 안전관리 관련 업체에서 이 정보를 가지고 서비스 제품을 만들어서 공단의 자료를 계속 제공하고 있음</li> <li>‘문예연감’은 보고서 출간이 30년 이상 된 만큼 데이터 상당히 축적되어 있음. 해당 데이터를 활용하고자 한다면, 이용 가치가 있고, 파급력이 클 것으로 예상됨</li> <li>문예위 시스템 데이터 중 해당 시스템의 데이터는 외부에 비공개 상태임. 다만, e나라도움시스템 도입 후 NCAS에 데이터에 축적이 되지 않은 상태임. 2021~2022년 사이, NCAS 정비 진행 예정</li> <li>문예위의 보유 데이터 현황 조사 결과, 개인정보 미포함으로 제출하였는데, 실제로 개인정보 포함여부에 대한 확인 필요함. 개인정보가 포함되어 있을 경우, 예술인복지재단의 데이터와 연계해서 활용 가능</li> </ul>
데이터 공동활용 한계	<ul style="list-style-type: none"> <li>데이터 법을 통해 협조요청을 오게 되면 제공되게 되어 있지만 개인정보 보호 문제가 워낙 사회적으로 예민하기 때문에 제공의 한계가 존재</li> <li>마스킹 기법 등 개인정보를 유출 안되는 정도에서 데이터를 받는 방법이 있지만 개인정보 보호를 이유로 제공 거절하거나 지나친 스크리닝과 필터링으로 아예 관련 정보를 받아도 활용 못하는 경우가 많음</li> <li>이로 인해 공약을 위해서 데이터 제공이 필요한 경우에도 데이터 공유가 이뤄지지 않고 있음</li> <li>실태조사를 진행하다보면 특정 집단(ex. 장애인, 여성기업)을 대상으로 조사를 진행하는데 이와 관련된 공공데이터가 보건복지부나 여성가족부 같은 곳에 당연히 있을 거 같은데 없음</li> <li>등록확인서와 같은 데이터가 있지만 일괄적 기준이 없으며, 등록확인서 발급에 있어 원하는 사람만 발급하다보니 실제 현실과의 차이가 발생</li> <li>그렇기에 등록확인서와 같은 것에 지나치게 의존하게 되면 특정 계층이 소외되는 현상이 발생가능</li> <li>빅데이터 산업이 빠르게 성장하고 있지만 그렇기에 이 빅데이터 조사범주 안에 들어오지 않는 사람들은 소외되는 상황이 발생할 수 있음</li> <li>예를 들어, 매출이 0원이라 사실상 폐업이지만 신고형태상 폐업이 아닌 상황 발생, 취직은 했지만 고용보험에 들지 못한 청년 등은 조사대상에서 제외되는 상황이 발생</li> <li>개인정보 노출 문제에 대해 빅데이터센터에 근무하는 참여자는 개인정보 등을 마스킹으로 처리하고 조사 및 연구의 꼭 필요한 사항만 오픈데이터로 돌리는 방향으로 데이터를 활용해야 함</li> <li>수동적 행정 처리를 극복하기 위해서는 공무원 대 공무원이 아닌 기관장 대 기관장으로 자료 활용에 대한 공식 MOU를 맺어야 함</li> <li>기관장 차원의 업무 협약이 이뤄져야 그 밑의 공무원들이 자유롭게 데이터를 요청, 제공하고 연계하여 활용하는 것이 가능할 것임</li> </ul>

분야	논의 내용
	<ul style="list-style-type: none"> <li>api 형태로 시스템 연계 제공 시, 시스템 개편 등의 변경 사항이 발생할 경우, api가 끊겨있는 경우들 발생. 이를 해결하기 위해서는 △ 정보화 부서와 데이터 부서 모두 있어야 하고, 표준화된 시스템이 구비가 되어 있어야 함 △ 데이터기반행정법에 의거, 데이터 연계 관련 공동연구를 진행하기 위해서는 이에 앞서 데이터 표준화가 선행되어야 함</li> </ul>
데이터 연계·결합	<ul style="list-style-type: none"> <li>공공데이터 사용 활성화가 어려운 이유 중에 하나는 다루고 있는 범주가 크다는 것</li> <li>범주가 크기에 대표 범위의 한계, 데이터의 양, 같은 범주 차원의 연계 데이터 부재 등으로 사용하고 있지 못하고 있음</li> <li>특정 계층을 타겟팅하여 그와 관련된 데이터만을 수집, 추출하여 연구하는 것이 공공 데이터의 장점을 살릴 수 있을 것</li> <li>한 사람의 주소를 알고 있지만 소득은 알지 못하는 상황에서 일반적으로 사용하는 방법 중 하나가 그 지역의 평균값으로 결측값을 대체하는 것</li> <li>이러한 2차 가공은 현실을 왜곡시킬 가능성이 높아 키값 매핑에 기반하여 필요한 정보와 정보가 정확히 매칭되는 1 to 1 데이터가 필요</li> <li>이렇듯 잘못된 빅데이터 활용은 배제, 왜곡현상을 가져올 수 있기에 대표성 측면에서 주의를 기울여야 함</li> <li>민간 기업 같은 경우, 자체 데이터를 가지고 있으며 이를 수익을 위해서 어떻게든 활용하려고 노력하고 있음</li> <li>단순 내부 데이터 뿐 아니라 외부데이터를 연계하여 분석하고 있으며 해외에서도 활용을 많이 하고 있음</li> <li>민간 기업에서도 그러나 결합을 했을 때 키값의 정밀함, 연계를 위한 관계의 어려움이 있기 때문에 대기업-대기업 단위의 MOU 체결을 통해 해결해가고 있지만 공공기관이 이런 식의 파트너십을 만들어가는 것은 쉽지 않을 것으로 예상</li> </ul>
데이터 공동활용 향후 발전 가능성	<ul style="list-style-type: none"> <li>데이터 연계 목적에 따라 발전 가능성은 크게 차이가 날 것으로 예상</li> <li>마케팅 쪽에서는 데이터의 대표성이 떨어지고 부정확해도 활용이 가능 하지만 공공기관, 특히 통계 생산의 측면에서 봤을 때는 한계가 있음</li> <li>개인정보 등에 대한 민감성, 대표성 문제가 있으며 특히 통계 가공, 생산의 측면에서 이 문제는 굉장히 큼</li> <li>공공데이터는 이를 만족시키기 힘들고 그렇다고 공공데이터를 위해 완전히 새로운 대표성 있는 데이터 생산을 하는 것은 여러 제약으로 어려움</li> <li>또한 단순한 경향, 의견파악이 아닌 정책 발굴, 예산투자과 연계된다면 공공데이터 활용에 대한 리스크는 엄청 커지게 될 것</li> <li>물론 공공기관도 활용목적에 따라 대표성을 포기하고 문제의식, 인사이트 도출 등을 위해서는 충분히 활용할 수 있을 것</li> <li>명확한 문제의식을 가지고 이러한 문제의식에 대한 답을 찾는 과정에서 어떠한 데이터가 필요한지 정리해야 함</li> <li>다양한 데이터를 모아서, '대단한' 인사이트를 도출하려고 하기 보다는, 기초적인 데이터를 전국단위로 제대로 모은다면 그 자체가 인사이트가 될 것으로 예상됨. 시범사업으로, 기초적인 데이터들을 전국 단위로 수집하는 방법을 추진해보는 방법을 제안</li> </ul>

## 1) 국민의 의견 수렴

	분야별 공동활용 대상 데이터	데이터(예시)
주요	문화예술인 민원·청원 등의 의견 수렴 및 삶의 수준을 보여주는 데이터	예술인등록DB, 민원 처리현황, 소득데이터(국세청), 고용보험DB, 주거DB, 건강보험DB, 문화예술인지원내역
부가	예술인의 의견을 수렴하는 과정에서 주요 데이터와 함께 활용할 수 있는 데이터	인구가구통계등록부, 주택통계등록부, 경제활동통계등록부, 실업률, 학적정보

### 《 분석 과제(안) 》

#### 분석 주제

문화예술인 사회적 포용·이동성 지표 개발·분석

#### 배경 및 목적

- 예술인들의 기본적 삶의 질 보장을 위한 최소수준과 집단 간 상대적 격차를 진단하는 지표
- 계층·집단 간의 이동성 현황을 파악하는 지표

#### 분석 대상 데이터(예시)

(공공데이터)

예술인등록DB(예술인복지재단), 문화예술인지원내역(한국문화예술위원회, 국가문화예술지원시스템)

- 소득데이터(근로소득, 사업소득 등, 국세청)
- 고용, 건강, 주거, 인구통계학적 정보(관련부처)

#### 분석 설명 및 방법

1. 문화예술인 소득계층별 고용수준/주거수준/건강수준 등의 현황을 파악하고 상대적 격차 파악
2. 예술인등록DB를 기준 자료로 활용하여 예술인지원내역, 민원, 관계부처별 DB를 공유·연계
3. 문화예술인 삶의 주요 영역에서 수준 및 격차를 보여주는 지표와 국가의 정책적 개입정도와 효과를 보여주는 대응지표로 구성

#### 예술인 × 영역별 교차분석

구분		소득·자산	고용	예술분야	교육	주거	건강·위험
지역별	강서구	-	-		-	-	-
성별	남성	-	-		-	-	-
연령별	40대	3분위	비정규직	문학	대졸	임차	-
가구특성별	1인가구	-	-		-	-	-

#### 기대효과

예술인들의 사회경제적 배경 또는 집단 특성 등 행정정보·데이터들을 연계·교차하여 지표 관련 다층적·심층적 상황 파악 및 분석에 기여

## 2) 특화된 대책 및 맞춤형 서비스

	분야별 공동활용 대상 데이터	데이터(예시)
주요	특정 계층 · 분야 관련 정보를 함유한 데이터	긴급 고용안정지원금 수급자, 소득 정보, 실업급여 및 고용보험 정보, 생활안정자금 신청 정보 등
부가	대상의 비교 및 분석을 위한 부가적 데이터	인구 통계학적 정보, 지역 정보, 취업 현황, 문화/예술/공연 등의 과거 종사 업종 정보 등

### 《 분석 과제(안) 》

#### 분석 주제

문화관광 분야별 프리랜서 / 특수 고용자들의 코로나 고용 불안정 현황 분석

#### 배경 및 목적

- 코로나로 인한 고용 불안정으로 어려움을 겪고 있는 문화계 프리랜서에 대한 현황 분석
- 공연/관광/체육 등 각 업종별 차별화된 POST-COVID 고용 안정화 지원 대책의 기초 마련

#### 분석 대상 데이터(예시)

##### (공공데이터)

- 정부/지자체 코로나 지원금 수급 내역
- 실업 급여 및 고용 보험 정보

##### (민간데이터)

- 여가 분야별 지출액 추이 (코로나 이후)
- 각 지역별 문화관광관련 업체 개폐업 정보

#### 〈직종별 지원금 신청자 현황〉

〈직종별 신청자 수〉

직종	신청자 수(명)	비율(%)	직종	신청자 수(명)	비율(%)
가사	141,936	24.5	마케팅/홍보/광고	6,852	1.2
보험설매자	164,526	17.8	경영기획/경영지원	4,296	0.7
교육원/전통시장	101,267	17.6	대외협력/외무	4,096	0.7
서비스관리/지원	38,581	6.6	정보통신/가사	3,466	0.6
판매원/전통시장	25,967	4.7	정보통신/경영지원	2,963	0.5
방송기자	22,744	3.9	정보통신/경영지원	2,451	0.4
내려문서/가사	22,581	3.8	대외협력/외무	2,090	0.3
방송기자	22,511	3.8	대외협력/외무	1,507	0.3
방송기자/경영지원	19,253	3.3	가정/정보통신/가사	1,234	0.2
방송기자	11,243	1.9	정보통신/가사	633	0.1
방송기자/전통시장	8,996	1.5	정보통신/가사	596,892	100

\* 산재보험 적용 14개 특고 직종은 파란색으로 표시

#### 분석 설명 및 방법

1. 문화계 종사 비정규직 / 프리랜서 종사자들의 분야별 정부/지자체 지원금 수급 현황 분석
2. 분야별로 고용 피해 정도 비교 분석  
(ex. 관광 > 실내공연/관람 > 실외 스포츠)
3. 국민의 여가별 지출액 추이 및 세부 업종별 개폐업 정보를 활용한 세부 업종별 회복력 지수 개발

#### 기대효과

세부 직종별 코로나 피해 정도에 따른 차별화된 재취업 정책 수립 및 지원책 마련

#### 〈세부 업종별 개폐업 추이〉

##### 비디오·영화·음악 업체 수 순증\* 추이

(2019~2020년 2~4월)



	분야별 공동활용 대상 데이터	데이터(예시)
주요	<ul style="list-style-type: none"> <li>성(Gender)·연령별 소비 품목 및 평균 지출 금액 데이터</li> </ul>	<ul style="list-style-type: none"> <li>성·연령대별 인기 순 소비 품목 데이터</li> <li>품목별 평균 소비 금액</li> <li>연령대 별 이동에 따른 소비 데이터</li> </ul>
부가	<ul style="list-style-type: none"> <li>지역별 동종 업종별 매장 데이터</li> </ul>	<ul style="list-style-type: none"> <li>업종별 매출 데이터</li> <li>업종별 매장 수 데이터</li> </ul>

### 《 분석 과제(안) 》

#### 분석 주제

지역경제 활성화를 위한 맞춤형 타겟 마케팅 분석

#### 배경 및 목적

- '20년 관광공사는 국내 여행트렌드가 개인 맞춤형으로 변화
- 외부 관광객을 대상으로 한 소비 데이터를 분석하여 업종별 주요 고객층 정보 구축

#### 분석 대상 데이터(예시)

- (공공데이터 - 정형)
- 지역별 업종별 매장 정보
- (민간데이터 - 정형)
- 최근 2년간 소비 매출 데이터

#### 분석 설명 및 방법

- 성·연령대별 주요 소비 품목 및 평균 금액과 업종명, 매장 수 등의 포함한 상관관계 분석
- 성·연령대별 주요 소비 키워드 통계화
  - 소비패턴 변화 분석
  - 분기별 업종별 이탈/잔존/잠재적 고객 예측 모델링

#### 기대효과

- 이탈 예측 타겟 대상 마케팅 수행으로 잔존율 증대
- 소비 트렌드 분석을 통해 장기 마케팅 전략 수립
- 지속적인 고객 확보를 통한 안정적인 시장 환경 구성

	분야별 공동활용 대상 데이터	데이터(예시)
주요	문화체육관광분야 고용 관련 정보를 함유한 데이터	문화체육관광 종사자 고용이력, 임금, 구직활동(고용센터, 직업상담, 콜센터, SNS, 고용뉴스, 기타 이용현황 등
부가	대상의 비교 및 분석을 위한 부가적 데이터	종사자 동향분석, 노동시장 변화, 고용동향 및 연결망 분석, 직업 트렌드, 지역일자리사업 효과 등

## 《 분석 과제(안) 》

분석 주제

1. 고용보험 자료를 활용한 문화체육관광 종사자 동향분석  
2. 문화체육관광 고용관련 비정형데이터 분석  
3. 코로나19로 인한 문화체육관광 고용변화

배경 및 목적

문화체육관광 특수 종사자, 지역일자리, 뉴딜 사업 등 일자리 변화, 동향 분석 필요  
고용부와 문제부 연계 “청년디지털 일자리 사업” 등 고용과 문화의 연계한 사회현상 분석 필요

분석 대상 데이터(예시)

(공공데이터)  
- 문제부 데이터 연계( 예술인활동증명DB, 예술분야 고용보험DB, 사업체노동력조사, 예술인 소득정보DB)  
- 고용센터 콜 상담자료(문화체육관광 관련)  
- 워크넷 구인정보, 입사지원, 직업상담자료  
- 고용보험DB

분석 설명 및 방법

1. 고용보험 자료를 활용한 문화체육관광 종사자 동향 분석  
- 문화체육관광 업종, 직종별 분류  
- 문제부 데이터 연계( 예술인활동증명DB, 예술분야 고용보험DB, 사업체노동력조사, 예술인 소득정보DB)  
- 노동시장의 변화(기존 문화체육관광 종사자 이직현황, 고용형태의 변화, 임금관련 소득)  
- 워크넷 유효구직자 현황, 이직활동 동향, 고용보험 및 취업 실태, 위기 극복과 트렌드 변화  
- 문화체육관광분야 종사자 고용 동향 및 연결망 분석  
2. 문화체육관광 고용관련 비정형데이터 분석  
- 문화체육관광 종사자 직업심리검사, 고용센터 콜센터, 취업관련 뉴스, SNS 등 비정형 데이터 분석  
- 비정형 분석(토픽모델링, 온톨로지 활용)  
- 문화체육관광 관련 직업 트렌드  
3. 코로나19로 인한 문화체육관광 고용변화  
- 문화체육관광 특수 종사자, 지역일자리, 뉴딜 사업 등 대상 분류에 따른 일자리 변화, 동향 분석  
- 특히, 코로나로 인한 심각한 타격을 받고 있는 관광, 여행, 숙박, 기타 등 타 부처의 플랫폼 노동자가 있다면 문화체육관광의 특수직을 분류하여 고용관련 동향 분석

기대효과

문화체육관광 종사자 동향 분석  
문화체육관광 관련 직업 트렌드 분석  
코로나19로 인한 문화체육관광 고용변화

〈고용 동향분석〉

지역별 산업구조와 고용구조의 특징

- 고용위기지역의 특징
- 지역별 산업구조 및 고용구조 검토

고용행정DB를 활용한 분석

〈고용 사회연결망 분석〉

〈고용 비정형데이터 분석〉

	분야별 공동활용 대상 데이터	데이터(예시)
주요	특정 지역 관련 정보를 함유한 데이터	문화체육관광부의 관광특구 지정자료
부가	대상 지역의 현황을 파악하기 위한 기초 데이터	행정구역별 사업체현황, 종사자수 등의 사업체 자료

## 《 분석 과제(안) 》

### 분석 주제

33개 전국의 관광특구 산업현황

### 배경 및 목적

관광특구에서 산업활동을 하는 사업체의 산업현황을 파악하여 지원정책 수립에 활용

### 분석 대상 데이터(예시)

(문화체육관광부)

- 관광특구 지정현황 자료

(통계청)

- 전국사업체조사

- 통계지리정보서비스(SGIS)

### 분석 설명 및 방법

1. 문화계 종사 비정규직 / 프리랜서 종사자들의 분야별 정부/지자체 지원금 수급 현황 분석
2. 분야별로 고용 피해 정도 비교 분석  
(ex. 관광 > 실내공연/관람 > 실외 스포츠)
3. 국민의 여가별 지출액 추이 및 세부 업종별 개폐업 정보를 활용한 세부 업종별 회복력 지수 개발

### 기대효과

관광특구의 기초통계자료 파악

### 〈직종별 지원금 신청자 현황〉

〈직종별 신청자 수〉

직종	신청자 수(명)	비율(%)	직종	신청자 수(명)	비율(%)
기타	141,836	24.5	대중문화행사인	6,852	1.2
보통일용직	164,526	17.8	건물기계조작직	4,206	0.7
고급전문직종사	161,267	17.6	대중문화행사인	4,096	0.7
서비스관련종사	38,581	6.6	특수기능직	3,466	0.6
판매관련종사	23,987	4.1	산업보안직종사	2,961	0.5
특수기능직	22,144	3.9	산업보안직종사	2,851	0.4
대리운전기사	22,581	3.8	식품안전직	2,070	0.3
문화체육관광	22,511	3.8	체육지도자	1,507	0.3
문화체육관광	19,253	3.3	가정서비스직종사	1,234	0.2
영주교사	11,243	1.9	문화보존직	633	0.1
문화체육관광	8,996	1.5	문화체육관광	586,892	100

\* 산재보험 적용 14개 특고 직종은 파란색으로 표시

### 〈세부 업종별 개폐업 추이〉

#### 비디오·영화·음악 업체 수 순증 추이

(2017년~2020년 2~4월)



	분야별 공동활용 대상 데이터	데이터(예시)
주요	특정 계층·지역·분야 관련 정보를 함유한 데이터	국민체력100센터 측정DB
부가	대상의 비교 및 분석을 위한 부가적 데이터	국민건강보험, 통계청(사망기록), 건강보험심사평가원 국립암센터 빅데이터 등

### 《 분석 과제(안) 》

**분석 주제**

국민체력 빅데이터 기반 선제적 맞춤 건강유지 정책 마련

**배경 및 목적**

- 바쁜 사회, 환경적 요인으로 인한 개인건강을 위한 체력유지 미비
- 의학의 발전으로 기대수명은 늘어나지만 건강수명은 기대에 미치지 못함

**분석 대상 데이터(예시)**

(공공데이터)

- 국민체력100센터 측정DB
- 국민건강보험 빅데이터
- 건강보험심사평가원 빅데이터
- 국립암센터 암환자 빅데이터
- 통계청(사망기록)

**분석 설명 및 방법**

국민체력100센터 DB를 바탕으로 체력등급과 국민건강보험 등 다양한 의료빅데이터를 연계하여 체력등급 및 유지가 개인의 건강과 어떠한 연관성이 있는지를 분석


(예시)

- 체력등급별 만성질환 유병률 차이 비교
- 체력등급별 특정암의 유병률 비교 등등
- 체력정도가 개인의 건강에 미치는 다양한 요인 분석


**기대효과**

기대수명과 더불어 건강수명 증대를 통한 개인 삶의 질 증대 & 국가의 사회적 비용(건보재정) 절감

〈국민건강보험 데이터 분석(자살 사망자)〉



〈통계청 사망기록 분석〉





### 3) 위험요소 예측 제거

	분야별 공동활용 대상 데이터	데이터(예시)
주요	국민의 안전과 유관하고, 예측에 활용할 수 있는 데이터	카지노 이용 고객 DB, 소득 정보, 개인 파산 신청 이력 정보 등
부가	예측 효과 개선을 위해 함께 활용할 수 있는 데이터	개인 신용도 정보, 카드 소비 패턴 정보, 금융 활동 정보 등

#### 《 분석 과제(안) 》

##### 분석 주제

개인의 여가 활동 기반의 개인 파산 고위험 고객군 사전 예측

##### 배경 및 목적

- 최근 가계 대출 증가로 인해 전반적인 재무 건전성 및 리스크 관리 중요성 대두
- 개인의 선호 문화 / 여가 활동에 기반한 개인 파산 가능 고위험군 예측 및 집중 관리

##### 분석 대상 데이터(예시)

- (공공데이터)
- 내국인의 카지노 이용 DB
  - 개인 파산 신청 이용 고객
- (민간데이터)
- 신용평가사 신용 평점 데이터
  - 금융 (DTI 등) 활동 상세 정보

##### 분석 설명 및 방법

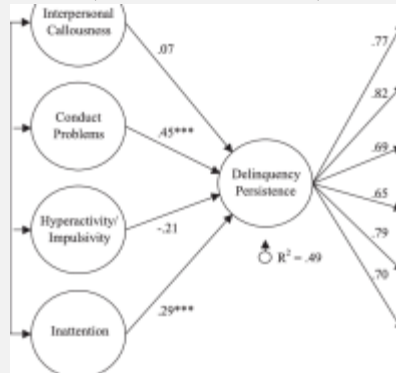
개인정보 비식별화를 통해 카지노를 포함한 gambling 여가 (경마, 경륜 등) 의 이용 고객에 대한 정보를 민간 신용 평가사의 데이터와 결합하여 개인의 신용도 혹은 개인 파산 위험 고객군을 사전에 예측할 수 있는 지수 개발

- 민간 데이터와 여가 활동 정보를 개별 변수화 하여 개인 파산 예측 모형 개발 (ex. imbalanced data 처리 기법 및 XG Boost 기법 활용)

##### 기대효과

개인의 재무 건전성 및 리스크 관리 활용 가능

##### 〈개인 파산 예측 모델 예시〉



##### 〈국내 신용평가사와의 비식별화 결합 (예시)〉



	분야별 공동활용 대상 데이터	데이터(예시)
주요	공연, 전시, 관람 등 예매 판매기간 및 환불 기간에 대한 정보	예매율(지역별, 성별 등), 취소율, 영역별 예매율, 취소율 등
부가	지역별 공연, 전시, 영화 정보, 마케팅 정보 등	공연, 전시 관련 sns분석 데이터

### 《 분석 과제(안) 》

#### 분석 주제

공연, 전시 데이터 분석을 통한 취약점 도출 및 대응 전략 마련

#### 배경 및 목적

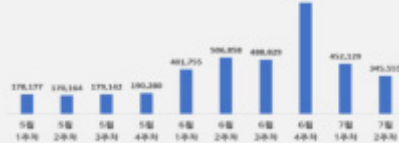
- 공연 예술 분야 예약, 판매, 공연 등의 데이터 분석을 통한 취약 부분 선정 및 원인 분석
- 취약 부분 지원 및 활성화 방안 도출을 통한 개선방안 도출

#### 분석 대상 데이터(예시)

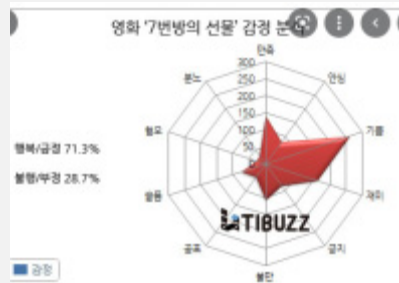
(정형·비정형 데이터)

- 최근 5년간 공연, 전시, 관람 데이터 분석
- 관련 공연, 전시, 관람관련 sns분석(뉴스, instagram, Facebook 등)

#### 〈주말 예매율 추이〉



#### 〈감성분석〉



#### 분석 설명 및 방법

- 기초자료 분석(정량적데이터): 취약 분야 도출
- 요인분석(공연, 전시, 관람 등 국민 참여 요인 분석)
- 취약점 보완에 따른 분야별 활성화 대책 마련

#### 기대효과

- 기간에 따른 취약 부분 도출을 통한 활성화 대책 마련
- 취약점 분석을 통한 취약 요인 방지

#### 4) 미래수요에 대한 선제적 대응

	분야별 공동활용 대상 데이터	데이터(예시)
주요	해외 관광 수요 예측이나, 대책 수립에 활용할 수 있는 데이터	출입국 통계, 외국인 국내 관광 실태 조사, 주요 국가별 코로나 백신 접종 현황 등
부가	해외 관광 수요 대응과 직·간접적으로 관련이 있는 데이터	주요 출신 국가별 국내 카드 소비 데이터, 외국인의 자국 외 국가의 관광 소비 패턴 등

#### 《 분석 과제(안) 》

##### 분석 주제

주요 국가별 외국인의 국내 관광 수요 예측

##### 배경 및 목적

- 전세계적으로 코로나 백신 접종을 증가에 따라 Return to Travel 에 대한 관심도 증가
- 주요 출신 국가별 외국인의 국내 수요를 사전에 예측함으로써 맞춤형 관광 프로그램 개발

##### 분석 대상 데이터(예시)

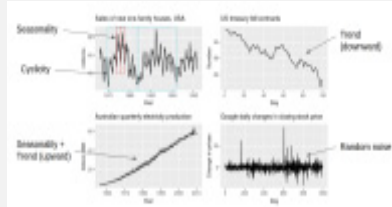
(공공데이터)

- 국가별 출입국 통계
- 외국인 국내 관광 실태 조사

(민간데이터)

- 외국인의 국내 소비 데이터
- 외국인의 자국 외 관광 관련 소비 데이터

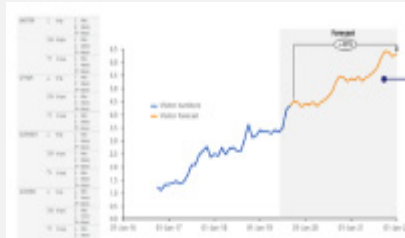
##### 〈머신러닝을 이용한 시계열 예측 모델 예시〉



##### 분석 설명 및 방법

주요 출신 국가별 입국자수 통계 자료와 외국인의 국내 관광 관련 업종의 소비 관련 데이터, 국가별 백신접종을 추이, 외국인의 자국 외 기타 국가에서의 관광 관련 소비 데이터를 활용하여 시계열 예측 모델 기반의 외국인 관광객 예측  
- ARIMA 기법을 활용한 시계열 예측 분석

##### 〈출신 국가별 월별 관광객 규모 예측〉



##### 기대효과

고수요 예측 국가의 외국인을 대상으로 한 맞춤형 관광 코스 개발 지원 및 국가별 관광 홍보 차별화 가능

## 5) 행정업무 경제성·효율성 개선

	분야별 공동활용 대상 데이터	데이터(예시)
주요	행정서비스 대상 데이터 및 개선 대상 데이터	전국 지자체 단위 둘레길 코스 정보, 둘레길 완주 인증 고객 현황, 둘레길과 관련한 민원 정보 등
부가	행정업무 내용 및 행정서비스 대상과 직·간접적으로 관련이 있는 데이터	유동인구 정보, 대중교통 이용정보, 둘레길 코스 주변 지역의 업종별 소비 규모 정보 등

### 《 분석 과제(안) 》

#### 분석 주제

지자체 단위 둘레길 코스 및 활성화 아이디어 발굴을 위한 이용객 분석

#### 배경 및 목적

- 최근 국민 여가 활동의 트렌드가 실내형 → 아웃도어로 변화하고 있음
- 각 지자체가 운영하고 있는 둘레길 (트레킹) 의 이용 현황 분석을 통해 개선안 도출

#### 분석 대상 데이터(예시)

##### (공공데이터)

- 둘레길 코스 현황
- 지자체별 완주자 증감 현황
- 둘레길 관련 민원 키워드

##### (민간데이터)

- 유동 인구 데이터
- 코스 인근 지역 대중교통 이용 정보

#### 〈유동인구 및 대중교통 분포도〉



#### 분석 설명 및 방법

지역별 둘레길 코스의 난이도 \* 접근성에 따른 이용객 추이 비교 분석

둘레길과 관련된 주요 민원 키워드 분석에 기반한 개선 포인트 마련 (ex. 완주 인증 불편, 종이형 지도와 스탬프 휴대 불편 등)

#### 〈관련 키워드 분석〉



#### 기대효과

- 효과적인 둘레길 보완/정비 계획 수립 가능
- 코스 재정비, 디지털 인증 시스템 구축 등

	분야별 공동활용 대상 데이터	데이터(예시)
주요	지역별(지자체) 체육시설 현황 인구 통계 정보	체육시설 리스트(야구장, 수영장, 테니스장, 축구장 등) 소재지 및 운영기관(연락처 포함) 지역별 주거시설 현황 및 연령별 인구
부가	지역별(지자체) 체육시설 이용현황 체육시설에 대한 SNS 내용	주간, 월간 체육시설 이용 현황(비용 포함) 체육장비 구비 현황 SNS 언급 내용

### 《 분석 과제(안) 》

<div> <div>분석 주제</div> <div>체육시설별 이용 현황 분석을 통한 체육장비 보급 및 추가 시설 위치 선정</div> </div>
<div> <div>배경 및 목적</div> <div> <ul style="list-style-type: none"> <li>- 건강에 대한 중요성으로 체육시설을 활용한 운동 인구 증가</li> <li>- 지역별로 체육시설 편중, 장비 부족에 따른 주민의 불만 증가</li> </ul> </div> </div>
<div> <div>분석 대상 데이터(예시)</div> <div> (공공데이터 - 정형) <ul style="list-style-type: none"> <li>- 지자체별 체육시설 현황</li> <li>- 체육시설별 이용객 현황 및 장비 보유 현황</li> </ul> (민간데이터 - 비정형) <ul style="list-style-type: none"> <li>- SNS에 언급된 체육시설 현황 및 필요사항 등</li> </ul> </div> </div>
<div> <div>분석 설명 및 방법</div> <div> 지역별 체육시설 현황과 보유 장비 및 이용객 현황 등을 분석 <ul style="list-style-type: none"> <li>- 지역별 체육시설 현황 분석</li> <li>- 체육시설 이용 변화 추이 분석(비용 포함)</li> <li>- SNS에 언급된 긍·부정 내용, 필요사항 등 분석을 통한 선호도, 추가 시설 요청 사항 분석</li> </ul> </div> </div>
<div> <div>기대효과</div> <div> <ul style="list-style-type: none"> <li>- 체육시설 이용 현황 패턴 분석을 통한 지역별 체육시설 목표 및 방향성 재정립</li> <li>- 지역별 추가 체육시설 마련을 위한 근거 마련</li> </ul> </div> </div>

	분야별 공동활용 대상 데이터	데이터(예시)
주요	지역(지자체)별 관광지 SNS 화제어	화제어 언급 키워드 언급 빈도수
부가	주요 관광지 현황 지역(지자체)별 음식점(모범) 현황	내외국인별 입장객수, 관광지명 음식점 주소 및 업소명 업태 및 주메뉴, 전화번호 등

### 《 분석 과제(안) 》

<div> <div>분석 주제</div> <div>지역별 주요관광지에 위치한 음식점(모범)들의 SNS 화제어를 파악하여 홍보 방안 제시</div> </div>
<div> <div>배경 및 목적</div> <div> <ul style="list-style-type: none"> <li>- COVID-19 확산으로 인해 해외여행이 어려워 국내여행으로 대체됨</li> <li>- 지자체는 이러한 국내여행객을 유치하기 위해 사활을 걸고 있는 상황</li> </ul> </div> </div>
<div> <div>분석 대상 데이터(예시)</div> <div> <p>(공공데이터 - 정형)</p> <ul style="list-style-type: none"> <li>- 음식점(모범)업 주요 현황</li> <li>- 주요 관광지별 입장객(월별, 주간별) 현황</li> </ul> <p>(민간데이터 - 비정형)</p> <ul style="list-style-type: none"> <li>- SNS에 노출된 관광지별 음식점 등 화제어 현황</li> </ul> </div> </div>
<div> <div>분석 설명 및 방법</div> <div> <p>지역별 주요 관광지와 관광지에 위치한 주요 음식점의 현황 등을 분석</p> <ul style="list-style-type: none"> <li>- 지역별 음식점 별 현황 분석</li> <li>- 주요 관광지별 입장객 현황 및 SNS 화제어 등 노출 현황 분석</li> <li>- 지역별 음식점들에 대한 효과성 분석</li> </ul> </div> </div>
<div> <div>기대효과</div> <div> <ul style="list-style-type: none"> <li>- 지역별 주요관광지에 위치한 음식점을 파악하여 관광 활성화 방안에 적용</li> </ul> </div> </div>

### 3. 조사결과 활용상의 유의점

보유 데이터 조사결과만으로 문화체육관광부 소속 및 산하기관 전체 보유 데이터 현황을 대표함과 동시에 데이터별 특징을 유형화할 있는 결과로 활용하기에는 제약점이 있다. 본 조사에는 문화체육관광부를 제외한 조사대상 46개 기관 중 21개 기관만이 조사에 응답했기 때문에 조사에 참여하지 않은 나머지 25개 기관들이 외부에 개방하지 않은 데이터 정보가 없는 상황이다. 따라서 본 데이터 전수조사 결과만으로 문화체육관광 분야 데이터 소재 및 현황을 일반화하고 대표할 수 있는 결과로 한정 짓는 것은 현실과 다른 왜곡된 결과를 초래할 수 도 있을 것이다.

본 연구에서 수행한 보유 데이터 조사결과는 문화체육관광 분야 데이터 총괄 현황을 파악할 수 있는 참고자료 또는 각 기관이 고유 업무 수행을 통해 신규로 보유 또는 생산·관리 되고 있는 데이터의 분야, 특성, 유형 등을 판단할 수 있는 기초자료 수준으로 활용하는 것이 적절한 활용 방안이 될 것이다.

보유 데이터 전수조사와 행정안전부 공공데이터포털에서 개방하고 있는 문화체육관광부 소속 및 산하기관 데이터를 결합한 결과는 문화체육관광 분야 데이터 소재 및 현황을 총괄적으로 파악할 수 있는 객관적 기초자료로 활용될 수 있을 것으로 판단하고 있다. 현재까지 문화체육관광 분야 공공데이터를 행정 업무 단위로 체계화한 정부기능분류, 공동활용 정책분류, 문화체육관광부 조직분류 등으로 유형화한 연구 또는 데이터가 없는 실정이므로, 본 연구를 통해 결합된 데이터 총괄(전수+공공) 자료는 문화체육관광 분야 데이터 소재 및 현황을 파악할 수 있는 기초자료로 활용 될 수 있을 것이다.

문체부 소속 및 산하기관에서 법령 등에서 정하는 고유의 행정업무 수행을 통해 생성, 취득하여 운영되고 있는 데이터현황 파악을 목적으로 하는 데이터조사가 정확성 및 신뢰성을 담보하기 위해서는 무엇보다 데이터 제공·협조와 관련한 법·제도 정비가 우선적으로 검토되어야 한다. 이를 통해 기관 간 데이터 공동활용을 제약하는 장애요소 발굴 및 개선사항이 검토되어야 하며, 조사된 데이터 현황이 전수화될 수 있도록 모든 소속 및 산하기관이 참여할 수 있도록 하는 관련 법·제도에 대한 정비가 필요하다. 더불어, 조사된 결과가 균질화 및 표준화될 수 있도록 데이터 분류, 데이터 정의, 컬럼 정의, 데이터 활용 이력 등 공동활용 데이터에 대한 항목 정의 및 비정형 데이터(텍스트, 이미지 등)에 대한 지속적인 검토·보완이 필요할 것으로 보인다.





문화·관광·콘텐츠 데이터 공동활용 기반 구축 연구

제5장

데이터 공동활용 활성화  
방안 마련



## 제1절 데이터 관리 및 공동활용 기반 마련

---

데이터의 공동활용을 위해서는 필요한 데이터의 소재 파악과 해당 데이터를 연계·활용할 수 있는 기반이 마련되어 있어야 한다. 이는 곧 각 기관에서 보유하고 있는 데이터와 업무상 수요 데이터를 파악하고, 기관 간 보유하고 있는 데이터를 연계할 수 있는 시스템, 연계된 데이터를 조회하거나 분석할 수 있는 시스템 등을 구축해야하며, 전문 인력이 수집된 데이터를 분석하고 활용해야하는 것과 같다.

정부는 지난 2020년 6월 9일 제정된 데이터기반행정 활성화에 관한 법률(이하 데이터기반행정법) 상에 데이터와 메타데이터의 수집, 표준화, 데이터 연계 및 공동활용할 수 있는 데이터통합관리 플랫폼의 구축·운영, 데이터 관련 전문인력을 양성할 것을 명시하고 있다. 이에 따라 행정안전부는 범정부 차원에서 기관메타 시스템을 구축하도록 하였고, 개별 기관의 기관메타 시스템을 중앙메타와 연동함으로써 모든 데이터를 중앙으로 수집하기 위한 기반을 마련하였다. 이는 중앙으로 수집된 데이터를 분류하고 메타데이터 정보를 제공함으로써 범정부 데이터의 소재지를 찾고 연관 데이터 목록까지 확인할 수 있도록 ‘국가데이터맵’ 서비스를 제공함으로써 데이터기반행정을 실행하기 위함이다.

본 절에서는 이러한 데이터기반행정을 실현하기 위하여 선행되어야 하는 데이터 관리 및 기관 간 데이터를 공동활용하기 위한 기반 마련 방안에 대해 제도적·기술적 측면에서 서술한다.

### 1. 제도적 측면

#### 1.1. 법·제도 정비의 필요성

데이터기반행정법이 시행됨에 따라 데이터기반행정을 수행하기 위해서는 기관의 특

성에 맞추어 데이터의 수집·활용 계획의 수립여부, 적절성 등을 사전에 확인하는 제도 도입 등 관련 업무 규정이나 기타 제도 등의 정비 필요하다. 문화체육관광부는 데이터 기반행정법 제정을 근거하여 2020년 9월 25일 문화체육관광부 데이터 관리 규정을 제정하였다(문화체육관광부훈령 제430호).

〈표 5-1〉 문화체육관광부의 법·제도 정비

법령(훈령)	주요 내용	비고
문화체육관광부 데이터 관리 규정 (훈령 제430호)	<ul style="list-style-type: none"> <li>데이터총괄책임관 지정 및 운영, 정책분석팀 업무 명시 등 데이터 관리 업무 체계 정립</li> <li>효율적 관리 및 중복방지를 위한 데이터 예산 검토·조정사항</li> <li>데이터분석센터의 설치 및 운영, 빅데이터 활용, 가명정보 처리 등 데이터의 제공 및 관리체계 정립</li> </ul>	2020.9.25. 제정 2021년 개정 예정

문화체육관광부 데이터 관리 규정은 ‘데이터 기반 행정법’ 등 관련 법령의 내용을 반영하고, 데이터의 공동 활용에 대한 전반적인 사항의 정비를 위함을 목적으로 두고 있다. 규정에서는 매 3년마다 문화체육관광데이터 기본계획을 수립하되, 해당 기본계획 내에 문화체육관광데이터의 연계·분석·개방 및 공동활용에 대한 사항을 포함하도록 하고 있으며, 정책분석팀장의 역할 중 하나로 데이터와 관련하여 데이터관리부서간의 협조와 데이터의 공동활용이 활성화될 수 있도록 필요한 조치를 취하도록 하고 있다. 그리고 데이터 역량 강화를 위한 관련 교육을 실시하고, 전문인력 양성을 위한 시책추진, 민간 참여 경진대회 개최 등 데이터 활용의 활성화를 위한 교육 및 전문인력 양성에 대해 규정하고 있다. 또한 데이터의 활용과 이를 위한 역량 강화 외에도 데이터 총괄 전담 조직을 신설하는 등 데이터기반행정의 제도적 기반을 정립하고 있으며, 문화 분야 정책수립에 데이터 활용을 촉진하기 위하여 관련 규정의 개정 등을 추진할 예정이다.

하지만 그럼에도 불구하고 데이터의 공동활용을 위한 법적, 제도적 기반은 아직 부족한 현실이다. 본 연구에서 추진한 데이터 업무활용 실태조사의 결과를 보면, 타 기관의 데이터 이용 경험자 중 데이터 제공이 원활하지 않았다 응답한 사람에게 데이터 제공이 원활하지 않은 이유를 묻는 질문에 [그림 4-8]과 같이 ‘규정 없음(61.5%)’을 가장 많이 꼽고 있다. 이는 타 기관에 데이터를 제공할 때 역시 마찬가지로 내/외부 기관에 데이터 제공 시 원활하지 않은 이유로 ‘규정 없음(70.0%)’이 가장 높은 비율을 나타냈다([그림 4-12]).

데이터기반행정법은 강제성을 가지고 있지 않기 때문에, 데이터의 공동활용을 거부하거나 데이터기반행정을 하지 않아도 법적 처벌이나 벌금 등의 불이익은 존재하지 않는다. 또한 이상적 법률 조항과 현실적 구현 간의 괴리가 존재함을 부정할 수 없다. 예를 들어, 데이터기반행정법에서는 데이터기반행정 활성화 기본계획 수립 시 데이터의 연계·제공 및 공동활용에 필요한 체계를 구축하고 전문인력을 양성하는 내용을 포함하도록 되어 있으나 데이터기반행정을 위한 예산 또는 전문인력의 추가 지원은 부족한 실정으로, 데이터 공동활용 체계를 구축하기 위하여 기존 인력의 업무만 가중된 것과 다름없다. 유사한 예는 데이터 업무활용 실태조사의 결과에서도 나타나는데, 데이터기반행정법 제 24조(데이터 관련 전문인력 양성)에서는 전문인력 양성 교육훈련 프로그램을 개발하고 보급 지원하도록 명시되어 있으나, [그림 4-14]에서는 데이터 교육 경험이 있는 경우가 불과 절반이 되지 않는 40.1%로 나타났다.

이와 같이 데이터의 공동활용을 위해서는 현실적인 행정절차의 개선이 필요하며, 데이터 공동활용을 위한 세부적인 가이드라인 등의 제공도 빼놓을 수 없다. 현재는 외부에 데이터를 제공하기 위하여 어떠한 데이터를 어느 수준으로 어떻게 제공할지에 대한 세부적인 가이드라인 없이 여러 상위 기관으로부터 데이터 제공에 대한 요구만 있는 상황이다. 기관별로 보유하고 있는 데이터의 성격과 기관 내 데이터 담당자가 이해하고 있는 데이터의 수준이 모두 다름에도 불구하고 구체적인 가이드라인 없이 데이터 보유 현황에 대한 조사 및 제공에 대한 요청 등이 이루어지고 있기 때문에 이는 곧 담당자의 업무 가중으로 이어질 수밖에 없다. 따라서 데이터 공동활용을 위하여 필요한 구체적이고 세부적이 사항을 담은 가이드라인의 제공이 필요하다. 더불어 데이터 공동활용에 대한 업무 협조에 따른 적절한 보상체계도 함께 만들어져야 한다. 데이터 제공에 대한 업무는 전담/담당인력이 없는 상황에서 실무자에게 있어서는 부가적인 업무일 뿐 의무와 보상이 없는 업무에 해당하기 때문에 업무의 중요도 대비 적극성을 저하시키는 원인이 되기도 한다. 이러한 보상체계는 금전적 보상이 아니더라도 사기 진작을 시킬 수 있는 측면에서의 보상체계를 고려하여야 한다. 이와 같이 데이터의 공동활용 및 데이터기반행정을 제한하는 장애요소와 법적 제약사항 등은 도처에 산재하여 있다. 따라서 이러한 요소들을 지속적으로 발굴하여 꾸준히 정비하고 개선해나가야 한다.

또한 문화체육관광부의 소속 및 소관 공공기관 간 업무 협력이 원활하게 이루어질 수 있도록 업무협력이 원활하지 않은 원인을 파악하고, 데이터의 공동활용을 저해하는 장애

요소를 파악하여, 상위 기관으로서 하위 기관에 제도적 지원을 해야 하며, 문화체육관광부 내부적으로 데이터 공동활용에 대한 강제성을 가지는 규정을 추가하거나 제도 및 업무절차를 개선해야 할 필요가 있다.

## 1.2. 조직과 인력의 강화

데이터를 공동활용하고 나아가서는 데이터기반행정을 위해서는 데이터를 분석하고 활용할 수 있도록 데이터 관련 업무를 전담 또는 담당하는 조직의 기능을 강화하고 해당 조직 내에 데이터를 전문적으로 다룰 수 있는 데이터 전문 인력의 확충은 필수적이다. 이러한 데이터 전담 또는 담당 조직은 데이터를 수집하고 총괄 관리하며, 분석할 수 있어야 한다. 세부적으로는 기관 차원의 데이터 표준화 체계 수립 및 품질을 관리하며, 기관에서 생산하는 데이터를 총괄 관리하고, 타 기관과의 데이터 연계·공동활용의 업무뿐만 아니라, 경우에 따라서는 데이터 분석까지도 담당할 수 있어야 한다.

정부는 범정부 주요 현안 및 국가적 차원의 데이터 분석을 수행하고, 기관별 자체 분석센터와의 역할 정립 및 소통을 강화하기 위하여 범정부 데이터 분석을 지원하는 정부통합데이터분석센터 조직을 신설하였으며, 데이터기반행정 업무추진에 필요한 기관별 소요인력 등을 바탕으로 데이터 전문인력(데이터 직류 공무원 등)을 지원하고, 기관별 데이터기반행정 책임관(CDO)을 지정하여 기관 내 데이터기반행정을 촉진하고 대외 소통·협력 채널로 활용하기 위한 방안을 제시하였다(제1차 데이터기반행정 활성화 기본계획('21년~'23년), 2021.2., 관계부처 합동).

문화체육관광부는 데이터 총괄 전담조직인 정책분석팀을 신설하고 신규 인력을 5명 추가배치 하였으며, 한국문화예술위원회 내에 한국판뉴딜성과창출TF, 한국관광공사의 관광빅데이터팀, 한국문화정보원의 문화빅데이터부를 신설하였다. 또한 데이터기반행정법 제20조(데이터분석센터)와 문화체육관광부 데이터 관리 규정 제12조(데이터분석센터의 설치 및 운영)에 의거하여 한국문화관광연구원을 문화체육관광부 데이터분석센터로 지정하고(2020.12.10.), 문화·체육·관광 분야 공공기관의 문화데이터를 총괄(데이터 수집·연계·융복합 분석 등)하도록 운영 중이다. 이에 한국문화관광연구원은 데이터 생산·수집·결합·동향분석·심층 보고서 분석 등을 수행할 수 있도록 정책정보센터 내에 데이터분석팀을 신설하여 기관 내 조직을 정비하였으며, 전문인력을 충원하였다.

[그림 5-1] 정책정보센터 조직도



출처: 한국문화관광연구원 연구기획팀 내부 보고자료

하지만 이러한 조직 및 인력의 확충은 터무니없이 부족한 실정이다. 2021년 3월 기준, 문화체육관광부 소속·소관 공공기관의 데이터 관리 및 분석 담당 부서 운영 현황은 소속기관 1개(국립중앙박물관), 산하기관 6개로 총 7개 기관에서 데이터 업무 담당 부서가 운영되고 있다. 직제상 데이터가 명시된 부서명을 보유한 한국문화관광연구원, 한국문화정보원 및 한국언론진흥재단으로 확인된다(문화체육관광부 2021년도 데이터기반 행정 활성화 시행계획안 2021.4.). 이는 본부를 제외한 소속·소관 공공기관 48개 의 15%에 불과한 수준이다.

〈표 5-2〉 문화체육관광부 소속·산하기관 데이터 관리 및 분석을 위한 부서 운영 현황

기관명	실국 및 과명	현원(명) (전담+겸임)	전담 인력(명)	'21년 확보 인력(예정)	비 고
국립중앙박물관	디지털박물관과	-	-	4	
한국언론진흥재단	뉴스유통국 뉴스빅데이터팀, 미디어연구센터 총 2개 부서	21	3	1	
한국관광공사	관광디지털본부	-	-	-	신설('21.1.)
한국문화정보원	문화빅데이터부 데이터기획팀	5	5	2	신설('21.1.)
	문화빅데이터부 빅데이터팀	4	4	1	
한국문화 관광연구원	데이터분석팀	6	3	4	신설('21.3.)
	정책정보센터	8	6	0	
(재)국악방송	미디어사업부	1	0	1	
대한체육회	기획조정본부 정보화팀	5	0	1	
합계		50	21	14	

이러한 데이터 전담 또는 담당조직 및 인력의 부재로 인한 영향은 실무에서 나타난다. 문화체육관광부의 소속·소관 공공기관의 대부분은 기존의 인력이 데이터 제공 및 공동 활용에 대한 업무를 추가적으로 처리하고 있기 때문에 대내외적인 요구사항을 처리하기에는 무리가 있다. 조직 차원에서 데이터를 체계적으로 관리할 수 있는 전담 부서나 인력의 부재로 인해 기관 내 보유하고 있는 데이터가 관리되지 못하고, 데이터의 중복이나 부실 여부 또한 점검할 수가 없으며, 기관 내 각 부서에서 필요로 하는 데이터에 대한 수요조차 파악이 어렵다. 즉 기관에서 생산하는 많은 데이터를 활용하기 위한 체계가 구비되어 있지 못하며, 담당자가 변경 시에는 업무 연속성이 떨어지게 되어 동일 데이터를 수집함에도 불구하고 기존과 상이한 데이터가 수집되기도 한다. 조직 내 부서 간에도 업무 장벽이 존재하여 데이터에 대한 전수 조사 시 모든 데이터의 취합이 되지 않는 경우가 빈번하며, 데이터 관련 법이 시행될 때 필요한 데이터의 제공이 기관 차원이 아닌 개별 부서 차원에서 이루어지기도 한다.

따라서 문화체육관광부 소속·소관 공공기관의 데이터 공동활용을 위해서는 이러한 데이터 전담 또는 담당 조직이 없는 기관을 대상으로 관련 조직을 신설하고 역할을 명확히 해야 한다. 예를 들어, 정보화 부서의 경우 데이터를 데이터베이스화하여 개인의 PC가 아닌 중앙에서 기술적 관리를 위한 업무와 외부 데이터 요구에 대해 데이터를 기술적으로 연계하는 업무 등을 담당하고, 데이터 부서의 경우 기관 내 보유하고 있는 데이터 현황 관리 및 표준화 방안 수립 등의 데이터의 품질을 관리하며 데이터를 목적에 맞게 분석하는 업무를 담당하는 형태와 같이 조직의 기능을 강화해야 할 필요가 있다. 또한 이와 같이 조직의 역할에 맞게 데이터를 기술적으로 관리하고 품질을 관리하며 분석하기 위해서는 데이터 관련 전문인력의 확보는 불가피하기 때문에 원활한 데이터 공동활용을 위하여 조직과 인력 구성을 위한 상위기관의 적극적 지원이 절실하다.

### 1.3. 데이터 거버넌스 체계 마련

데이터 거버넌스는 데이터에 대한 정책 및 지침, 표준화, 운영조직 및 책임 등의 표준화된 관리체계를 수립하고 운영을 위한 프레임워크 및 저장소를 구축하는 등 데이터와 관련하여 전반적인 것을 아우른다. 데이터 거버넌스의 구성요소는 원칙(데이터를 유지 및 관리하기 위한 지침과 가이드)과 조직(데이터를 관리할 조직의 역할과 책임), 프로세



스(데이터 관리를 위한 활동과 체계)이다. 따라서 데이터 표준용어 설정, 명명 규칙 수립, 메타데이터 구축, 데이터 사전 구축, 데이터 생명주기 관리 등의 업무로 구성된 데이터 표준화, 데이터 정합성 및 활용의 효율성을 위해 표준데이터를 포함한 메타 데이터와 데이터 사전의 관리 원칙을 수립하며, 수립된 원칙에 근거하여 상세한 프로세스 생성 및 관리, 운영을 위한 담당자, 조직의 역할과 책임을 명시하는 데이터 관리체계, 메타데이터 및 표준데이터를 관리하기 위한 전사 차원의 저장소를 구성하는 데이터 저장소 관리, 표준 준수 여부를 주기적으로 점검하고 모니터링을 실시하며, 거버넌스의 조직 내 안정성 장착을 위한 지속적인 변화 관리 및 주기적 교육 등의 표준화 활동이 데이터 거버넌스 체계라 할 수 있다.

데이터의 공동활용을 위해서는 범정부 차원의 데이터 거버넌스 체계 수립과 지속적인 관리가 필요하다. 따라서 각 기관으로부터 수집된 데이터의 효율적·효과적인 활용을 위한 데이터의 특성 파악 및 표준화 실시, 데이터 품질 관리 등 데이터 거버넌스 체계에 따라 데이터를 관리할 필요가 있다. 특히 기관별로 데이터에 대한 기준 체계가 다르기 때문에 다른 기관의 데이터를 공동활용하기 위해서 선행되어야 하는 것은 데이터 간의 표준화이며, 이는 데이터 그 자체를 비롯하여 용어 등의 표준화를 모두 포함한다.

행정안전부는 이와 같은 데이터 거버넌스의 역할을 자처하며, 데이터의 중앙으로 수집, 데이터의 표준화, 기관별 메타 관리, 메타관리시스템을 통한 데이터 사전 및 국가데이터맵 구축 등 데이터에 대한 정책과 지침을 수립하고, 표준화 방안을 마련하여 제공하고 있으며, 데이터를 중앙에서 관리하는 등의 역할을 수행하고 있다.

문화체육관광부 역시 한국문화정보원을 통해 문화데이터를 수집하고 관리하며 데이터의 저장소 등 관리 체계를 갖추고 있지만, 보유 데이터가 문화 분야에 치중되어 있으며 소속·소관 공공기관의 전사적 데이터를 수집하는 체계가 부족한 상황이다. 따라서 문화체육관광부의 데이터 거버넌스 역할을 수행하기 위해 본부를 비롯하여 소속·소관 공공기관의 데이터를 유지·관리하기 위한 지침과 가이드를 수립하여 제공해야 할 필요가 있다. 여기에는 문화체육관광분야에 특화된 데이터 표준화 지침이 포함되어야 하며, 데이터를 관리하기 위한 본부와 소속·소관 공공기관의 역할을 명확히 하여 문화체육관광 데이터가 본부에 모두 수집될 수 있는 체계를 마련해야 한다.

## 1.4. 예산 확보

데이터의 공동활용을 활성화하기 위해 구축되는 인프라(조직, 인력, 시스템 등)의 운영 및 유지를 위해 예산의 확보는 필수적이다. 메타관리시스템을 비롯하여 데이터를 수집·관리하고, 공동활용 데이터를 등록하고, 수집된 데이터를 분석·활용하기 위한 시스템의 구축·운영뿐만 아니라 데이터 전문 인력의 확보 및 역량 강화, 데이터 전담 조직의 운영 등을 위한 예산은 기존에 기관별로 편성되었던 예산과는 별도로 추가 확보되어야 한다. 특히 기관의 규모가 작을수록 이러한 예산 확보는 더욱 중요하다.

기관의 데이터를 수집·관리·운용하기 위한 시스템을 구축·운영하며, 수집된 데이터를 활용하기 위한 전문인력을 확보하는 등의 일련의 작업은 모두 예산이 수반되어야 가능하다.

## 2. 기술적 측면

### 2.1. 데이터의 표준화 및 품질관리

#### 가. 데이터 표준화

정부는 「전자정부법」 제50조 및 같은 법 시행령 제59조를 통해 행정정보의 공동이용을 위한 행정코드 및 각종 표준·기준 등에 대한 표준화에 대한 법적 근거를 마련하였으며, 「공공데이터의 제공 및 이용활성화에 관한 법률」 제23조와 「공공기관의 데이터베이스 표준화 지침」에 따라 공공데이터를 누구나 같은 의미로 이해하고 같은 방식으로 사용할 수 있도록 공통표준용어를 정의하는 등 행정코드나 송수신체계, 용어 등의 표준화를 위한 방안을 제시하며, 정보시스템 구축 및 개선 시에 해당 표준을 따르도록 한다.

〈표 5-3〉 표준화의 법적 근거

법령명	본문
전자정부법	제50조(표준화) 중앙사무관장기관의 장은 국회규칙, 대법원규칙, 헌법재판소규칙, 중앙선거관리위원회규칙 및 대통령령으로 정하는 바에 따라 전자문서, 행정코드 및 행정기관등에서 공통적으로 사용되는 행정업무용 컴퓨터 등의 표준화를 위하여 필요한 조치를 할 수 있다.
전자정부법 시행령	제59조(표준화) ① 행정안전부장관은 법 제50조에 따라 다음 각 호의 사항에 대하여 표준을 정하여 관보에 고시할 수 있다. 1. 전자문서 및 전자화문서 2. 행정정보 공동이용에 필요한 각종 행정코드 및 행정업무용 정보시스템 간의 연계방식

법령명	본문
	3. 중앙행정기관등이 사용하는 행정업무용 소프트웨어의 규격 4. 중앙행정기관등이 공통적으로 사용하는 행정업무용 컴퓨터(단말기 및 주변기기를 포함한다)의 성능 및 이용기술 5. 그 밖에 사무처리 등의 효율화를 위하여 필요한 사항 ② 행정안전부장관은 제1항에 따른 표준 외에 행정정보의 공동이용을 위하여 필요한 각종 표준·기준, 그 밖의 세부 지침을 정할 수 있다. 다만, 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」 제8조제1항에 따라 과학기술정보통신부장관이 고시하는 사항은 제외한다. ③ 중앙행정기관등의 장은 제1항에 따라 제정된 표준을 준수하여야 하고, 행정안전부장관은 중앙행정기관등에 대하여 표준의 준수 여부 등 표준화 현황조사를 실시할 수 있다. ④ 행정안전부장관은 표준화를 촉진하기 위하여 제1항에 따른 표준을 준수하는 각종 상용제품을 선정하여 그 이용을 촉진하는 등 전자정부 표준화에 필요한 정책을 수립·시행할 수 있다.
공공데이터의 제공 및 이용활성화에 관한 법률	제23조(공공데이터의 표준화) ① 행정안전부장관은 과학기술정보통신부장관과 협의하여 공공데이터의 제공 및 이용을 활성화하고 효율적인 관리를 위하여 다음 각 호의 사항에 대한 표준을 제정·시행하여야 한다. 다만, 공공데이터 표준과 관련된 사항이 「산업표준화법」에 따른 한국산업표준으로 제정되어 있는 경우에는 그러하지 아니하다. 1. 공공데이터의 제공형태 및 제공기술 2. 공공데이터 제공을 위한 분류체계 3. 그 밖에 공공데이터의 제공 및 이용 활성화를 위하여 필요한 사항 ② 공공기관의 장은 제1항에 따른 표준을 준수하여야 하고, 행정안전부장관은 그 준수 여부를 조사하여 시정을 요구할 수 있다. 시정요구를 받은 공공기관의 장은 특별한 사유가 없는 한 이에 따라야 한다.

데이터 표준화는 데이터의 정보 요소의 명칭과 정의, 형식, 규칙 등에 대한 원칙을 수립하여 지속적으로 관리하기 위한 활동이다. 대표적인 데이터 표준화 활동으로는 코드 표준화와 도메인 표준화, 명명규칙 표준화 등을 예로 들 수 있다. 코드 표준화는 데이터가 가진 코드값을 표준화 하여 관리하는 것으로, 범정부차원의 ‘행정기관 코드표준화 추진지침(행정안전부고시 제2021-6호)’이 있으며, 해당 지침에서 제시되지 않은 코드에 대해서는 개별 기관에서는 기관 특성에 맞추어 코드를 정의할 필요가 있다. 도메인 표준화는 데이터의 항목(컬럼)이 갖는 공통적인 특성들(허용값의 범위, 데이터 타입, 길이 등)을 정의하는 것을 의미하며, 명명규칙 표준화는 테이블(데이터셋)의 유형(코드테이블, 마스터테이블, 이력테이블 등)에 따라 테이블명에 일정한 명명규칙을 정의하여 테이블명을 통해 테이블의 유형이나 성격 등을 용이하게 파악할 수 있다.

이렇듯 데이터의 품질을 보장하기 위한 기본 활동인 데이터 표준화는 개별 기관이 가진 데이터의 구조와 기준 등은 기관의 특성에 따라 구성되어 있기 때문에 이를 통합 관리하기 위해서는 중앙으로부터의 표준화 방안 수립이 필수적이다. 이를 위해 행안부에서는 「공공기관의 데이터베이스 표준화 지침(행정안전부고시 제2021-32호)」을 수립하여,

공통표준용어, 공통표준단어, 또한 공통표준도메인 등 모든 공공기관들의 업무에 공동 활용이 가능한 단어 및 용어의 표준을 정하였다.

하지만 개별 기관은 각자의 필요에 의해 데이터를 생성·수집하는 체계를 이미 구축하고 그에 따라 관리하고 있으며, 이는 데이터의 공동활용을 저해하는 요소이기도 하다.<sup>6)</sup> 또한 상황에 따라서는 기관 내에서도 통일되지 않은 채로 시스템 또는 데이터 별로 관리되기도 한다. 기관 차원의 표준화 방안이 부재하는 것은 자체적으로 표준화 방침을 수립하기에 인력과 예산이 부족하기 때문이다. 표준화 방침은 정보화 또는 데이터 담당 인력/조직이 수립해야하나, 기관의 특성에 따라 정보화 또는 데이터 부서가 별도로 존재하지 않거나 인력이 과소한 경우도 있어 기관 차원의 관리가 이루어지지 않는 것이다. 그밖에도 기관 간 데이터를 연계할 시에 데이터의 용어나 단위 등의 이해가 기관별, 담당자별로 상이하기 때문에 기관 간의 상이한 표준에 대해 매칭하기 위한 작업도 필요하다.

이와 같이 데이터의 표준화는 매우 중요하지만, 표준화를 추진할 때 주의해야하는 점이 있다. 이는 기 구축되어 있는 데이터 체계와 데이터베이스(DB)에 대해 표준화를 하는 것에 시간적·경제적인 어려움이 있다는 것이다. 단기적인 표준화 추진은 잘못된 표준화를 초래하거나 경제적 손실 등이 클 수 있으므로 지양해야 한다. 따라서 단기적인 관점에서의 표준화 방안과 장기적인 관점에서의 표준화 방안을 각각 수립할 필요가 있다. 단기적으로는 현재 가진 데이터 체계 및 데이터베이스를 재구축하기 보다는 중앙에서 데이터를 수집할 때, 표준화된 양식으로 가져오는 방식을 고려하는 것이 하나의 대안일 수 있으며, 시스템적 표준화 추진은 장기적으로 노후화로 인한 차세대 추진 또는 신규 시스템 구축 등을 통해 단계적으로 중앙 표준 기준을 적용하는 방식으로 진행할 필요가 있다. 단, 개별 기관의 데이터는 그대로 두고 중앙으로 수집 시에 표준화를 진행하는 경우, 데이터를 변환하면서 데이터 무결성 규칙을 준수하는지에 대한 검사를 통해 데이터 정합성이 보장되는지에 대한 검토가 필요하므로 변환 시에 유의할 필요가 있다.

6) 데이터 업무활용 실태조사 결과에 따르면 데이터의 공동활용을 저해하는 제약조건에 대해 응답자의 70.3%가 데이터 간 표준화가 미흡하다 답변하였다(그림 4-31)).

## 나. 메타데이터

메타데이터는 일명 ‘데이터의 데이터’로, 데이터가 가진 속성 정보를 구조화한 데이터를 일컫는다. 따라서 보유하고 있는 데이터(테이블) 목록과 위치, 데이터의 항목들이 가지고 있는 속성(명칭, 설명, 데이터 타입, 데이터 길이 등)에 대한 정보를 나타낸다. 이러한 메타데이터는 데이터의 구조와 속성, 특성, 이력 등의 정보를 포함하고 있기 때문에 체계적인 관리와 검색 및 활용이 가능하다.

행안부는 「범정부 데이터 플랫폼 구축 사업」을 통해 중앙행정기관을 비롯하여 지자체와 공공기관을 대상으로 기관메타시스템의 도입 및 중앙메타관리시스템과의 연계를 추진하였다. 이를 통하여 국가 차원에서 데이터를 수집하고, 연계체계를 구축하여 기관별 데이터 정보를 중앙으로 연계함으로써 통합관리가 가능한 데이터 수집 관리를 체계화하였다. 즉, 각 기관이 보유하고 있는 모든 데이터와 소재, 데이터의 속성 등을 중앙에서 파악할 수 있는 기반이 마련되었다. 또한 개별 기관에서는 공통표준용어나 공통표준단어, 공통표준도메인 등의 조회 및 관리가 용이해졌으며, 각 기관의 메타시스템에 등록된 메타데이터 간 연관도를 그래프 형태로 시각화하여 제공하기 위한 국가데이터맵 검색이 가능해졌다. 또한 상위 기관은 하위 기관의 메타정보 조회가 가능하며, 각 기관에서 운영 중인 정보시스템 및 DB의 관리가 가능하다.

이에 문체부에서는 데이터기반활성화 시행계획의 2021년 세부 추진 과제로 각 기관에서 운영 중인 정보시스템과 DB의 전수조사 및 메타데이터 관리 강화를 내세움으로써 메타데이터의 관리에 대한 중요성을 강조하고 있다.

이와 같이 메타관리시스템은 수립된 데이터 표준에 따라 데이터베이스에 적용하고, 데이터의 표준관리 체계(표준화 관련 규정과 절차, 담당자 등 포함)를 수립 및 운영할 수 있도록 지원해주지만, 일부 기관은 규모가 작거나 예산의 부족, 인력의 부재 등으로 인해 메타관리시스템을 구축하지 못하였으며, 설사 구축하였다 하더라도 표준 체계를 수립하여 운영하고 있지 못하는 실정이다. 이는 기관 내 조직과 전문인력의 부재, 예산의 부족으로 인한 것으로, 앞서 언급하였던 것처럼 예산을 확보하고 조직과 인력을 확충할 필요가 있다. 또한 표준화 측면에서 이미 기 구축된 DB에 대한 표준화는 장기적인 계획을 세워 추진하는 것이 바람직하다고 볼 수 있다. 따라서 조직과 인력, 예산 확보 이후에는 표준화된 메타데이터 관리체계를 구축하고, 이에 따라 순차적으로 단계별 개선이 필요하다.

## 다. 데이터 카탈로그

행안부는 중앙으로 수집된 데이터를 유형별로 관리하고, 데이터를 탐색하며, 데이터 간의 연관관계를 용이하게 파악할 수 있는 데이터 카탈로그를 제공함으로써, 데이터의 공동활용에 대한 편의성을 제공한다.

데이터 카탈로그는 조직 내에서 구성된 데이터 자산 목록으로 메타데이터를 사용하여 조직의 데이터 관리 작업을 지원한다. 이는 데이터 전문가가 데이터를 검색하고 데이터 거버넌스를 지원하는데 도움이 되는 메타데이터 관리 서비스이다.

데이터 카탈로그 서비스의 예로, 한국보건산업진흥원의 보건의료 빅데이터 플랫폼<sup>7)</sup>을 통해 각 기관에서 제공하는 데이터의 데이터셋과 변수 설명, 코드값 등에 대한 내용을 제공함으로써, 해당 카탈로그를 통해 연구자는 연구에 필요한 데이터를 확인하고 이 용신청할 수 있도록 서비스를 제공하고 있다.

행안부는 공동활용데이터 등록관리시스템을 통해 필요한 공동활용 데이터셋의 위치와 규모, 생성일 등 분석·활용을 위한 정보를 빠르고 쉽게 조회하고, 데이터를 유형별로 관리하고 데이터 탐색 및 데이터간의 연관관계 등을 쉽게 파악할 수 있도록 데이터 카탈로그를 제공하고 있다. 공동활용데이터 등록관리시스템(insight.go.kr)이란, 데이터의 공동활용을 지원하기 위하여 행안부에서 구축하고 서비스하는 시스템으로, 공동활용데이터 등록관리시스템을 통해 필요한 공동활용 데이터셋의 위치나 규모 등 활용하기 위한 정보를 조회할 수 있도록 하고 있으나, 아직까지 그 활용도가 높진 않다.

문화체육관광부 역시 데이터를 공동활용하기 위해서 소속 및 소관 공공기관이 가진 데이터를 목록화하여 제공하고, 데이터가 가진 항목과 속성 등의 메타 항목을 체계적으로 관리할 필요가 있다. 문화체육관광부가 보유한 데이터의 수집, 데이터의 특성 및 소재파악 등 체계적인 관리, 서비스 제공 등에 활용하기 위한 분류체계를 정의하고, 데이터의 유형과 특성에 따라 세분화함으로써 체계적인 관리가 가능하다. 문화체육관광부는 기관의 특성 상 여러 분야의 데이터를 다루기 때문에 이러한 분류체계의 정립은 무엇보다 중요하다.

7) 보건의료 빅데이터 플랫폼(<https://hcdl.mohw.go.kr>): 보건의료 분야 4개 기관(질병관리청, 국민건강보험공단, 건강보험심사평가원, 국립암센터)의 데이터를 개인 단위로 연계, 공공적 목적 연구에 활용할 수 있도록 연구자에 개방하는 시범사업을 지원하기 위한 정보시스템으로, 보건의료 빅데이터의 신청, 심의결과 확인, 연계, 비식별화, 개방(폐쇄환경 열람)의 기능 제공

한국문화정보원은 다부처 연계 플랫폼을 구축하여 문체부 소속·산하 기관의 콘텐츠 메타 정보를 관리하고 있다. 다부처 연계 플랫폼은 개별 콘텐츠에 대한 데이터를 나타내는 콘텐츠 메타정보를 관리함으로써 공동활용을 정보를 쉽게 확인할 수 있다. 콘텐츠 메타 정보는 해당 콘텐츠 데이터에 대한 정의, 데이터가 가진 항목(변수 또는 컬럼)에 대한 정의 외에도 물리적으로 해당 콘텐츠가 존재하는 방식을 함께 정의함으로써 데이터를 보유한 기관은 체계적인 관리가 가능하고, 해당 데이터를 이용하고자 하는 기관에서는 데이터에 대한 이해도가 향상되어 활용이 용이할 것이다. 하지만 문화체육관광부 소속·소관 공공기관의 적극적인 협조가 부족한 상황으로 문화체육관광부는 상위 기관으로써 적극적으로 협조할 수 있는 체계를 마련해야 할 필요가 있다.

## 2.2. 공동활용 데이터의 관리체계

### 가. 데이터 수집 체계의 정립

행정안전부와 문화체육관광부의 데이터 수집방법은 개별 기관이 가진 데이터를 그대로 등록하는 형태로 되어 있다. 개별 기관에서 모든 데이터를 등록하지 않거나 모든 항목을 등록하지 않는 경우에 대처하기 위한 방안을 마련할 필요가 있다. 따라서 데이터를 수집하는 체계에 대한 정립이 필요하다. 예를 들어 설문조사 데이터를 중앙으로 수집하는 경우, 필수 수집 항목에 대한 정의를 함으로써 데이터 등록 시에 해당 항목을 필수로 등록하게끔 진행을 하는 것이다. 데이터의 유형에 따라 필수 수집 항목을 정의하는 것만으로도 데이터를 연계하거나 분석·활용함에 있어 편의성이나 효율성이 높아질 것이다. 따라서 데이터의 유형을 파악하여, 필수 수집 항목 등에 대해 개별 부처(문체부)만의 표준을 정립하고, 이를 중앙(행안부)으로 취합하여 범정부적 수집표준이 정립될 것이다.

수집 표준에 대해 정립하는 것 외에도, 데이터를 수집하기 위한 기준을 수립해야 한다. 데이터 수집 시의 어려움은 공공기관의 구조적인 특성에서 비롯된다고 볼 수 있다. 국회 및 기재부, 행안부 또는 상위기관 등에서 동일한 데이터를 중복으로 요청하는 경우가 허다하다. 이러한 동일 데이터의 중복 요청은 데이터를 담당하는 업무 담당자들의 피로도를 높이고 업무를 과중하게 만드는 요인이라 할 수 있다. 많은 공공기관에서 데이터를 제공하는 업무를 전담하는 인력을 구성하고 있지 않기 때문에, 데이터 제공 요청에 대한 업무는 부가적인 업무라고 할 수 있다. 이에 일부 데이터 제공 업무 담당자들은

요청을 받을 때, 데이터 제공을 최소화하여 회신하는 경우도 존재한다. 따라서 중복 업무의 최소화를 위하여 수집할 데이터와 세부 항목들에 대한 기준도 함께 수립할 필요가 있다.

또한 시스템을 통해 구축되지 않는 많은 데이터가 기관 중앙에서 관리되지 못하고 해당 업무 담당자의 PC에 별도 저장되어 있는 경우가 많다. 즉, 생산되는 데이터의 수집 체계가 기관 자체적으로 구비되어 있지 않은 경우가 많다. 외부로부터 데이터에 대한 요청을 받을 시에도 기관 차원의 취합이 별도로 이루어지지 않고, 대다수의 기관에서는 개별 부서 수준에서 회신을 하고 있는 실정이다. 따라서 개별 기관에서 데이터에 대한 수집과 개방을 총괄하기 위한 방법과 관리체계도 수립해야 한다.

## 나. 공동활용 데이터 등록

공동활용 데이터는 다른 공공기관에서 정책을 수립하거나 의사결정에 활용할 수 있도록 제공하는 데이터를 말한다. 이는 민간에게 제공·서비스하는 공공데이터와는 달리 데이터를 공개하는 대상이 민간이 아닌 공공기관이라는 점에서 공공데이터 개방과는 차이가 있다.

행안부는 공동활용 데이터의 등록 유형을 기관이 자율적으로 공동활용 필요성을 판단하여 등록하는 자율등록과 행안부에서 지정하여 각 기관에서 등록하는 지정등록, 행안부가 인터넷에 공개된 데이터를 원활한 활용을 위해 등록하는 수집등록 이렇게 3가지 유형으로 분류하였다. 공동활용 데이터의 등록은 기관메타관리시스템을 통해 데이터 공동활용 설정 및 추가정보 입력을 통해 등록할 수 있고, 기관메타관리시스템에 등록되지 않은 데이터의 경우는 ‘공동활용데이터 등록관리시스템(insight.go.kr)’을 통해 직접 등록이 가능하다. 또한 공공부문에서의 민간데이터의 제공 및 연계·공동활용 방안을 마련하기 위하여 민간데이터의 구매·활용 지원시스템을 구축하고 운영방안을 수립하고 있으며, 민간데이터의 활용 촉진을 위한 법·제도 개선을 방안을 검토<sup>8)</sup>하는 등 데이터의 공동활용을 위한 방법을 적극적으로 모색하고 있다.

8) 제3자 단가계약을 통한 민간데이터 구매, 데이터 전문계약제도 신설 등 법제도 개선방안 검토(출처: 제1차 데이터기반행정 활성화 기본계획('21년~'23년), 2021.2., 관계부처 합동)



문체부 또한 기관 간 데이터 공동활용을 위하여 행정데이터를 비롯한 이미지·텍스트 등의 비정형 데이터의 공동활용 데이터 등록을 단계별로 추진할 계획을 수립하였다. 또한 기관에서 생산된 데이터 외에도 수요 예측 및 정책 관련 의사결정 등에 활용하기 위한 민간 데이터를 구매하여 활용하는 사업을 추진하고 있다.

이와 같이 데이터기반행정법이 발효된 이후 공동활용 데이터를 중앙에서 수집하고, 관리하기 위한 체계를 수립하고 관련 제도를 개선해나가며 시스템을 구축하는 등 다양한 작업이 이루어지고 있다. 행안부는 ‘데이터 공동활용 가이드라인’과 ‘데이터 역량강화 가이드라인’을 통해 행정기관 등 공공기관 간 데이터 공동활용에 필요한 세부절차 및 공공기관에서 필요한 데이터 역량의 자가진단 및 필요 역량을 확보할 수 있는 절차와 실행방법을 제시하고 있다. 하지만, 이러한 가이드라인의 홍보 및 제공이 부족하고, 상세한 방법에 대한 설명 부족 등으로 개별 기관이 어떠한 데이터를 공동활용하기 위한 데이터로 등록해야하는지, 등록 방법은 무엇인지 등에 대해 혼란을 가지는 경우가 많은 실정이다. 또한 데이터 수집 시의 사용하는 용어에 대해서도 기관별, 담당자별로 이해하는 바가 상이하며, 공동활용 데이터의 등록 건수가 기관의 실적으로 판단될 경우에는 자칫 데이터의 품질보다는 제공 건수에 초점이 맞추어져 데이터기반행정이라는 목적에 맞지 않는 상황이 발생할 수 있다. 실제 공공데이터 개방 시, 공공데이터 개방 건수를 기관의 실적으로 평가를 함에 따라 공공데이터 포털에 개방된 데이터는 실적을 채우기 위한 단순 리스트 형태의 파일이 대부분으로 데이터의 품질이 낮은 경우가 빈번하며, 이러한 점은 공공데이터 간의 연계나 활용 측면에서 한계가 발생한다. 이러한 것을 극복하기 위해서는 기관의 데이터 관리 책임자와 실무자를 대상으로 공동활용 데이터의 등록 및 방법, 데이터의 품질과 구조 등에 대한 보다 상세한 업무매뉴얼을 제시해야할 필요가 있으며, 데이터를 공동활용하고자 하는 목적과 공동활용의 결과물을 고려하여 공동활용 데이터로 등록하고, 공동활용 데이터의 건수를 기관의 실적으로 판단하는 것은 지양함으로써 무분별한 공동활용 데이터 등록을 방지해야 한다. 또한 부처 차원에서 정책 수립 또는 의사결정자의 수요를 파악하여 해당 부처의 소속 및 소관 공공기관 등을 대상으로 데이터를 수집한다면 공동활용 데이터의 등록이 좀 더 명확하게 이루어질 것이다.

그밖에도 공동활용 데이터 등록 시 유의해야할 사항은 해당 데이터에 보안 또는 개인 정보보호 관련 이슈사항이 존재하는지 파악하여 해당 정보를 제거하는 등 정보보안 및 개인정보보호에 주의하는 것이다. 데이터기반행정법 제3조(국가 등의 책무) 제5항은 데

이터기반행정을 수행하는 과정에서 개인정보 보호를 우선적으로 고려하여 대책을 마련하도록 하며, 제4조(다른 법률과의 관계)에서는 개인정보가 포함된 경우 「개인정보 보호법」에 따르도록 명시하고 있다. 다만, 개인정보 보호와 관련된 사회적 민감성으로 인해 데이터 제공에 한계가 존재하며, 많은 기관에서 마스킹을 통해 개인정보보호를 고려하여 데이터를 제공하고 있으나, 일부 기관에서는 데이터 제공을 거절하거나, 과도한 필터링 등으로 인하여 제공받은 데이터를 활용하지 못하는 문제도 발생한다. 따라서 개인정보를 보호하며 데이터를 공동활용할 수 있는 방안을 찾기 위한 노력이 지속되어야 한다.

#### 다. 데이터시스템 구축

문화체육관광부를 비롯하여 소속 및 소관 공공기관에서 관리하는 정보시스템은 600여개에 육박하며, 이의 구축 및 운영에 많은 예산이 투입되었음에도 불구하고 문제부처원의 전사적인 데이터 공유 체계가 부족한 실정이다. 이에 문체부에서는 문화 데이터의 수집·관리·분석을 일원화하고, 대국민 문화 데이터 수요에 맞는 플랫폼 등의 40개 시스템을 구축하고, 운영하기 위한 과제를 추진할 계획이다.

해당 과제를 통하여 문화·체육·관광 분야의 데이터를 제공하기 위한 기반을 구축할 수 있다. 데이터의 수집 및 제공, 특성 및 소재파악 등 데이터를 체계적으로 관리하고 서비스를 제공할 수 있는 메타데이터를 구축하고, 신규 시스템의 구축 시에는 시스템 간의 연계성 및 확장성을 고려하여 데이터베이스를 설계하고 구축할 수 있다. DB가 아닌 파일 데이터의 구축 시에도 다른 기관에 데이터를 제공하는 창구와의 연계성을 고려하여 데이터를 생성하고 관리하여야 한다.

이외에도 데이터시스템을 구축할 때 시스템을 통해 구축되지 않는 데이터를 시스템에 탑재하기 위한 방안을 마련하여야 한다. 기관차원에서 관리되지 못하고 업무 담당자 PC에 저장되어 있는 데이터를 기관 차원에서 수집해야 하며, 이를 위하여 기관의 중앙 관리 데이터 수집 시스템의 구축 및 관리 체계가 필요하다. 이에 따라 업무 담당자의 변경에 따른 데이터 소실 등이 발생하지 않을 것이다.

또한 중앙에서 데이터의 수집 및 연계의 현황을 파악할 수 있는 모니터링 시스템을 함께 구축해야 할 필요가 있다. 모니터링 시스템은 데이터 보유 기관의 데이터 갱신 주기, 갱신 일자, 현행화 여부 및 수요기관의 정보, 데이터 제공 시기, 대상 데이터, 제공방법

등의 제공현황 등을 관리·모니터링 함으로써 데이터의 최신성, 상호연계성, 정합성 등이 유지될 수 있도록 데이터를 현행화하고 운영할 수 있다.

정형 데이터 외에도 문화·관광·체육 분야의 행정데이터, 이미지·텍스트 등 비정형 데이터를 부 차원에서 중앙 수집·관리함으로써 부의 소속 및 소관 공공기관 뿐만 아니라 부처 간의 데이터 공동활용이 활성화 될 것이다. 이를 위해 기 구축·운영 중인 문화 빅데이터 플랫폼과 한국관광 데이터랩 등의 데이터 플랫폼 고도화와 연계가 필요하다.

공공데이터법 시행령에서는 제공 중인 공공데이터의 변경이 발생하는 경우 이용자의 안정적 이용을 보장하기 위하여 변경일부터 3개월까지 변경 전 공공데이터를 제공하도록 명시하고 있다. 데이터기반행정법에 따른 데이터의 공동활용에 있어서도 이와 같이 각 기관에서 보유하고 있는 데이터의 내용이 변경될 경우, 데이터 이용자의 안정적인 이용을 보장하기 위한 변경 전 데이터의 병행 제공하는 등의 서비스 제공이 필요하다.

#### 〈표 5-4〉 공공데이터의 제공 및 이용 활성화에 관한 법률 시행령 제16조

- 제16조(공공데이터 포털의 구축·관리 및 활용촉진 등)** ① 공공기관의 장은 법 제21조제1항에 따른 공공데이터 포털에 연계·제공하는 소관 공공데이터의 최신성, 정확성 및 상호연계성이 유지되도록 하여야 한다.
- ② 공공기관의 장은 법 제19조제4항 및 제21조제2항에 따라 공공데이터 포털에 연계·제공하는 소관 공공데이터의 내용을 변경하거나 그 제공을 중단하려는 경우에는 변경일 또는 중단일 3개월 전까지 행정안전부장관에게 통보하여야 한다. 다만, 공공데이터의 내용을 변경하는 경우에는 변경일부터 3개월까지는 변경 전 공공데이터를 제공하여야 한다.
- ③ 행정안전부장관은 공공데이터 포털에 이용자가 공공데이터의 등록 누락, 이용 불편사항 및 품질오류 등을 신고할 수 있도록 하며, 오류 시정 등 필요한 경우에는 해당 공공기관의 장에게 시정요구 등을 할 수 있다.
- ④ 행정안전부장관은 공공데이터 포털을 통하여 제공되는 공공데이터에 대하여 이용자 방문기록·제공현황 및 검색어 분석과 설문조사 등을 할 수 있다

## 라. 데이터허브 구축

데이터 허브는 여러 곳으로부터 수집된 데이터를 공유하는 것을 목적으로 만들어진 공간으로, 개별 기관에서는 다양한 분야의 데이터를 수집 및 결합·분석, 활용하기 위하여 데이터 허브를 통해 중앙에서 집결된 데이터를 활용할 수 있다. 행안부에서 운영하는 해안 포털이 그 대표적인 예로, 해안 포털은 공공부문의 빅데이터 분석 활용을 지원하기 위한 빅데이터 분석지원 플랫폼이다. 공동활용데이터 등록관리시스템을 이용하기 위해서는 해안포털 회원가입이 필수적이다. 중앙부처와 지자체 공무원, 공공기관 업무 담당

자들을 대상으로 공동활용데이터의 활용(조회, 요청) 기능을 제공하고 있다. 또한 중앙행정기관 기능분류체계(3단계 분류, 정책분야 15개 분야, 정책영역 65개 분야, 대기능 484개 분야로 분류)에 따라 데이터를 분류하여 제공한다.

문체부는 ‘문화데이터포털’을 통해 문체부를 비롯하여 소속 및 소관 공공기관이 보유하고 있는 문화, 체육, 관광 등의 다양한 문화분야의 데이터를 수집 관리하고, 오픈API 등의 형태를 통해 데이터를 제공하고 있다. 또한 한국관광 데이터랩을 통해 관광 전반에 걸친 데이터뿐만 아니라, 민간데이터(카드, 통신 데이터 등)를 확보하여 관광 빅데이터 분석 및 융합분석 등의 서비스를 제공하고 있다.

이와 같이 데이터 허브를 통한 데이터를 제공 시, 데이터 활용 기관 및 활용 목적 등의 정보를 관리함으로써 공동활용 데이터의 활용현황을 파악하며 정해진 목적 외의 용도로 이용하지 않도록 필요 조치를 취할 필요가 있다. 또한 협력 네트워크를 구성하여 협업을 유도하고, 분석기술 및 분석자원을 지원함으로써 데이터를 기반으로 하는 정책수립 및 의사결정이 가능해질 것이다.

## 제2절 데이터 공동활용 활성화

---

앞서 데이터 공동활용의 기반을 마련하기 위하여 제도적 측면에서 법과 제도의 정비, 조직과 인력의 확충, 데이터 거버넌스의 체계 마련, 예산 확보를 살펴보았고, 기술적 측면에서 데이터의 표준화와 품질관리, 데이터 수집체계의 정립, 공동활용 데이터 등록 등의 공동활용 데이터의 관리체계를 살펴보았다. 본 절에서는 데이터 공동활용의 활성화를 위한 데이터 활용역량 강화 측면과 데이터 활용문화 정착의 필요성 등에 대해 서술한다.

### 1. 데이터 활용역량 강화

#### 1.1. 데이터 역량강화를 위한 진단 및 평가 방안

##### 가. 데이터 역량강화 가이드

데이터기반행정법이 2020년 12월 10일 시행됨에 따라, 2021년은 해당 법을 적용한 관련 주요 사업을 시행하는 실질적인 첫 해이다.

데이터기반행정의 안정적이고, 전략적인 추진을 위해서는 관련 계획 및 시행계획이 수반되어야 함은 물론, 이를 토대로 주요 사항에 대한 추진 현황을 점검하는 방안 또한 필요하다. 무엇보다 “데이터”의 개념이 행정 및 정책 결정이라는 영역에 제도적으로 본격적으로 편입되었다. 그러나 현실적으로 공공기관의 기관별 특성에 따라 데이터 전문인력의 보유 여부 등은 기관별 편차가 클 것으로 예상된다. 이에 데이터기반행정의 안착과 실제 해당 제도를 통해 운영되는 공공기관의 데이터 역량 강화는 매우 중요한 사안이다.

행정안전부 디지털정부국은 지난 2021년 2월 19일, ‘제1회 데이터기반행정 활성화 위원회 회의’를 통해 ‘공공기관 데이터 역량강화 가이드라인’을 발표하였다. 해당 가이드라인은 미국 관리예산처(Office of Management and Budget, OMB)를 중심으로 수

립된 FDS(Federal Data Strategy)의 ‘데이터역량 가이드라인(Improving Agency Data Skills Playbook)’과 전문가 의견을 수렴하여 작성되었다.

공공기관 데이터 역량강화 가이드라인은 “공공기관이 데이터기반행정 활성화를 위해 필요한 데이터 역량을 측정하고 보강할 수 있는 절차와 기준”을 안내하기 위해 작성된 지침이다.

## 나. 데이터 역량강화 방향과 절차

데이터 역량강화에 대한 추진 근거는 데이터기반행정 활성화 및 이를 위한 행정적·기술적·재정적 조치 마련을 명시한 「데이터기반행정법」 제3조이다. 데이터 역량강화 추진의 적용대상은 중앙행정기관, 지자체 등 「데이터기반행정법」 제2조제3호 및 동법 시행령 제2조에 따른 공공기관이다.

데이터 역량강화 추진을 위해서는 다음과 같은 절차를 거쳐야 한다.

첫째, 데이터 역량지표체계 정립해야 한다. 데이터 역량지표는 미국 FDS 기준(데이터 라이프 사이클)으로 대분류(6개)를 구분하며, 데이터 역량에 따라 기획, 수집, 정제, 분석, 시각화, 정책활용으로의 선순환적 구조를 가지고 움직여야 한다.

〈표 5-5〉 데이터 라이프 사이클에 따른 데이터 역량(대분류)

데이터 역량	주요 내용	FDS의 데이터 라이프 사이클
1. 기획	기관의 데이터 전략을 수립하고 주요업무 관련 분석과제의 발굴·기획	
2. 수집	데이터 특성·구조를 이해하고 내외부 데이터의 연계·저장	
3. 정제	수집한 데이터를 분석 가능한 품질의 데이터로 변경·결합	
4. 분석	정책에 활용가능한 통찰력을 제공하도록 정제된 데이터를 분석	
5. 시각화	의사결정권자가 이해할 수 있는 수준으로 분석결과를 시각화	
6. 정책 활용	정책 의사결정에 분석결과를 활용	

\* 출처: 행정안전부(2021), 공공기관 데이터 역량강화 가이드라인

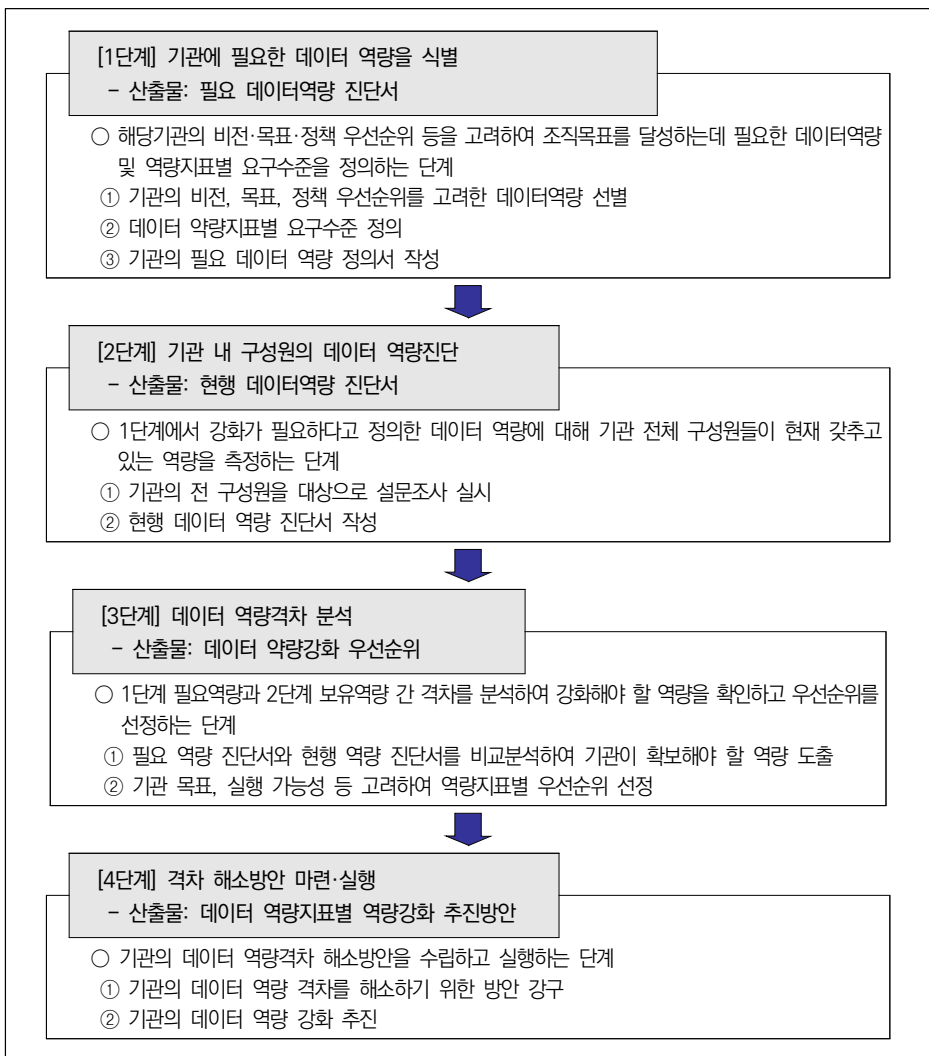
데이터 역량지표체계의 대분류는 중분류(15개), 소분류(43)개로 세분화하여 데이터 라이프 사이클에 따라 데이터역량을 측정하고 보완할 수 있도록 구성되어 있다.

〈표 5-6〉 데이터 역량지표 체계

대분류(6개)	중분류(15개)	소분류(43개)
1. 기획 2. 수집 3. 정제 4. 분석 5. 시각화 6. 정책 활용	① 데이터 전략 수립 ② 데이터 중심 조직관리 ⋮ ⑭ 분석 결과 해석 ⑮ 데이터기반 의사결정	① 데이터 전략 이해 ② 기관의 데이터 전략 수립 ⋮ ④ 의사결정에 분석결과 활용 ④ 정책반영 및 영향 파악

\* 출처: 행정안전부(2021), 공공기관 데이터 역량강화 가이드라인

[그림 5-2] 데이터 역량강화 4단계 절차



둘째, 데이터 역량강화 4단계 절차에 따라 데이터역량을 측정하고 보완을 추진한다. 데이터 역량지표 체계를 바탕으로 기관별로 필요한 데이터 역량지표와 지표별 요구수준을 정의해야 한다. 이를 바탕으로 기관의 현재 역량 측정, 역량격차 분석, 격차 해소방안 마련 등 4단계 절차를 통해 단계별 데이터 역량 현황을 측정하도록 한다.

1단계에서는 데이터기반행정 책임관의 역할이 중요하다. 데이터기반행정 책임관을 중심으로 TF를 조직하여, 기관의 비전·목표·정책 우선순위 등을 고려한 데이터역량 및 역량지표별 요구수준을 정의해야 한다. 특히, 데이터 전략 수립, 분석과제 기획, 분석결과 해석, 데이터기반 의사결정 역량을 가능한 한 필수역량으로 반영하여야 하며, 이에 대한 결과는 ‘필요 데이터역량 진단서’로 작성한다.

2단계에서는 1단계 결과를 토대로 기관 전 구성원을 대상으로 데이터 역량 자체진단 설문조사를 진행한다. 데이터 역량 요구수준(5점 기준) 대비, 기관 내 측정값을 토대로 ‘데이터역량 진단서’를 작성한다.

3단계에서는 1,2단계의 진단서를 비교, 분석하여 기관 목표, 실행가능성 등을 고려하여 역량지표별 우선순위를 선정한다.

4단계에서는 역량격차 해소 방안과 ‘데이터 역량강화 우선순위’를 매칭하여 ‘데이터 역량지표별 역량강화 추진방안’을 마련하도록 한다. 예산, 소요기간 등을 고려하여 채용, 교육, 사업발주 등의 업무 추진 방안을 마련하고, 역량강화 추진방안별 세부 이행계획을 수립한다.

일련의 과정을 통해 최종적으로는 기관 전체 구성원이 데이터 역량강화에 참여하도록 독려하고, 데이터 역량강화 성과를 환류하여 다음연도 데이터 역량강화를 위한 예산확보, 소요정원 요구, 제도 개선 등을 추진하도록 한다.

〈표 5-7〉 데이터 역량지표별 역량강화 추진방안(예시(안))

번호	데이터 역량지표	채용	교육	사업발주
1	데이터기반 의사결정		데이터기반행정 정책활용·사례교육 실시	
2	데이터 전략 수립		데이터 전략 수립 관련 교육 실시	
3	분석과제 기획		빅데이터 기획 역량강화 교육 실시	
4	분석 수행	데이터 분석 인력 00명 채용		데이터 분석 사업 발주



번호	데이터 역량지표	채용	교육	사업발주
5	데이터 이해		데이터 특성, 저장구조 등 교육 실시	
6	분석 결과 해석		분석결과 해석 관련 교육 실시	
7	분석모형 설계	연말까지 분석모형 설계 인력 00명 채용		
8	데이터 수집			데이터 수집 자동화 시스템 구축사업 발주
9	시각화 기획		데이터 시각화 관련 전문가 초청 강연 실시	
10	데이터 시각화		시각화 실습 전문교육 실시	

\* 출처: 행정안전부(2021), 공공기관 데이터 역량강화 가이드라인

## 1.2. 데이터 활용 교육

### 가. 데이터 교육의 필요성 및 교육 계획(안)

#### 1) 데이터 교육 실태 현황 및 교육의 필요성

본 연구에서는 문화체육관광부 소속·산하기관의 데이터 및 정책업무 담당자, 연구원을 대상으로 ‘데이터 업무 활용 실태조사’를 진행하였다.

데이터기반행정 활성화를 위해 필요한 것의 중요도 평균 점수(5점 기준)는 ‘데이터 관련 전문인력 확보’ 4.4점, ‘데이터 관련 부서 간 협력’ 4.4점, ‘데이터 관련 예산 확보’ 4.3점, ‘정책에 활용할 데이터 관련 기획’ 4.3점, ‘데이터를 중시하는 조직 문화’ 4.3점 등의 순으로 나타났다. 전반적으로 큰 차이를 보이지 않은 가운데에서도 인력, 예산, 교육에 대한 부분에서 중요도가 더 높게 나타난 것을 확인할 수 있다.

문화체육관광부 소속 및 산하기관 데이터 및 정책업무 담당자와 연구원을 대상으로 ‘데이터 기반 행정 강화를 위한 데이터 공동활용 중요도’를 조사한 결과, 전체 응답자 10명 중 8명 이상(87.5%)이 중요하다고 응답하였다. 반면, ‘데이터 공동활용 정책 및 사업 참여 경험’을 묻은 결과, 전체 응답자 10명 중 8명 이상(88.0%)이 ‘없다’고 응답하였다. 데이터 공동활용의 중요도 대비 실제로 데이터 공동활용 업무에 참여한 비율이 12.5%인 것은 실로 중요도 대비 현실의 간극을 보여주는 예시라고 볼 수 있다. 이는 기관 내 데이터 전문인력이 부재함을 나타내주는 또 다른 모습이라고 보여진다. 데이터

공동활용의 제약조건에 대한 질문에서도 데이터 공동활용을 위한 전문인력 부족(74%, 1순위)과, 데이터 관련 교육 지원 미흡(62.5%, 3순위)이 제약조건 심각성 정도가 높게 나타났다.

데이터 업무 활용 실태조사 결과, 최근 1년 간 ‘데이터 교육을 받은 경험’ 여부를 묻은 결과, ‘있다’고 응답한 경험률은 절반이 되지 않는 40.1%였다. 데이터 관련 ‘자격증 보유 여부’에서 ‘있다’고 응답한 경우는 불과 9.4%에 불과하다. 데이터 역량의 중요성을 인지하고 있는 정도 대비 문화체육관광부 소속·산하기관 재직자들의 현실은 데이터 관련 교육의 실제 접근성이 매우 낮은 편이라고 볼 수 있다.

데이터 업무 활용 실태조사의 결과는 <표 5-7> 데이터 역량지표별 역량강화 추진방안(예시(안))과 연계하여 살펴볼 수 있다. ‘데이터 역량지표별 역량강화 추진방안’을 마련을 위해 기관 차원에서 예산, 소요기간 등을 고려하여 채용, 교육, 사업발주 등의 업무 추진 방안을 마련하고, 주요 사항별 세부 이행계획을 수립이 필요하다.

<표 5-8> 데이터 역량강화를 위하여 보완되어야 할 주요 요건

구분	데이터 역량지표별 역량강화 추진방안(안)	데이터 업무 활용 실태조사 결과
주요 요건	<ul style="list-style-type: none"> <li>• 예산</li> <li>• 채용</li> <li>• 교육</li> <li>• 사업발주</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터 관련 전문인력 확보</li> <li>• 데이터관련 부서 간 협력</li> <li>• 데이터 관련 예산 확보</li> <li>• 정책에 활용한 데이터 관련 기획</li> <li>• 데이터를 중시하는 조직문화</li> </ul>

\* 출처: 연구진 재구성

데이터 관련 예산과 인력 채용, 데이터를 중시하는 조직 문화, 사업발주 등에 대해서는 특히 전사적인 노력이 필요하며, 단기적으로 해결되기 어려운 점이 있다. 다만, 현재의 인력을 대상으로 데이터 활용 역량을 강화할 수 있는 제반 여건을 조성하고, 단계적으로 데이터를 사용할 수 있는 역량을 향상시킬 수 있는 관련 분야별 ‘데이터 교육’을 실시하여, 데이터 업무 관련 실무활동에 적용 할 수 있을 것이다.

최근 1년 간 ‘데이터 교육을 받은 경험자(40.1%)’를 대상으로 교육 주관기관을 확인한 결과, ‘외부 기관’(40.3%), ‘재직 중인 기관 교육’(32.5%), ‘재직 중인 기관 + 외부 기관’(27.3%)인 것으로 나타나 대내외적으로 큰 편향은 없었다. 교육 방식은 ‘온라인 교육’(62.3%) 비중이 가장 높았으며, ‘온라인 + 오프라인 교육’(27.3%), ‘오프라인 교

육’(10.4%) 순이었다. 교육 방법은 ‘강의 중심’(90.9%), ‘실습 중심’(16.9%), ‘토론 중심’(6.5%)의 비율로 나타났다.

따라서 상기와 같은 교육 현황 결과에 따라 기관의 데이터 교육은 ① 교육 주관기관 기준에 따라 내부 교육과 외부 교육, ② 교육 방식 기준으로 온라인 교육과 오프라인 교육, ③ 교육 방법 기준으로 강의 중심, 실습 중심, 토론 중심 교육, ④ 교육 담당자는 관련 분야 내부 전문가 또는 외부 전문가 중에 선정하여 진행한다.

## 2) 전사적인 차원의 데이터 인식제고 교육(내부 교육)

데이터 관련 교육은 정책을 담당하는 모든 공무원 및 공공기관 종사자에게 요구되는 사항이다. 현재의 인력을 대상으로 데이터 활용 역량을 강화할 수 있는 제반 여건을 조성하고, 단계적으로 데이터를 사용할 수 있는 역량을 향상시킬 수 있는 관련 분야별 ‘데이터 교육’을 실시하여, 데이터 업무 관련 실무활동에 적용할 수 있을 것이다.

앞서 데이터 역량강화를 위한 교육(안)은 데이터 역량지표별 역량강화 추진방안(예시(안))의 내용 차용이 가능하다. 특히, 데이터 역량지표와 연계하여 분야별 교육을 실시한다면, 영역별 수요를 고려하여 체계적인 실무자 양성에 도움이 될 수 있을 것이다.

데이터 업무 활용 조사에서 확인한 바와 같이 현재의 데이터 교육 경험률(40.1%)은 이 조사의 응답자 대부분이 데이터 관련 업무 종사자임을 고려했을 때도 절반이 안 되는 미약한 수준이다. 따라서 전사적인 차원에서 기관 내 종사자들의 데이터에 대한 이해 제고 차원의 데이터 인식제고 교육을 시행하는 방안을 우선 추진이 필요하다.

업무 영역 변경, 업무 환경 및 담당자 변경 등을 고려하여 7가지 영역에 대해서는 2년 내에 적어도 1회 이상 경험이 가능하도록 교육 프로그램을 구성하는 방법을 제시하고자 한다. 공통 분야 교육(데이터기반 의사결정, 데이터 이해 및 시각화 기획 등)을 설계하여, 조직 구성원들이 데이터 역량강화에 대한 전반적인 흐름을 체득할 수 있도록 한다. 그리고 조직 구성원들의 업무 특성을 반영하여 선행적으로 이수해야 할 분야를 선택하도록 한다. 이를 테면, 전략과 기획 분야 업무가 메인인 대상자는 데이터 기획 분야 교육을 우선 실시하고, 연구, 조사 및 데이터 업무가 메인인 경우에는 데이터 분석 분야 교육을 선행하도록 한다. 그리고 다음 해에는 각 업무 담당자들이 전년도에 미이수 분야의 교육을 이수하도록 하여 분야별 공백이 없도록 한다.

〈표 5-9〉 데이터 역량강화를 위한 교육(안)

구분	데이터 역량지표 기준 분야	관련 교육(안)
공동 분야 (매년)	데이터 이해	데이터 특성, 저장구조 등 교육
	데이터기반 의사결정	데이터기반행정 정책활용·사례교육
	시각화 기획	데이터 시각화 관련 전문가 초청 강연
데이터 기획 분야 (격년)	데이터 전략 수립	데이터 전략 수립 관련 교육
	분석과제 기획	빅데이터 기획 역량강화 교육
데이터 분석 분야 (격년)	분석 결과 해석	분석결과 해석 관련 교육
	데이터 시각화	시각화 실습 전문교육

\* 출처: 연구진 재구성

데이터 인식강화를 위한 전사적인 교육은 ① 기관 내부 교육으로 진행하며, ② 교육 방식은 기관의 여건을 고려하여 온라인 교육과 오프라인 교육을 병행하고, ③ 교육 방법은 강의 중심으로 진행하되, 기관의 제반 현황에 따라 실습과 토론 교육을 실시하도록 한다.

## 나. 전문인력 양성과정 운영

데이터 관련 전문인력 양성에 대해서는 데이터기반행정법 제24조에 명문화되어 있다. 행정안전부 및 과학기술정보통신부 장관이 관련 시책 마련 주체자로서 할 수 있는 역할들에 대해서 확인이 가능하다.

〈표 5-10〉 데이터기반행정법 제24조(데이터 관련 전문인력 양성)

- ① 행정안전부장관과 과학기술정보통신부장관은 데이터 관련 전문인력 양성을 위하여 필요한 다음 각 호의 시책을 마련할 수 있다.
  1. 전문인력의 수요 실태 파악 및 중장기 수급 전망 수립
  2. 전문인력 양성 교육훈련 프로그램의 개발 및 보급 지원
  3. 데이터 활용 관련 직무표준의 마련 및 자격제도의 정착 지원
  4. 전문인력 고용창출 지원
  5. 그 밖에 전문인력 양성에 필요한 사항
- ② 행정안전부장관과 과학기술정보통신부장관은 데이터 관련 전문인력을 효율적으로 활용하기 위하여 필요한 경우 관련 연구기관이나 민간단체 등과 협력할 수 있다.
- ③ 그 밖에 데이터 관련 전문인력 양성에 필요한 사항은 대통령령으로 정한다.

마찬가지로 문화체육관광부 데이터 관리 규정에서도 문화체육관광부 장관은 공무원들의 데이터 역량 강화를 위하여 데이터 관련 교육 실시 및 개설, 전문인력을 양성하기 위한 시책 추진, 예산 등을 지원할 수 있다고 명시되어 있다.

〈표 5-11〉 문화체육관광부 데이터 관리 규정 제8조(교육 및 전문인력 양성)

- ① 장관은 공무원들의 데이터 역량 강화를 위하여 데이터 등과 관련된 교육을 실시할 수 있고, 필요하다고 인정 되는 경우 교육기관에 교육과정의 개설을 요청할 수 있다.
- ② 장관은 데이터 등의 활용에 필요한 전문인력을 양성하기 위한 시책을 추진할 수 있다.
- ③ 장관은 데이터 등의 교육과 전문인력 양성에 필요한 경비는 예산의 범위에서 지원할 수 있다.
- ④ 장관은 데이터 등의 정책 활용과 유관산업 육성을 지원하기 위하여 민간이 참여하는 경진대회를 개최할 수 있으며, 우수자에게는 상장과 포상금을 지급할 수 있다.

앞서 데이터 업무 활용 실태조사의 결과는 〈표 5-8〉 데이터 역량강화를 위하여 보완 되어야 할 주요 요건에서 확인한 바와 전문인력의 확보는 데이터기반행정에서 매우 중요한 조건이다. 그러나 전문인력의 확보는 예산 및 인력 확충의 조건이 필요충분적으로 마련되어야 가능하다. 따라서 전문인력이 충원되는 조건이 충족되지 않는 한, 현재 기관에 재직 중인 인력을 대상으로 데이터 역량을 강화하는 방안이 필요하다.

데이터 활용 실태조사 결과, 문화체육관광 소속·산하 기관의 데이터 관련 응답자 중 ‘자격증을 보유하고 있다’는 응답률은 9.4%에 불과하다. 업무로 인해 교육여건이 여의치 않아 교육에 투자하지 못하는 경우가 적지 않을 것이다. 이에, 외부 전문가 및 기관을 활용한 데이터 전문분야 특화 교육을 추진하는 하되, 교육 희망자가 교육 이수를 위해, 교육 기관을 방문하기 어려울 수 있는 상황 등을 고려하여, 기관 내부에서 교육을 받을 수 있는 프로그램을 운영하는 방법을 고려할 필요가 있다. 이미 국내에서도 SAS Korea, 한국생산성본부 등 일부 교육기관에서는 기관이 원하는 장소에서 기업이 원하는 교육을 단기적으로 제공하는 맞춤형 교육을 진행하고 있다.

데이터 전문분야 특화 교육은 외부 전문가를 초빙하여 ① 교육 장소에 따라 내부교육과 외부 교육, ② 교육 방식 기준으로 온라인 교육과 오프라인 교육, ③ 교육 방법 기준으로 강의 중심, 실습 중심, 토론 중심 교육 중에 진행 방식을 선택하여 추진한다.

#### 1) 외부 전문가의 기관 방문 교육 과정

데이터 교육의 필요성은 인정하지만 업무로 인해 교육여건이 여의치 않아 교육에 투자하지 못하는 기관을 직접 방문하는 방식으로 재직자 교육을 추진할 필요가 있다. 이미 국내에서도 SAS Korea, 한국생산성본부 등 일부 교육기관에서 기관이 원하는 장소에서 기업이 원하는 교육을 단기적으로 제공하는 맞춤형 교육을 진행하고 있다.

외부에서 진행되는 교육은 시간적 그리고 공간적인 제약으로 인하여 업무처리에 우선 순위를 뺏기는 경우가 빈번히 발생한다. 따라서 외부 교육의 추진보다는 내부에서 진행하는 교육을 활성화하는 것이 더 바람직할 것이다. 내부에서 진행하는 데이터 관련 교육을 활성화하기 위해서 기관의 특성에 맞추어 교육과정을 체계적으로 수립해야한다. 또한 조직 내 학습을 독려하는 분위기가 조성되어야 한다. 이는 교육의 성패를 가를 수 있는 중요한 항목이다. 이와 같은 교육은 기관 내 데이터 전문인력 또는 부서를 중심으로 교육과정을 점검하고 추진해야 하며, 변화하는 환경과 기술을 반영하여 지속적인 최신화가 필요하다. 외부 전문가의 기관 방문 교육은 이론교육 보다는 실습 위주의 교육을 통해 채용 즉시 현장에 투입될 수 있는 인력양성 사업을 추진하는 것을 우선으로 한다. 외부 전문가의 기관 방문 교육 교육은 ① 내부교육, ② 오프라인 교육, ③ 강의 중심, 실습 및 토론 중심 등의 혼합 방식으로 추진 가능하다.

## 2) 외부 전문기관에서 운영하는 고급인력 양성과정

재직자를 대상으로 한 데이터 과학자 양성 프로그램을 적극 개발하여 직원들이 데이터 과학자 역량을 갖추도록 하는 한편, 자격증 과정을 도입하여 데이터 인력의 공신력을 확보할 필요가 있다. 데이터 역량은 일정 수준 정도 기관에서의 내재화가 필요하다. 데이터 분석의 경우 기관에 축적된 데이터를 기관 재직자가 분석할 때, 결과의 활용성을 보다 향상시킬 수 있다. 따라서 각 분야의 최신 전문지식을 기반으로 역량을 갖춘 인력을 확보하기 위한 재직자 교육 프로그램을 도입하거나, 전문 교육기관을 통한 인력양성을 추진해야 한다.

과학기술정보통신부와 한국데이터산업진흥원은 산업계 재직자를 대상으로 빅데이터 전문가 양성 프로그램을 운영하고 있다.

교육프로그램은 ① 빅데이터 직무 수준별 구분에 따라 빅데이터 기획전문가(중급), 빅데이터 기술 전문가(중급/고급), 빅데이터 분석 전문가(중급/고급), ② 팀 프로젝트 기반의 이론+실습+멘토링 교육이 결합된 산업융합 과정, ③ 전문 분야별 구분에 따른 데이터 시각화 전문가, 데이터 거래 중개 전문가, 데이터 가공 전문가 양성 교육 과정으로 구성되어 있다. 세부적으로는 아래와 같이 교육 과정을 운영하고 있다.

[그림 5-3] 빅데이터 아카데미 과정

<b>ANALYSIS</b> <span>중급</span> <b>빅데이터 분석 전문가 33기</b> 교육기간 : '21.06.28(수) ~ 07.09(금) 프로젝트 : '21.07.12(월) ~ 08.02(목) 산형 기간 : 21.05.30(월) ~ 06.14(화) #Python #SQL #R #데이터분석 #데이터과학 #인공지능 #빅데이터	<b>ANALYSIS</b> <span>중급</span> <b>빅데이터 분석 전문가 34기</b> 교육기간 : '21.06.28(수) ~ 07.09(금) 프로젝트 : '21.07.12(월) ~ 08.02(목) 산형 기간 : 21.05.31(화) ~ 06.22(화) #Python #SQL #R #데이터분석 #데이터과학 #인공지능 #빅데이터	<b>ANALYSIS</b> <span>중급</span> <b>빅데이터 분석 전문가 35기</b> 교육기간 : '21.06.16(수) ~ 07.07(금) 프로젝트 : '21.06.30(월) ~ 07.05(목) 산형 기간 : 21.06.17(목) ~ 06.03(화) #Python #SQL #R #데이터분석 #데이터과학 #인공지능 #빅데이터	<b>PROCESSING</b> <span>중급</span> <b>빅데이터 기술 전문가 21기</b> 교육기간 : '21.06.28(수) ~ 07.09(금) 프로젝트 : '21.07.12(월) ~ 08.01(수) 산형 기간 : 21.05.31(화) ~ 06.22(화) #DataSearch #hadoop #Kafka #Spark #기타빅데이터
<b>PROCESSING</b> <span>고급</span> <b>빅데이터 기술 전문가 22기 (데이터 재테크 구성 학습 과정)</b> 교육기간 : '21.06.27(월) ~ 07.09(금) 프로젝트 : '21.07.11(화) ~ 07.11(목) 산형 기간 : 21.05.30(월) ~ 06.02(화) #DataSearch #hadoop #Kafka #Spark #기타빅데이터	<b>PLANNING</b> <span>중급</span> <b>빅데이터 기획 전문가</b> 교육기간 : '21.07.12(월) ~ 07.16(금) 프로젝트 : '21.07.26(목) ~ 08.05(목) 산형 기간 : 21.05.31(화) ~ 06.28(화) #빅데이터 기획 #빅데이터 서비스 분석 #빅데이터 기술 #빅데이터 기업 #빅데이터 산업	<b>INDUSTRY</b> <span>중급</span> <b>산업 융합형 과정 1기</b> 교육기간 : '21.06.18(월) ~ 07.09(금) 프로젝트 : '21.07.03(화) ~ 07.08(목) 산형 기간 : 21.06.21(화) ~ 06.26(화) #빅데이터로 만들기 #빅데이터 분석 #빅데이터 #빅데이터 #빅데이터 #빅데이터 #빅데이터 #빅데이터	<b>INDUSTRY</b> <span>중급</span> <b>산업 융합형 과정 2기</b> 교육기간 : '21.06.18(월) ~ 07.09(금) 프로젝트 : '21.07.03(화) ~ 07.08(목) 산형 기간 : 21.06.21(화) ~ 06.26(화) #빅데이터로 만들기 #빅데이터 분석 #빅데이터 #빅데이터 #빅데이터 #빅데이터 #빅데이터 #빅데이터
<b>VISUALIZATION</b> <b>데이터 시각화 전문가</b> 교육기간 : '21.06.06(수) ~ 06.10(금) 프로젝트 : '21.06.13(월) ~ 06.04(수) 산형 기간 : 21.05.30(월) ~ 06.02(화) #시각화 #데이터 #데이터 #데이터 #데이터 #데이터 #데이터 #데이터	<b>REMAKER</b> <b>데이터 가공 전문가</b> 교육기간 : '21.06.06(수) ~ 06.10(금) 프로젝트 : '21.06.13(월) ~ 06.04(수) 산형 기간 : 21.05.31(화) ~ 06.01(수) #데이터 #데이터 #데이터 #데이터 #데이터 #데이터 #데이터 #데이터	<b>CURATOR</b> <b>데이터 기획 설계 전문가</b> 교육기간 : '21.06.04(수) ~ 06.10(금) 프로젝트 : '21.06.18(월) ~ 06.04(수) 산형 기간 : 21.05.31(화) ~ 06.01(수) #데이터 #데이터 #데이터 #데이터 #데이터 #데이터 #데이터 #데이터	

\* 출처: 한국데이터산업진흥원 빅데이터 아카데미(<https://dataonair.or.kr/bigdata/apply2021/>)

본 연구에서는 데이터 전문가 양성을 위한 교육 프로그램의 예시로 빅데이터 아카데미 과정을 활용하였다. 이와 같은 교육은 민간 등에서도 유사한 프로그램을 운영 중에 있다. 외부 전문기관에서 운영하는 고급인력 양성과정은 ① 외부교육, ② 오프라인 교육 (경우에 따라 온라인 교육), ③ 강의 중심, 실습 및 토론 중심 등의 혼합 방식으로 추진 가능하다.

## 2. 데이터기반행정 활성화 문화 조성

정부부처 및 공공기관 등의 기존 업무 프로세스는 데이터의 활용이 미흡하고 데이터 기반행정에 대한 인식 부족 등의 이유로 데이터기반행정이 지속성과 실효성을 확보하는데 한계가 존재한다(제1차 데이터기반행정 활성화 기본계획('21년~'23년), 2021.2., 관계부처 합동). 또한 관련 시스템의 구축·운영, 데이터 전문 인력의 추가 확보, 데이터 활용 역량 강화를 위한 교육 등을 실시하고 있으나 아직까지 데이터기반행정 활용 사례나 그 실효성에 대해서는 보장하고 있지 못하다. 또한 데이터에 대한 중요성은 인식하고 있어도 업무 담당자들이 다양한 분야의 행정업무와 데이터를 접목시키는 작업에 익숙하지 않으며, 데이터 활용에 있어서의 개인정보보호, 정보보안 등의 법적·윤리적 이슈에 대해 파악이 어려운 상황이다. 이에 데이터 관련 업무 담당자뿐만 아니라, 전사적 업무에 걸쳐 데이터를 올바르게 이해하고 활용할 수 있는 기반환경을 조성할 필요가 있다.

정부는 제1차 데이터기반행정 활성화 기본계획을 통하여 정책결정에 데이터를 적극 활용하는 과학적 행정 구현과 국민이 신뢰하고 공감하는 지능형 행정서비스 제공이라는 목표를 실현하기 위한 추진전략으로 ① 데이터 통합 기반 구축으로 데이터 공동활용, ② 데이터기반행정의 활성화를 위한 제도 확립, ③ 지능형 서비스 제공을 위한 데이터 분석·지원, ④ 데이터기반의 일하는 방식으로 혁신이라는 4가지 전략을 내세웠다. 이 중 데이터기반의 일하는 방식으로 혁신에 대한 세부 추진과제를 보면, 첫 번째 데이터 관점의 행정 프로세스 혁신이다. 정책수립-집행-평가의 쉼 과정에 데이터 활용 촉진하기 위하여 각 단계에서 데이터를 활용하여 맞춤형 정책을 수립하고 집행상황을 점검하는 것이다. 수립단계에서 필요한 데이터를 정의하고 집행단계에서 데이터 분석(민원분석, 소셜 데이터분석 등)을 통해 정책 모니터링 및 향후 방향 설정, 평가단계에서는 정책집행 결과의 효과성을 검증하는 것이다.

〈표 5-12〉 정책 라이프사이클 빅데이터 활용(예시)

- 
- ◇ (수립단계) 정책 수립 타당성 및 효과성 평가(예측) 등
    - (예시) CCTV, 공공자전거 거처대 입지 선정 시 데이터에 기반한 입지 타당성 및 효과성 평가
  - ◇ (집행단계) 정책 모니터링 및 영향성 평가 등
    - (예시) 선정된 입지에 대한 소셜데이터 분석으로 입지 조정 등 개선방안 도출
  - ◇ (평가단계) 정책 성과 점검 및 개선사항 도출 등
    - (예시) 범죄발생 감소, 공공자전거 이용 활성화 등 데이터기반 성과 평가
- 

출처: 제1차 데이터기반행정 활성화 기본계획('21년~'23년), 2021.2., 관계부처 합동



두 번째는 데이터기반행정의 동력 확보를 위한 조직과 인력을 강화하고, 기관별 데이터 역량 제고를 위한 프로그램을 개발 및 제공하며, 체계적인 단계별 교육 제공을 통해 데이터기반 혁신역량을 강화하는 것이다. 마지막으로 데이터기반행정 우수사례를 발굴·홍보하고, 데이터기반 문제해결 방안에 대한 국민참여를 촉진함으로써 데이터기반 과학적 행정 문화를 확산하는 것이다. 데이터기반행정의 필요성과 주요 추진과제를 적극 홍보하고, 데이터기반행정에 대한 인센티브를 제공하는 것 또한 데이터기반행정 문화 조성에 도움이 된다.

이와 같이 정책을 수립하는 행정 프로세스를 개선하고, 인력 및 역량을 강화하며, 데이터기반행정 문화를 확산하는 것 외에도 중요한 것은 업무책임자(리더)의 역량과 데이터기반행정에 대한 의지이다. 업무책임자가 데이터에 대한 분석능력이 있고, 데이터기반행정에 대해 적극적인 태도와 의지를 가지고 있다면 해당 조직은 조직 내부에서부터 자발적으로 데이터기반의 업무 수행 문화가 조성될 것이다.

### 3. 공동활용 제약 요소 발굴 및 개선사항 마련

데이터의 공동활용을 제한하는 장애요소는 데이터 분석 전문 인력의 부재 또는 데이터 분석 역량의 부족, 데이터를 활용·분석할 수 있는 시스템의 부재, 법적·제도적 기반 부족 등과 같이 다양하게 존재한다. 필요한 데이터의 존재여부나 위치를 파악하기 어렵다는 것도 하나의 제약요소로 판단할 수 있다. 이러한 점은 행정안전부에서 제공하는 공동활용데이터 등록관리시스템 또는 국가데이터맵 서비스를 통해 어느 정도 대응이 가능하나, 데이터 자체를 보고 판단하는 것은 아니기 때문에 여전히 데이터의 특성을 파악하기에는 어려움이 존재한다는 단점이 존재한다.

또한 데이터의 존재를 알고 해당 기관에 데이터를 요청하여도 해당 데이터를 제공받지 못하거나, 제공받기에 어려움이 있다. 이는 본 연구에서 진행한 ‘데이터 업무 활용 실태조사’ 결과를 살펴보면, 좀 더 명확하게 나타난다.

[그림 4-31]을 살펴보면, 데이터 공동활용의 제약조건의 심각성에 대해 데이터 공동활용을 위한 전문인력 부족이 심각하다고 응답한 사람은 응답자의 74.0%로 가장 높은 수치를 기록했다. 이를 통해 데이터를 분석하고 활용할 수 있는 전문인력의 확보가 필요

하며, 기 보유 인력에 대해서는 데이터의 활용 측면에 대한 역량 강화가 필요함을 알 수 있다. 전문인력 부족 다음으로 심각성이 높은 제약조건은 데이터 간 표준화 미흡으로 나타났다. 표준화 미흡이 심각하다고 응답한 사람은 70.3%로 이는 데이터의 속성 기준과 용어의 정의 등이 기관별 담당자별로 상이하기 때문일 것으로 판단된다. 따라서 필요한 데이터를 취득하여 활용하고자 하여도 데이터가 수집된 기준이 활용하고자 하는 기준과 상이할 수 있으며, 데이터 간 표준화가 되어 있지 않아 데이터 연계 시에도 어려움이 발생한다. 이러한 표준화 부분에 대해서는 앞서 언급한 것과 같이 장기적으로는 표준화 체계를 수립하여 추진하더라도 단기적으로 중앙에서 수집할 때 표준화된 형태로 데이터를 수집하는 등의 수집 체계 개선이 필요하다.

그 외에도 필요 데이터의 미흡(61.5%), 공동활용 관련 교육 지원 미흡(62.5%), 개인정보보호 문제(63.5%) 등의 제약요인이 심각하다고 응답하였다. 이러한 제약요인들은 앞서 모두 언급했던 것으로, 요인별로 기관의 특성에 맞춰 개선사항을 마련해야 하며 특히 개인정보보호의 이슈에 대해서는 정부 차원에서 데이터 공동활용을 위한 방안을 마련해야 한다.

이외에도 데이터 보유 기관에서 데이터의 공동활용을 거부하는 다양한 사례를 분석하여 원인을 파악하고, 공동활용이 가능하도록 개선사항을 도출할 필요가 있다.

## 제3절 소결

---

데이터의 공동활용 활성화를 위해서는 공동활용하기 위한 데이터의 관리 체계를 수립하고, 데이터를 수집하며, 이용자들이 접근하여 이용 및 활용할 수 있는 시스템을 구축해야 한다.

데이터의 관리 체계를 수립하기 위해서는 기관별로 보유한 데이터를 파악하여 수집하되, 데이터 수집 기준과 체계를 먼저 정립해야 한다. 정부부처 별로 데이터를 중복하여 요청하는 것을 지양하기 위하여 정부에서 데이터를 수집하는 기준과 체계를 수집해야 하며, 기관의 규모나 특성에 따라 데이터에 대한 기관 차원의 중앙 관리가 되지 못하고 업무 담당자별로 데이터를 관리하는 경우가 발생하고 있으므로 기관별 데이터 수집 및 개방을 총괄하기 위한 방법과 관리체계 수립이 필요하다. 또한 기관 간 및 기관 내 데이터가 가진 용어와 도메인, 코드 등의 표준화를 통해 다른 기관의 데이터를 활용함에 있어 데이터에 대한 이해도를 높이고 오사용을 줄일 수 있다. 단, 기 구축되어 있는 데이터 체계와 DB에 대한 표준화는 시간적·경제적인 어려움이 있으므로, 장기적인 관점에서 표준화를 진행해야 하며 단기적으로는 중앙 수집 시의 표준화를 고려할 수 있다. 메타데이터의 관리를 통해 이러한 데이터의 표준화가 용이해지며, 데이터 카탈로그를 통해 데이터의 소재, 특성 등의 파악이 가능하며, 체계적인 관리 및 서비스 제공이 가능하다.

데이터의 공동활용을 위하여 행안부는 3가지 방식을 통하여 공동활용 데이터를 등록하게끔 유도하고 있다. 기관 자체적 등록인 자율등록과 행안부가 기관에 등록 요청하는 지정등록, 행안부에서 자체적으로 수집하여 등록하는 수집등록이다. 공동활용데이터 등록관리시스템 내에 기관 내에서 직접 등록이 가능하며, 기관 내 메타관리시스템에 개방으로 설정되어 있는 데이터는 자동으로 수집되는 체계로 구성되어 있다. 행안부는 ‘데이터 공동활용 가이드라인’과 ‘데이터 역량강화 가이드라인’을 제공하고 있으나, 홍보 등이 부족하여 업무담당자의 혼란이 야기되는 경우가 많으며, 기관 차원의 공동활용 데이터

등록 건수를 기관 실적으로 처리하는 경우 등록 건수 위주의 과도한 등록을 야기할 수 있으므로 지양해야 한다. 기관 차원의 데이터를 수집할 수 있는 데이터시스템을 구축하여, 데이터가 업무담당자에 의존되지 않도록 해야 한다. 또한 수집된 데이터는 중앙의 데이터허브를 통해 활용이 가능하며, 제공하는 데이터에 대하여 데이터 활용 기관, 목적 등을 관리함으로써 정해진 목적에 의해 활용되도록 해야 한다.

데이터기반행정의 추진기반을 마련하고 역량을 강화하기 위하여 데이터기반행정의 관리체계를 정립하고 데이터의 활용역량을 강화해야 하며 데이터기반행정이 활성화되는 문화를 조성하여야 한다.

데이터기반행정의 정착을 위하여 데이터 수집·활용 계획의 수립여부나 적절성 등을 사전에 확인하는 제도를 도입하는 등 관련 법과 제도를 정비해야 하고, 데이터기반행정 법에 따른 책임관의 임명 등 기관별 내부 추진체계를 마련해야 한다. 또한 데이터 관련 업무를 담당하는 조직을 신설하거나 기능을 강화하고 데이터에 대한 전문 인력이 확충되어야 할 뿐만 아니라 관련 인프라를 운영 및 유지하기 위한 예산의 확보 또한 중요하다.

또한 데이터의 분석 및 활용에 대한 역량을 진단하고 단계별 교육을 실시하여 지속적으로 강화해 나가야 한다. 행정안전부는 ‘공공기관 데이터 역량강화 가이드라인’을 발표하여 공공기관이 데이터기반행정 활성화를 위해 필요한 데이터 역량을 측정하고 보강할 수 있도록 절차와 기준을 안내하고 있다. 해당 가이드라인에서는 단계별로 데이터역량을 측정하고 보완할 것을 제시하고 있다. 또한 데이터 활용에 대한 지속적인 교육과 전문인력을 양성하기 위한 교육과정을 제공해야 할 필요가 있다.

또한 그 무엇보다 중요한 것은 데이터기반행정 문화가 정착하는 것이다. 그러기 위해서는 데이터기반행정 활성화 문화를 조성해야 하므로, 이를 위해 행정프로세스를 데이터 관점으로 개선하며, 전문 인력을 확충하고, 데이터기반행정에 대한 우수사례 등을 발굴하고 홍보해야 한다. 또한 데이터의 공동활용에 대해 거부하는 사례 등 제약사항을 지속적으로 분석하여 꾸준히 개선해 나가야 한다.

## 결론 및 제언



## 제1절 결론

---

### 가. 연구의 요약

본 연구는 데이터기반행정 강화를 위한 문화·관광·콘텐츠 데이터 공동활용 기반 마련을 목적으로 수행한 과제이다. 이를 위해 정책 환경 분석, 사례 분석, 데이터 현황조사를 실시하였고 이를 바탕으로 데이터 공동활용 활성화를 위한 정책 방안을 제시하였다. 이론적 논의에서는 데이터기반행정 추진여건 및 환경분석을 통해 데이터행정의 이론적 의의를 살펴보고, 증거기반정책과 같은 과학적이며 합리적인 정책과정에 대한 요구가 증대되면서 데이터기반행정의 필요성이 증대되고 있음을 확인하였다.

국내외 사례 분석을 통해 데이터 공동활용의 중요성을 확인할 수 있다. 국내사례는 정부부처 및 지방자치단체가 실시한 주요 사업 중 데이터 공동활용의 우수사례로 평가받은 사례를 중심으로 배경과 추진내용, 그리고 데이터기반행정에 있어서 그 사업이 가지는 의의를 살펴보았다. 국내사례의 경우 주요 이슈와 관련된 활용사업을 개발하는 과정에서 데이터를 통합이나 연계 없이 사업을 계획하거나 추진하는 단계별로 필요한 개별적인 통계 수치 및 데이터를 활용하여 사업 및 정책을 만들었다는 특징이 있었다. 국외사례에서는 데이터기반행정의 모범적인 사례로 평가받고 있는 영국, 스웨덴, 네덜란드의 데이터기반행정 및 데이터 공동활용 주요 사례를 중심으로 데이터기반행정의 정책환경 및 해당 사업의 추진배경 및 추진내용 등을 살펴보았다. 데이터기반행정의 모범 사례로 평가받고 있는 나라의 경우 구체적인 데이터 관련 법·제도를 기반으로 적극적인 데이터 개방 및 부처 간 협력체계가 구축되어 있었다.

기관별 데이터 보유 현황 및 분석업무에 필요한 타 기관 데이터 수요를 파악하여 데이터 기반행정 활성화의 추진기반의 기초자료로 활용하고자 데이터 현황조사(공공데이터 현황조사, 보유 및 수요 데이터 조사, 데이터 업무 활용 실태조사)를 실시하였다.

문화체육관광 분야 공공데이터 포털 개방 현황을 살펴보면, 총 26,979개의 공공데이터 중 문화재 관련 공공데이터가 10,584개로 가장 많았고, 다음으로 관광 관련 공공데이터가 5,880개, 문화예술 관련 공공데이터가 5,099개 등의 순으로 나타났다. 문체부 소속 및 산하기관이 공공데이터포털에 개방한 데이터와 행정안전부 공공데이터포털에서 개방하고 있는 문화체육관광 분야 전체 데이터를 비교해 보면, 문체부 소속 및 산하기관에서 제공하고 있는 데이터의 비중은 문화체육관광 분야 전체 데이터의 3.2% 수준으로 매우 낮은 비중을 보이는 것으로 나타났다.

문화체육관광 분야 조직 분류를 기준으로 보유 데이터 현황을 공동활용 정책분야별로 구분하여 조사결과를 살펴보았다. 전체 문화체육관광 분야 데이터 1,364개 중 문화예술 데이터가 560개로 가장 많은 것으로 나타난 가운데, 문화예술을 다른 조직 분류(분야)와 비교해서 살펴볼 경우에도 5가지 공동활용 정책분야 모두에서 문화예술 관련 업무 데이터의 비중이 가장 높은 것으로 나타났다. 행정안전부에 권고하고 있는 5가지의 공동활용 정책분야별 데이터 비중을 살펴보면, “특화된 대책 및 맞춤형 서비스” 관련 데이터가 1,164개(85.5%)로 가장 많았고, 다음으로 “행정업무 경제성·효율성 개선” 112개(8.2%), “미래수요에 대한 선제적 대응” 40개(2.9%) 순으로 공동활용 정책분야 데이터 비중이 높은 것으로 조사됐다.

데이터 업무 활용 실태조사 결과를 종합해보면, 문화체육관광부 소속 및 산하기관 데이터 및 정책업무 담당자들은 담당 업무를 수행할 때 데이터를 많이 활용하고 있으며, 데이터기반행정 활성화를 위해서는 인력, 예산, 장비·시설의 확충이 중요하다고 생각하고 있었다. 그러나 데이터기반행정 실현 정도는 중요도가 높았던 항목들을 충분하지 않다고 생각하고 있음을 알 수 있다. 또한 데이터기반행정 강화를 위한 데이터 공동활용의 중요성은 높게 생각했으나 실현 정도는 잘 이뤄지지 않고 있다고 보았다. 그리고 공동활용의 제약조건으로는 ‘전문인력이 부족’하고, ‘데이터의 표준화가 미흡’하며, ‘개인정보 보호의 문제’가 있음을 알 수 있었다.

본 연구에서 실시한 데이터 현황조사(공공데이터 현황조사, 보유 및 수요 데이터 조사, 데이터 업무 활용 실태조사) 결과를 활용하여 데이터기반행정법에 명시된 정책분야에 부합하는 공동활용 데이터 목록화 작업을 진행하였다. 목록화 작업을 위해 기관별 보유 데이터 외에 기관 담당자별로 업무 수행 시 필요한 공동활용 데이터 수요를 파악하였고 필요시 기관이 보유한 공동활용 데이터 목록 및 시스템 정보 등 부가정보 요청도



병행하여 추진하였다. 이어서 목록화된 데이터를 기준으로 정책분야별 분석활용 과제(안)을 도출하였다. 행정안전부가 권고하고 있는 공동활용 정책분야별로 활용될 수 있는 문화체육관광 분야 데이터를 기준으로 타 부처 및 민간 데이터와 공동활용할 수 있는 대상 데이터를 설정하였고 이를 토대로 행정내부 서비스 및 국민 서비스 개선을 위해 효율적으로 활용 될 수 있는 정책분야별 분석활용(안)을 제시하였다. 제시한 정책분야별 분석활용(안) 중 문화체육관광 분야 기획과제 및 기관 간 협업 과제 등으로 발굴될 수 있는 주요 분석과제를 분석주제, 분석 배경 및 목적, 공동활용 대상 데이터, 분석 설명 및 방법, 기대효과 등으로 구체화하여 분석과제(안)를 주요 정책분야별로 제시하였다.

데이터의 공동활용 활성화를 위해서는 공동활용하기 위한 데이터의 관리 체계를 수립하고, 데이터를 수집하며, 이용자들이 접근하여 이용 및 활용할 수 있는 시스템을 구축해야 한다. 데이터의 관리 체계를 수립하기 위해서는 기관별로 보유한 데이터를 파악하여 수집하되, 데이터 수집 기준과 체계를 먼저 정립해야 한다. 정부부처 별로 데이터를 중복하여 요청하는 것을 지양하기 위하여 정부에서 데이터를 수집하는 기준과 체계를 수집해야 하며, 기관의 규모나 특성에 따라 데이터에 대한 기관 차원의 중앙 관리가 되지 못하고 업무 담당자별로 데이터를 관리하는 경우가 발생하고 있으므로 기관별 데이터 수집 및 개방을 총괄하기 위한 방법과 관리체계 수립이 필요하다.

데이터기반행정의 정착을 위하여 데이터 수집·활용 계획의 수립여부나 적절성 등을 사전에 확인하는 제도를 도입하는 등 관련 법·제도를 정비해야 하고, 데이터기반행정법에 따른 책임관의 임명 등 기관별 내부 추진체계를 마련해야 한다. 또한 데이터 관련 업무를 담당하는 조직을 신설하거나 기능을 강화하고 데이터에 대한 전문 인력이 확충되어야 한다. 이어서, 데이터기반행정의 추진기반을 마련하고 역량을 강화하기 위하여 데이터기반행정의 관리체계를 정립하고 데이터의 활용역량을 강화해야 하며 데이터기반행정이 활성화되는 문화가 조성되어야 할 것이다. 마지막으로 데이터기반행정의 정착을 위하여 데이터 수집·활용 계획의 수립여부나 적절성 등을 사전에 확인하는 제도를 도입하는 등 관련 법과 제도를 정비해야 하고, 데이터기반행정법에 따른 책임관의 임명 등 기관별 내부 추진체계를 마련해야 한다. 또한 데이터 관련 업무를 담당하는 조직을 신설하거나 기능을 강화하고 데이터에 대한 전문 인력이 확충되어야 한다.

데이터 공동활용 기반 구축을 논의함에 있어 가장 중요한 것은 공동활용이 추구하는 목적이다. 데이터 공동활용은 데이터기반행정의 기본 전제임을 간과해서는 안 된다. 경험과 직관이 아닌, 객관적인 결과를 도출하고 이를 통해 과학적 행정을 구현하는 것이 데이터 공동활용의 목적인 것이다. 데이터를 효율적으로 활용하기 위한 방안으로 데이터 간 연계·결합 등을 시도해 보는 것도 의미가 있지만, 이보다 선행되어야 할 것이 데이터 공동활용의 목적은 무엇인지, 데이터 연계·결합을 어느 범위까지 진행할 것인지에 대한 구체적이고 명확한 목표설정이다. 이를 위해서는 공동활용을 위해서 필요한 데이터는 무엇인지, 활용 대상 데이터의 소재 및 현황은 어떻게 되는지를 파악하는 것이 중요하다. 또한 기관 간, 부서 간 데이터를 공동활용하고, 요청·제거하여 보이지 않은 칸막이를 제거해나가야 한다. 이를 통해 기관 간, 부서 간 보유하고 있는 데이터를 효율적으로 공동활용할 수 있는 데이터기반의 행정 문화를 조성하고 확산해 나갈 필요가 있다.

앞으로는 경험과 직관에 의한 업무 추진을 지양하고 객관성이 담보된 데이터에 기반한 과학적인 행정 구현이 무엇보다 중요한 시점이다. 향후에는 데이터기반행정 활성화를 위한 실행전략을 구체화 시킬 수 있는 단계별 데이터 공동활용 활성화 연구가 필요하다. 구체적으로는 데이터 공동활용 기반 마련, 데이터기반행정 추진기반 및 역량 강화 등의 실행전략을 추진과제로 세분화하여 연구계획을 수립하고 중장기적 관점의 정책 연구가 진행되어야 할 것이다. 이 외에도 공공데이터 생성·연계·활용 부분의 타당성을 검토하는 연구가 필요하다. 특히 오늘날 민간에서 생산하는 데이터가 폭발적으로 증가하는 시점에서 공공영역의 데이터만이 아니라 민간영역의 데이터 조사를 통해 데이터 공동활용의 적절성을 평가해볼 필요가 있다. 또한 데이터기반행정과 데이터 공동활용 수준을 국제비교하여 우리나라 수준을 가늠할 수 있는 연구 수행도 필요할 것으로 보인다. 이를 통해 우리나라의 수준 파악과 그에 따른 대응이 가능하게 될 것이다. 더불어, 지속적인 데이터기반의 행정추진 가운데 각 기관에 데이터 역량과 데이터기반 문화가 조성됨으로써 객관적이며 과학적인 행정 및 의사결정이 가능해질 것이다.

## 나. 한계점

본 연구는 문체부 소속 및 산하기관에서 분산 축적·관리되고 있는 다양한 유형의 데이터 정보를 파악하고, 데이터 공동활용 기반을 마련하였다는 점에서 의의가 있다. 그럼에

도 다소간의 한계를 남겼는데 그 내용을 정리하면 다음과 같다.

본 연구에서 수행한 보유 데이터 전수조사에는 조사대상 46개 기관 중 21개 기관만이 조사에 응답했기 때문에 조사에 참여하지 않은 나머지 25개 기관들이 외부에 개방하지 않은 데이터 정보가 없는 상황이다. 이처럼 다수의 기관이 참여하지 않은 원인에 대해 기관별 담당자 심층인터뷰 및 전문가 자문회의를 진행하여 다음과 같이 정리할 수 있었다.

첫째, 조사 시 부서 간 협력이 제대로 이루어지지 않아 데이터 전수조사에 대한 회신 이 없는 기관이 다수 발생하였다. 둘째, 최근 지속적인 데이터 관련 유사 조사의 중복 요청(국회, 행안부, 문체부 등에서 유사 조사 진행)으로 관련 업무 담당자들의 피로도가 매우 높았으며, 데이터 제공 및 자료 작성을 담당하는 실무자에게 별도의 보상이 없는 부분이 업무 중요도 대비 조사 참여율을 저하 시키는 원인으로 부각되고 있었다. 셋째, 빈번한 담당자 변경으로 업무 연속이 떨어지는 경우에도 수집된 데이터 품질 저하와 더불어 조사 협조도가 낮아지는 것으로 나타났다.

데이터 현황조사 결과를 기반으로 기본적인 빈도분석 및 교차분석은 수행하였으나, 조사결과를 보다 입체적으로 살필 수 있는 질적 분석은 수행하지 못했다. 실제 데이터 전수 결과를 토대로 다양한 추가 정보들이 있었으나, 추가적인 분석까지는 진행하기에는 많은 제약점이 존재하였다. 예컨대, ‘데이터 최초 작성 시점’, ‘데이터 구축 운영 방법’, ‘데이터 확장자’, ‘컬럼명’ 등의 세부 정보를 입수 할 수는 있었다. 하지만 이들 대부분의 정보가 기관별, 담당자별로 데이터 및 용어에 대한 이해정도가 상이하여 실제 분석 자료로 활용하기에는 적절치 못한 경우가 많았다. 따라서 향후 관련 조사 수행 시에는 조사된 결과가 균질화 및 표준화될 수 있도록 데이터 분류, 데이터 정의, 컬럼 정의, 공동활용 데이터 등에 대한 명확한 개념 정의 및 응답 범위 설정에 대한 표준화 작업이 필요할 것으로 보인다.

마지막으로 데이터현황조사 결과를 기반으로 수집·정리된 공동활용 가능 데이터를 토대로 정책분야별 분석활용 과제(안)을 도출할 수 있었으나, 수집된 데이터 정보가 원 데이터가 아닌, 메타데이터 수준으로 실제 데이터를 연계·결합 여부를 판단하기에는 한계가 있었다. 향후 연구 진행시에는 데이터를 활용하고자 하는 목적 및 문제의식을 사전에 명확하게 설정한 후 관련 원 데이터를 입수하여 데이터 연계·결합·활용 가능성을 가늠할 수 있는 연구가 필요하다.

## 제2절 정책 제언

### 가. 데이터 거버넌스로써 제도적 기반 마련

데이터의 공동활용 활성화를 위해서는 연계 가능한 데이터를 확보할 수 있어야 하지만, 실상 유의미한 데이터를 확보하는데 어려움이 있다. 개방된 데이터를 검색하여 찾는 것은 문제가 없으나, 데이터를 보유하고 있는 기관에 데이터의 제공을 요청하는 경우 보유 기관에서 여러 이유로 이를 거부할 시, 해당 데이터를 확보할 수가 없다. 이를 해결하기 위해서 데이터 거버넌스는 데이터의 요청과 제공의 절차가 효과적으로 진행될 수 있도록 제도적인 기반을 마련하고, 지원해야 할 필요가 있다. 데이터를 효율적으로 확보하기 위하여 문화체육관광부 주도 하에 주기적·지속적으로 보유 및 수요 데이터의 목록화 관리가 이루어진다면 문화체육관광부만의 데이터 맵 구축이 가능하고, 이를 통해 데이터 공동활용 가능성 높은 데이터의 소재 확인과 데이터 유형 파악이 용이할 것이다.

향후 공동활용 데이터를 효율적으로 발굴·관리하고 활용하기 위한 기준과 방법 등을 체계화할 수 있는 방안 연구를 추진함으로써 문화체육관광부 차원의 데이터 공동활용에 대한 이해를 돕고, 공동활용 데이터 선정·발굴 과정의 상세한 검토를 통해 기관 간 데이터 공동활용 활성화 방안의 구체적인 역할 및 권한을 제도화·지침화하여야 한다.

### 나. 개인정보보호 문제의 해결방안 모색

데이터의 공동활용을 제한하는 여러 요인 중 빠지지 않고 언급되는 요인 중 하나가 개인정보보호 이슈이다. 정부와 지자체, 공공기관은 업무의 특성 상 국민의 개인정보를 포함한 데이터를 가지고 있는 경우가 대부분이며, 데이터 공동활용을 통하여 업무를 추진하기 위해서는 해당 개인정보를 통해 데이터를 연계·결합이 필요하다. 하지만 데이터 기반행정법에서는 데이터에 개인정보가 포함된 경우 「개인정보 보호법」에 따르도록 명시하고 있으며, 이로 인하여 데이터 보유기관에서는 공동활용을 추진할 수 없다. 따라서

확장성이 높은 데이터 공동활용을 위해서는 기관의 자발적 데이터 등록·신청 프로세스가 아닌 중앙에서 데이터를 결합하여 제공하는 프로세스를 도입하거나 문화체육관광부의 소속 및 산하기관 내에서는 개인정보에 대한 비밀유지 계약을 체결하여 데이터를 제공하는 등의 개인정보보호 문제를 해결해야할 필요가 있다.

#### 다. 데이터 표준화 지원

데이터를 공동으로 활용하기 위해서는 수집된 데이터의 체계가 동일해야하기 때문에 데이터 표준화는 필수불가결한 영역이다. 행정안전부는 ‘공공기관의 데이터베이스 표준화 지침’을 통해 용어, 단어, 도메인 등의 행정코드 표준을 제시하고 있다. 하지만 이미 구축되어 있는 데이터를 표준화하기란 매우 어려운 일이며, 단기적으로 표준화를 추진할 경우 잘못된 표준화를 초래할 수도 있어, 이에 따른 경제적 손실이 클 수 있다. 따라서 단기적인 관점과 장기적인 관점에서의 표준화 방안을 각각 수립할 필요가 있다. 단기적으로 기 구축된 데이터를 중앙으로 수집할 때 표준화된 양식으로 변환하여 수집하되, 이력관리 등에 문제가 발생할 수 있으므로 장기적 관점에서 차세대 시스템 구축 등의 전면 개편을 통해 표준화를 시행할 필요가 있다. 단, 차세대 추진 시에 데이터 표준화는 매우 엄밀한 작업이 필요하기 때문에 정부 차원의 컨설팅 등의 지원이 필요할 것이다.

#### 라. 데이터의 품질관리 및 관련 가이드 제시

데이터를 생성하고 수집 및 관리하는 전 과정에 대한 검토를 통하여 데이터의 품질을 관리할 수 있다. 데이터의 생성 시기, 생성 절차, 관리자, 현행화, 표준화 준수 여부 등 이와 같은 전(全)과정에 걸쳐 모니터링 등을 통해 관리해야 한다.

데이터기반행정부의 관점에서 데이터는 행정업무의 추진, 정책의 수립 및 모니터링 등 공공기관으로서 추진하는 모든 업무에 걸쳐 데이터를 기반으로 이루어져야하기 때문에 데이터에 대한 모든 영역에서의 품질관리가 이루어져야 한다. 이를 위해 데이터를 보유하는 기관은 자체적으로 품질을 관리할 수 있도록 정부차원의 가이드를 제시해야하며, 각 기관에서는 가이드에 따라 데이터의 품질을 관리하고 양질의 데이터를 유지할 수 있도록 해야 한다.

## 마. 데이터 전문인력 확보 및 역량 강화

데이터 공동활용을 위해서는 데이터에 대한 전문성이 필수적으로 요구된다. 데이터 전문성은 데이터 관련 업무의 영역별로 상이한데, 데이터 구축, 분석, 관리, 기반서비스 제공 등으로 구분할 수 있다. 하지만 공공기관에서 이와 같은 전문성 영역별로 모든 인력을 확보하기는 어렵기 때문에 적절한 구성과 배치를 통해 전문영역별 균형감을 갖도록 인력을 확충해야할 필요가 있다.

데이터 전문인력을 확충하는 것 외에도 기존 인력에 대해 데이터 전문성을 제고하는 것도 중요하다. 데이터 공동활용은 정책을 담당하는 모든 공무원 및 공공기관 종사자에게 요구되는 것이기 때문에 업무 수행에서 데이터 마인드가 형성되어야 한다. 따라서 입문·초급 수준의 데이터 인식제고 교육을 시작으로, 전문인력 양성 교육훈련 프로그램까지 단계별, 수준별로 교육을 강화해야한다.

## 바. 데이터기반행정 문화의 정착

데이터기반행정의 성공을 위해서는 당연히 데이터기반행정을 추진하는 조직 내 문화형성이 필요하다. 이러한 데이터기반행정 문화를 확산함에 있어 리더의 중요성을 무시할 수 없다. 데이터에 대한 이해도와 활용도가 높은 리더를 주축으로 데이터 기반의 행정 및 정책을 추진해 간다면 데이터기반행정 문화가 정착하는데 탁월할 것이다. 또한 데이터 관련 업무에 대한 부서 간 협력적·우호적 조직문화를 형성해 나갈 필요가 있다. 본 연구를 수행한 기관별 보유데이터 및 수요데이터 조사 시에도 부서 간 협력이 제대로 이루어지지 않아 데이터 전수조사에 대한 회신이 없는 기관이 다수 발생하였다. 따라서 부서 간 협업이 이루어질 수 있도록 조직문화를 개선해 나가야할 필요가 있다.

---

## 참고문헌

### 1. 국내외 문헌

- 강정묵(2017), 지방정부의 빅데이터 활용에 관한 영향요인 분석: 서울시 공무원의 인식을 중심으로, 국정관리연구, 12(1), 161-197.
- 개인정보보호위원회(2017.12), 데이터 연계·결합 지원제도 도입방안 연구, 개인정보보호위원회.
- 김근령·이대희(2018), 보건의료 빅데이터 활용에 관한 법적 검토: 개인정보보호를 중심으로, 과학기술법연구, 24(3): 57-88.
- 김혜련(2016), 공공 데이터와 민간 빅데이터 연계 방법론 및 사례, 「국가통계방법론심포지엄」, 통계청.
- 권용석·장재호·정군우·전민준(2019), 데이터기반행정 추진전략 및 로드맵 수립, 대구경북연구원.
- 권현영(2020), 데이터기반행정법과 데이터정책의 과제, KISO 저널, (40), 23-28.
- 권효진(2017), 치안분야에서의 Big Data 활용 사례와 바람직한 공공 연구조직 설계, 「한국기술혁신학회 추계학술대회」, 1245-1262.
- 박선우(2018), 빅데이터 시대와 데이터 융합, 정보통신방송정책, 30(1), 1-24.
- 배종면(2016), 결정적 자료연계에서 확률적 자료연계로, 대한보건연구, 42(3), 23-27.
- 손현(2015), 증거기반 과학행정 구현을 위한 데이터 관련 법제 정비방안 연구, 정부 3.0추진위원회.
- 손현(2017), 빅데이터기반의 과학행정 구현을 위한 입법법안 연구, 한국법제연구원
- 송명원·김상범(2006), 정보자원통합의 현황과 전망. 전자정부포커스, 2006(3), 1-21.
- 송석현·이삼열·신열·이재용(2017), 한국의 공공데이터 개방 정책의 효과에 대한 연구, 한국지역정보학회지, 20(4), 1-34.
- 신광영(2017), 스웨덴의 행정 데이터 통합과 활용에 관한 연구, 스칸디나비아 연구, 20, 83-108.
- 오미애(2015.9), 보건복지분야 데이터 연계 필요성 및 활용방안, 「보건복지포럼」, 제 227권, 17-28.

- 오세영·윤건·오균(2017), 증거기반정책을 위한 정부의 통계 구축 및 활용에 대한 현황 조사, 한국행정연구원.
- 오철호(2017), 문제제기: 데이터 기반 정책분석평가의 연구와 적용, 정책분석평가학회보, 27(2): 155-167.
- 유종성·전병유·신광영·이도훈·최성수(2020), 증거기반 정책연구를 위한 행정자료의 활용, 한국사회정책, 27(1), 5-37.
- 윤건(2018), 증거를 통하여 정책을 세워라, 통계의 창(통계청 통계교육원), 2018 겨울호.
- 윤건(2019), 데이터기반행정 강화 방안 연구: 공공데이터 융합(integration)을 중심으로, 한국행정연구원.
- 윤건·김윤희(2020), 공공데이터 융합의 영향요인과 매개효과에 관한 탐색적 연구: 데이터기반행정의 맥락을 중심으로, 한국지역정보학회지, 23(3), 1-22.
- 윤건·김철우(2020), 지자체 차원의 데이터기반행정 구현을 위한 요인 분석 -핵심 구성요소와 주요 참여자를 중심으로-, 한국지방자치학회보, 32(4), 71-101.
- 윤건·김철우(2020. 12), 한국지방자치학회보, 제32권 제4호(통권112호).
- 윤광석·이건(2017), 공공데이터 활용 행정 촉진을 위한 거버넌스 모색: 영국의 ADRN 사례를 중심으로, 국가정책연구, 31(1), 1-27.
- 윤상오·현지우(2019), 공공데이터 개방정책의 실태분석 및 개선방안에 관한 연구, 한국공공관리학보, 33(1), 219-247.
- 윤희정(2014), 조직적, 기술적, 환경적 요인이 빅데이터의 활용 및 효과에 미치는 영향, 경영관리연구, 7(2), 153-178.
- 이민호(2015), 정보공유 및 시스템 연계·통합 추진 효과성의 영향요인 탐색, 한국사회와 행정연구, 26(2), 23-52.
- 이민호(2015), 정보공유 및 시스템 연계·통합 활성화 방안, 한국행정연구원, ISSUE PAPER, 통권 32 2015-11.
- 이신·허유경·김혜미(2017), 빅데이터를 이용한 교통계획: 심야버스와 사고줄이기, 서울시 교통정책 패키지.
- 이원태(2016), 지속가능한 공공데이터 관리체계 연구, 한국정보화진흥원.
- 이재관(2004), 일본의 지능형 교통 시스템 ITS, 자동차공학회지, 2004년 10월호 제26권 제5호, 104-105.
- 한국과학기술정보연구원(2017), 과학기술 빅데이터 공유·융합체제 구축, 미래창조과학부.
- 한국데이터진흥원(2017.10), CKAN을 활용한 민간데이터 연계 활성화 방안 연구 결과 보고서, 한국데이터진흥원.



- 한국보건사회연구원(2015), 「보건복지포럼」, 2015년 09월호(통권 제227호)
- 한은희(2020), 영국의 연구를 목적으로 한 행정 데이터 연계 활용, 국제사회보장리뷰 2020년 가을 14호, 38-52.
- ADR UK.(2019), Administrative data is an invaluable resource for public good. Let's use it, Annual Report 2018-19, Retrieved from [https://www.adruk.org/fileadmin/uploads/adruk/Documents/ADRUK\\_Annual\\_Report\\_2018-19.pdf](https://www.adruk.org/fileadmin/uploads/adruk/Documents/ADRUK_Annual_Report_2018-19.pdf)
- Allard, S.W., Wiegand, E.R., Schlecht, C., Datta, A.R., Goerge, R.M., & Weigensberg, E.(2018). State agencies' use of administrative data for improved practice: Needs, challenges, and opportunities. Public Administrative Review, 78, 240-250.
- Mc Grath-Lone L, Blackburn R, Gilbert R. The Education and Child Health Insights from Linked Data (ECHILD) database: an introductory guide for researchers. London: University College London, 2021.
- Michael Guentherman, BS1,2, Ellie Hartman, PhD1, Amanda Schlegelmilch, MS3, Emily A. Brinck, PhD4, and Catherine A. Anderson, PhD5.
- Pew Charitable Trusts.(2018). How to state use data to inform decisions: A national review of the use of administrative data to improve state decision-making. Retrieved from [https://www.pewtrusts.org/-/media/assets/2018/02/dasa\\_how\\_states](https://www.pewtrusts.org/-/media/assets/2018/02/dasa_how_states)
- Risk Assessment and Horizon Scanning Experimentation Centre <https://www.dsta.gov.sg/docs/default-source/dsta-about/risk-assessment-and-horizon-scanning-experimentation-centre.pdf?sfvrsn=2>

## 2. 정부 자료

- 경찰청(2019.11.29.), 빅데이터를 통한 범죄예측, 첫 발을 내딛다(보도자료).
- 공공데이터전략위원회(행정자치부)(2016.2.), 공공 빅데이터 활용 활성화 추진계획(안).
- 공공데이터전략위원회(행정자치부)(2016.12.), 제2차('17~'19) 공공데이터 기본계획.
- 관계부처합동(과학기술정보통신부)(2019.1.) 데이터·AI경제 활성화 계획.
- 대한민국 정책브리핑(2020), 정책 위키, 「데이터경제」.
- 청와대(2018.8.31.), 데이터경제 활성화 규제혁신 현장방문 인사말.
- 한국지능정보사회진흥원(2021.6.), 공공데이터베이스 표준화 관리 매뉴얼.
- 행정안전부(2016.6.30.), 「개인정보 비식별 조치 가이드라인」 발간(보도자료).

- 행정안전부(2017.6.20.), 빅데이터로 자동차 보험사기, 병역면탈 근절(보도자료).
- 행정안전부(2017.7.28.), 온 마을이 나서서 아이를 돌보시다(보도자료).
- 행정안전부(2020.5.19.), 데이터기반행정 활성화에 관한 법률 제정(안) 입법예고.
- 행정안전부(2020.9.3.), 빅데이터가 만드는 똑똑한 행정. 표준분석모델로 더욱더 앞당긴다(보도자료).
- 행정안전부(2020.12.10.), 데이터를 기반으로 한 과학적 행정시대 본격화(보도자료).
- 행정안전부(2021), 제1차 데이터기반행정 활성화 기본계획(안).
- 행정안전부(2021.9), 제3차 「데이터기반행정 책임관 협의회」 개최 계획

### 3. 법률·고시 등

- 공공데이터 관리지침[행정안전부고시 제2019-71호]
- 공공기관의 데이터베이스 표준화 지침[행정안전부고시 제2021-32호]
- 공공데이터의 제공 및 이용 활성화에 관한 법률[법률 제17344호]
- 데이터기반행정 활성화에 관한 법률[법률 제17370호]
- 문화체육관광부 데이터 관리 규정[문화체육관광부훈령 제430호]
- 행정기관의 코드표준화 추진지침[행정안전부고시 제2021-6호]

### 4. 인터넷 자료

- 대한민국 정책브리핑, 데이터 3법(2020.3.30.)  
<https://www.korea.kr/special/policyCurationView.do?newsId=148867915>
- 대한민국 정책브리핑, 데이터 경제(2020.3.13.)  
<https://www.korea.kr/special/policyCurationView.do?newsId=148863563>
- 대한민국 정책브리핑, 청와대, 데이터경제 활성화 규제혁신 현장방문 관련 고민정 부대변인 서면 브리핑(2018.08.31.)  
<https://www.korea.kr/news/blueHouseView.do?newsId=148853630&pageIndex=1>
- 데이터 기반 비즈니스에 '데이터 카탈로그'가 필요한 이유(2021.05.30.)  
<https://www.ciokorea.com/news/141452>
- 일요서울i,(2020.02.10.), 창원시, 체납차량 빅데이터 분석...영치업무의 과학화  
<http://www.ilyoseoul.co.kr/news/articleView.html?idxno=366603>

---

# ABSTRACT

## **A Study on the Building of the Joint Utilization of Culture, Tourism, and Content Data**

Song, ChulJae

Act on the Promotion of Data-Driven Administration set three essential strategies to push forward, and they are promoting the joint utilization of data among agencies, establishment of a system for the activation of data-driven administration, and innovation of data-based working methods. The ultimate goals of the Act are to increase the accountability and reliability of public agencies and improve the quality of life for the people through scientific administration based on joint data utilization.

As it became an institution and legislation to promote the joint utilization of data among different agencies along with the exchange of data collected and possessed by public agencies, the central government departments and public agencies began to contemplate over what to exchange. It was the same with institutions and public agencies under the supervision of Ministry of Culture, Sports and Tourism in charge of producing, managing, and operating statistics and data related to culture, arts, content, tourism, and sports. Although massive amounts of public data are being accumulated thanks to the input of big budgets into data management by these institutions and public agencies, there are no grounds and systems to enable joint data utilizations due to the closed and exclusive management of each agency until recently and the lack of department-wide data sharing systems. Following the implementation of Act on the Promotion of Data-Driven Administration, it is thus important to request, provide, and utilize jointly data among different agencies by getting rid of the partitions between them and search for scientific plans to push

forward data-driven administration. Furthermore, it is emerging as an important task to build a foundation for people's happiness and improved quality of life by driving scientific administration whose objectivity and reliability is guaranteed instead of uniform administration driven by the old practices and past experiences.

This study aimed to establish a basis for the active joint utilization of data about culture, tourism, and content to reinforce data-driven administration. For this purpose, the study analyzed policy environments and cases and examined the current state of data. Using the results, the investigator proposed policy plans for the activation of joint data utilization. Theoretical discussions covered the theoretical significance of data administration by analyzing the conditions and environments to promote data-driven administration. As there was a growing demand for scientific and rational policy processes such as evidence-based policies, the need for data-driven administration was also increasing.

### **Keywords**

data, public data, data-driven administration, examination of the current state of data, joint data utilization

문화·관광·콘텐츠 데이터 공동활용 기반 구축 연구

부록

## 데이터 업무 활용 실태조사 조사표



## 부록. 데이터 업무 활용 실태조사 조사표

[조사근거] 「데이터기반행정 활성화에 관한 법률」제20조, 「문화체육관광부 데이터 관리 규정」제12조(문화체육관광부훈령 제430호)

### 문화체육관광 데이터 업무 활용 실태조사

안녕하십니까?

먼저 귀한 시간을 내어 본 설문에 응해주셔서 대단히 감사합니다.

문화체육관광부와 한국문화관광연구원은 「데이터기반행정 활성화에 관한 법률」제20조와 「문화체육관광부 데이터 관리 규정」제12조(문화체육관광부훈령 제430호)에 근거, 귀하가 소속된 기관의 정책 수립, 분석업무와 관련하여 데이터 활용 경험 및 제공 경험을 조사하고 있습니다.

귀하의 응답은 데이터기반행정 활성화를 위한 귀중한 자료로 활용될 예정이오니, 부디 빠짐없이 답해 주시기 바랍니다. 고맙습니다.

2021년 8월



한국문화관광연구원 정책정보센터  
데이터분석팀 송 철 재 책임전문위원  
통계관리팀 장 선 녀 차석전문위원  
(☎ 02-2669-8955)



주리서치랩  
연구1부 이 재 군 이사  
연구1부 이 정 진 연구원  
(☎ 02-3015-2125)

SQL) 귀하는 최근 1년 간(2020년 7월 ~ 2021년 6월) 업무 수행과 관련하여 데이터를 활용한 경험이 있습니까?

1. 예

2. 아니오(※조사 종료)

### I. 데이터 기반 행정 실태

■ 최근 1년 간(2020년 7월 ~ 2021년 6월) 업무 수행 시 귀하의 데이터 활용 경험을 토대로 응답해 주십시오.

문1) 귀하께서 주로 다룬 데이터 유형은 무엇입니까?

문1-1) 데이터 기본 유형(※ 해당 항목에 모두 응답해 주십시오.)

1. 통계조사 자료

2. 통계조사 자료를 제외한 수량(수치) 자료

3. 텍스트

4. 기타(적어 주세요: \_\_\_\_\_)

문1) 귀하께서 주로 다룬 데이터 유형은 무엇입니까?

문1-2) 정형/비정형 데이터

1. 정형 데이터

2. 비정형 데이터

3. 정형 + 비정형 데이터

**문2) 귀하가 주로 다루는 데이터 형태는 다음 중 어떻게 되십니까?**

1. 단일 데이터(**※문2-1**)                      2. 결합 데이터(**※문2-2**)

**문2-1) 귀하가 주로 다루는 단일 데이터 유형은 다음 중 어떻게 되십니까?**

**문2-1-1) 공공/민간 데이터**(※ 해당 항목에 모두 응답해 주십시오.)

1. 공공 데이터                      2. 민간 데이터

**문2-1) 귀하가 주로 다루는 단일 데이터 유형은 다음 중 어떻게 되십니까?**

**문2-1-2) 원시/가공 데이터**(※ 해당 항목에 모두 응답해 주십시오.)

1. 원시 데이터                      2. 가공 데이터

**문2-2) 귀하가 주로 다루는 결합 데이터 유형은 다음 중 어떻게 되십니까?**

1. 공공 + 공공 데이터    2. 공공 + 민간 데이터    3. 민간 + 민간 데이터(**※문7**)

**문3) 귀하가 다루는 공공 데이터의 제공 기관은 다음 중 어디에 속합니까?**

(※ 해당 항목에 모두 응답해 주십시오.)

1. 본인 기관 (**※문7**)  
2. 문화체육관광부 소속 및 산하기관  
3. 정부부처(중앙행정기관)  
4. 타부처 소속 및 산하기관  
5. 광역·기초자치단체  
6. 기타(적어 주세요: \_\_\_\_\_)

**문4) 귀하가 다루는 공공 데이터의 정책분야는 다음 중 어디에 속합니까?**

(※ 해당 항목에 모두 응답해 주십시오.)

- |                 |                       |
|-----------------|-----------------------|
| 1. 공공질서 및 안전 분야 | 2. 과학기술 분야            |
| 3. 교육 분야        | 4. 교통 및 물류 분야         |
| 5. 국방 분야        | 6. 농림 분야              |
| 7. 문화체육관광 분야    | 8. 보건 분야              |
| 9. 사회복지 분야      | 10. 산업통상중소기업 분야       |
| 11. 일반공공행정 분야   | 12. 재정세제금융 분야         |
| 13. 지역개발 분야     | 14. 통신 분야             |
| 15. 통일외교 분야     | 16. 해양수산 분야           |
| 17. 환경 분야       | 18. 기타(적어 주세요: _____) |



■ 최근 1년 간(2020년 7월 ~ 2021년 6월) 타 기관의 데이터 활용 경험을 토대로 응답해 주십시오

(문3의 2,3,4,5,6 응답자에게만 질문)

문5) 귀하가 다루는 타 기관 데이터의 이용 경로는 무엇입니까?(※ 해당 항목에 모두 응답해 주십시오)

1. 행정안전부 포털(홈페이지 포함)
2. 문화체육관광부 소속 및 산하기관 포털(홈페이지 포함)
3. 타 기관에서 보유하고 있는 포털(홈페이지 포함)
4. 개별 기관에 직접 요청
5. 데이터 구매
6. 기타(적어 주세요: \_\_\_\_\_)

(문3의 2,3,4,5 응답자에게만 질문)

문5-1) 타 기관의 데이터 이용 시 데이터 제공이 얼마나 원활하게 이루어졌습니까?

1. 전혀 원활하지 않음(☞문5-2)
2. 별로 원활하지 않음(☞문5-2)
3. 보통임(☞문6)
4. 약간 원활했음(☞문6)
5. 매우 원활했음(☞문6)

(문5-1의 1, 2 응답자에게만 질문)

문5-2) 데이터 제공이 원활하지 않았던 이유는 무엇이었습니까?

(※ 해당 항목에 모두 응답해 주십시오)

- |          |                      |
|----------|----------------------|
| 1. 국가 기밀 | 2. 개인정보 보호           |
| 3. 규정 없음 | 4. 기타(적어 주세요: _____) |

문6) 귀하께서 타 기관으로부터 데이터를 제공받았을 때, 평균적으로 어느 정도의 기간이 소요됐습니까?

- |             |             |
|-------------|-------------|
| 1. 요청 즉시    | 2. 15일 이내   |
| 3. 1개월 이내   | 4. 1~3개월 미만 |
| 5. 3~6개월 미만 | 6. 6개월 이상   |

■ 최근 1년 간(2020년 7월 ~ 2021년 6월) 내부/외부 기관에 대한 데이터 제공 경험을 토대로 응답해 주십시오

문7) 귀하의 기관이 보유하고 계신 데이터에 대해 외부 기관으로부터 데이터 제공 요청을 받은 경험이 있습니까?

- |            |              |
|------------|--------------|
| 1. 없다(☞문9) | 2. 있다(☞문7-1) |
|------------|--------------|

(문7의 2 응답자에게만 질문)

문7-1) 외부 기관에 대한 데이터 제공은 얼마나 원활하게 이루어졌습니까?

1. 전혀 원활하지 않음(☞문7-2)
2. 별로 원활하지 않음(☞문7-2)
3. 보통임(☞문8)

4. 약간 원활했음(※문8)

5. 매우 원활했음(※문8)

(문7-1의 1, 2 응답자에게만 질문)

문7-2) 데이터 제공이 원활하지 않았던 이유는 무엇이었습니까?

(※ 해당 항목에 모두 응답해 주십시오.)

1. 국가 기밀

2. 개인정보 보호

3. 규정 없음

4. 기타(적어 주세요: \_\_\_\_\_)

문8) 귀하의 기관이 외부 기관에 데이터를 제공하기까지 평균적으로 어느 정도의 기간이 소요됐습니까?

1. 요청 즉시

2. 15일 이내

3. 1개월 이내

4. 1~3개월 미만

5. 3~6개월 미만

6. 6개월 이상

■ 최근 1년 간(2020년 7월 ~ 2021년 6월) 데이터 교육 경험을 토대로 응답해 주십시오.

(※ 교육이 2개 이상일 경우 가장 최근에 참여하신 경험을 기준으로 작성하여 주십시오.)

문9) 귀하께서는 최근 1년 동안 데이터 교육을 받은 경험이 있습니까?

1. 없다(※문10)

2. 있다(※문9-1)

(문9의 2 응답자에게만 질문)

문9-1) 귀하께서 최근 1년 동안 받았던 데이터 교육은 어디에서 주관했습니까?

1. 재직 중인 기관 교육

2. 외부 기관

3. 재직 중인 기관 + 외부 기관

(문9의 2 응답자에게만 질문)

문9-2) 귀하께서 최근 1년 동안 받았던 데이터 교육 방식(환경) 선택하여 주시기 바랍니다.

1. 온라인 교육

2. 오프라인 교육

3. 온라인 + 오프라인 교육

(문9의 2 응답자에게만 질문)

문9-3) 귀하께서 최근 1년 동안 받았던 데이터 교육 방법은 어떻게 됩니까?

(※ 해당 항목에 모두 응답해 주십시오.)

1. 강의 중심

2. 토론 중심

3. 실습 중심

4. 기타(적어 주세요: \_\_\_\_\_)

(문9의 2 응답자에게만 질문)

문9-4) 최근 1년 동안 데이터 교육 내용은 어떻게 됩니까?

(※ 해당 항목에 모두 응답해 주십시오.)

1. 데이터 구축 방법

2. 데이터 분석 방법

3. 데이터 관리 방법

4. 데이터 기반 서비스 제공 방법

5. 기타(적어 주세요: \_\_\_\_\_)

☞ 보기카드

문10) 귀하께서는 데이터 관련 자격증을 보유하고 있습니까?

1. 없다(☞문11)

2. 있다(☞문10-1)

(문10의 2 응답자에게만 질문)

문10-1) 귀하께서 보유하고 계신 데이터 관련 자격증은 몇 개입니까?

보유개수 \_\_\_\_\_ 개

(문10의 2 응답자에게만 질문)

문10-2) 귀하께서 보유하고 계신 데이터 관련 자격증은 다음 항목 중 어디에 해당하십니까?

(※ 해당 항목에 모두 응답해 주십시오.)

1. 프로그래밍

2. 시스템

3. 네트워크

4. 데이터베이스

5. 보안/해킹

6. 통계데이터분석

7. 기타(적어 주세요: \_\_\_\_\_)

## II. 데이터 기반 행정 인식

■ 귀하의 담당업무와 귀 기관의 데이터 기반 행정과 관련하여 생각하시는 바를 응답해 주십시오.

문11) 귀하께서는 담당 업무를 수행하면서 데이터 활용이 얼마나 중요하다고 생각하십니까?

1. 전혀 중요하지 않다(☞문12)

2. 별로 중요하지 않다(☞문12)

3. 보통이다(☞문11-1)

4. 약간 중요하다(☞문11-1)

5. 매우 중요하다(☞문11-1)

(문11의 3, 4, 5 응답자에게만 질문)

문11-1) 귀하께서는 다음 업무 유형별로는 데이터를 활용하는 것이 얼마나 중요하다고 생각하십니까?

항목	전혀 중요하지 않다	별로 중요하지 않다	보통이다	약간 중요하다	매우 중요하다
1) 정책기획	1	2	3	4	5
2) 정책집행	1	2	3	4	5
3) 정책평가	1	2	3	4	5

문12) 귀하께서는 담당 업무를 수행하면서 데이터를 어느 정도 활용하고 있다고 생각하십니까?

1. 전혀 활용하지 않는다
2. 별로 활용하지 않는다
3. 보통이다
4. 약간 활용한다
5. 매우 많이 활용한다

문13) 다음 업무 유형별로는 담당 업무를 하면서 데이터를 어느 정도 활용하고 있다고 생각하십니까?

항목	전혀 활용하지 않다	별로 활용하지 않다	보통이다	약간 활용한다	매우 많이 활용한다	해당없음
1) 정책기획	1	2	3	4	5	8
2) 정책집행	1	2	3	4	5	8
3) 정책평가	1	2	3	4	5	8

문14) 귀하께서는 데이터 기반 행정 활성화를 위해 다음 각 항목들이 어느 정도 중요하다고 생각하십니까?

항목	전혀 중요하지 않다	별로 중요하지 않다	보통이다	중요하다	매우 중요하다
1) 정책에 활용할 데이터 관련 기획	1	2	3	4	5
2) 데이터 관련 예산 확보	1	2	3	4	5
3) 데이터 관련 전문인력 확보	1	2	3	4	5
4) 데이터 관련 교육 강화	1	2	3	4	5
5) 데이터 관련 장비·시설 확충	1	2	3	4	5
6) 데이터 관련 부서 간 협력	1	2	3	4	5
7) 조직 리더의 데이터 역량 강화	1	2	3	4	5
8) 일반 조직 구성원의 데이터 역량 강화	1	2	3	4	5
9) 데이터를 중시하는 조직 문화	1	2	3	4	5

문15) 귀 기관은 데이터 기반 행정 활성화를 위해 다음 각 항목들이 현재 어느 정도 이뤄지고 있다고 생각하십니까?

항목	전혀 충분하지 않다	별로 충분하지 않다	보통이다	충분하다	매우 충분하다
1) 정책에 활용할 데이터 관련 기획	1	2	3	4	5
2) 데이터 관련 예산 확보	1	2	3	4	5
3) 데이터 관련 전문인력 확보	1	2	3	4	5
4) 데이터 관련 교육 강화	1	2	3	4	5
5) 데이터 관련 장비·시설 확충	1	2	3	4	5
6) 데이터 관련 부서 간 협력	1	2	3	4	5
7) 조직 리더의 데이터 역량 강화	1	2	3	4	5
8) 일반 조직 구성원의 데이터 역량 강화	1	2	3	4	5
9) 데이터를 중시하는 조직 문화	1	2	3	4	5

### III. 데이터 공동활용 실태

■ 최근 1년 간(2020년 7월 ~ 2021년 6월) 귀하의 데이터 공동활용 경험을 토대로 응답해 주십시오

문16) 귀하께서는 데이터를 공동활용한 정책이나 사업에 참여하신 경험이 있습니까?

1. 없다(문17)

2. 있다(문16-1)

(문16의 2 응답자에게만 질문)

문16-1)~9) 데이터 공동활용 사업 관련 정보

(※ 사업이 2개 이상일 경우 가장 최근에 참여하신 경험을 기준으로 작성하여 주십시오.)

항목	내용
1) 사업명	
2) 사업목적	<div style="display: flex;"> <div style="flex: 1;">(     )</div> <div style="flex: 2;">           1. 데이터 구축            2. 데이터 분석            3. 데이터 관리            4. 데이터 기반 서비스 제공            5. 기타(적어 주세요: _____)            (※ 해당 항목에 모두 응답해 주십시오.)            보기카드         </div> </div>
3) 사업 기간	(     )년 (     )개월
4) 사업 예산	(     )만원
5) 투입 인력	(     )명
6) 공공데이터 공동활용 경험 및 데이터 보유 주체명	<div style="display: flex;"> <div style="flex: 1;"> <input type="checkbox"/> 있음  <input type="checkbox"/> 없음         </div> <div style="flex: 2;">           (     ), (     ), (     )            (예:국세청, 건강보험심사평가원, 한국고용정보원) 누락         </div> </div>
7) 민간데이터 공동활용 경험 및 데이터 보유 주체명	<div style="display: flex;"> <div style="flex: 1;"> <input type="checkbox"/> 있음  <input type="checkbox"/> 없음         </div> <div style="flex: 2;">           (     ), (     ), (     )            (예:KT, 신한카드) 누락         </div> </div>
8) 공동활용 정책분야 (※ 해당 항목에 모두 응답해 주십시오.)	<div style="display: flex; flex-wrap: wrap;"> <div style="flex: 50%;">1. 공공질서 및 안전 분야</div> <div style="flex: 50%;">2. 과학기술 분야</div> <div style="flex: 50%;">3. 교육 분야</div> <div style="flex: 50%;">4. 교통 및 물류 분야</div> <div style="flex: 50%;">5. 국방 분야</div> <div style="flex: 50%;">6. 농림 분야</div> <div style="flex: 50%;">7. 문화·체육·관광 분야</div> <div style="flex: 50%;">8. 보건 분야</div> <div style="flex: 50%;">9. 사회복지 분야</div> <div style="flex: 50%;">10. 산업·통상·중소기업 분야</div> <div style="flex: 50%;">11. 일반공공행정 분야</div> <div style="flex: 50%;">12. 재정·세제·금융 분야</div> <div style="flex: 50%;">13. 지역개발 분야</div> <div style="flex: 50%;">14. 통신 분야</div> <div style="flex: 50%;">15. 통일·외교 분야</div> <div style="flex: 50%;">16. 해양수산 분야</div> <div style="flex: 50%;">17. 환경 분야</div> <div style="flex: 50%;">18. 기타(적어 주세요:     )</div> </div>
9) 데이터 확보 채널 (※ 해당 항목에 모두 응답해 주십시오.)	1. 행정안전부 포털 2. 문화체육관광부 소속 및 산하기관 포털 3. 타 부처, 소속·기관 및 지자체에서 보유하고 있는 포털 4. 개별 기관에 직접 요청 5. 데이터 구매 6. 기타(적어 주세요: _____)

#### IV. 데이터 공동활용 인식

문17) 귀하께서는 데이터 기반 행정 강화를 위해 데이터 공동활용이 얼마나 중요하다고 보십니까?

1. 전혀 중요하지 않다
2. 별로 중요하지 않다
3. 보통이다
4. 약간 중요하다
5. 매우 중요하다

문18) 귀하께서는 현재 기관 간의 데이터 공동활용이 얼마나 잘 이루어지고 있다고 보십니까?

1. 전혀 이루어지지 않고 있다
2. 별로 이루어지지 않고 있다
3. 보통이다
4. 약간 이루어지고 있다
5. 매우 잘 이루어지고 있다

문19) 귀하께서는 데이터 공동활용에서 다음 제약조건이 얼마나 심각하다고 생각하십니까?

항목	전혀 심각하지 않다	별로 심각하지 않다	보통이다	약간 심각하다	매우 심각하다
1) 개인정보보호 문제	1	2	3	4	5
2) 데이터 담당자의 책임성 문제	1	2	3	4	5
3) 데이터 정확성 확보의 어려움	1	2	3	4	5
4) 데이터 보유기관의 이기주의	1	2	3	4	5
5) 데이터 제공 관련 법규정 미흡	1	2	3	4	5
6) 데이터 간 표준화 미흡	1	2	3	4	5
7) 데이터 공동활용을 위한 전문인력 부족	1	2	3	4	5
8) 데이터 공동활용 관련 교육 지원 미흡	1	2	3	4	5
9) 정책 수행에 필요한 데이터 검색 어려움	1	2	3	4	5
10) 정책 수행에 필요한 데이터 미흡	1	2	3	4	5

문20) 귀하께서 데이터 공동활용 활성화를 위해 필요하다고 생각하는 부분이 있다면 자유롭게 기술해 주십시오

## V. 응답자 특성

DQ1) 귀하의 전공은 어떻게 되십니까?

- |                      |                |
|----------------------|----------------|
| 1. 행정학               | 2. 통계학         |
| 3. 경제학               | 4. 경영학(MIS 포함) |
| 5. 전산/컴퓨터공학          | 6. 사회학         |
| 7. 기타(적어 주세요: _____) |                |

DQ2) 귀하의 근무처는 어디입니까?

(적어 주세요: \_\_\_\_\_)

DQ3) 귀하의 직무는 어떻게 되십니까?

- |                      |       |
|----------------------|-------|
| 1. 행정                | 2. 통계 |
| 3. 전산/정보             | 4. 연구 |
| 5. 기타(적어 주세요: _____) |       |

DQ4) 귀하의 데이터 관련 근무경력은 어떻게 되십니까?

- |           |          |
|-----------|----------|
| 1. 1년 미만  | 2. 1~2년  |
| 3. 3~5년   | 4. 6~10년 |
| 5. 10년 이상 |          |

◎ 응답해 주셔서 감사합니다.





## 집필내역

---

### 연구책임

송철재 한국문화관광연구원 책임전문위원: 제1장, 제4장, 제6장, 연구총괄

### 연구진

김수경 한국문화관광연구원 책임전문위원: 제2장, 제4장 제3절, 제5장 제2절

장선녀 한국문화관광연구원 차석전문원: 제3장, 제4장 제4절~제5절 일부

김정림 한국문화관광연구원 차석전문원: 제5장

## 문화·관광·콘텐츠 데이터 공동활용 기반 구축 연구

---

발행인 김 대 관

발행처 한국문화관광연구원

서울시 강서구 금남화로 154

전화 02-2669-9800 팩스 02-2669-9880

<http://www.kcti.re.kr>

인쇄일 2021년 10월 15일

발행일 2021년 10월 15일

인쇄인 (사)한국장애인이워크협회 일자리사업장

I S B N 978-89-6035-871-3 93300

DOI <https://doi.org/10.16937/kcti.rep.2021.e9>

이 연구보고서를 인용하실 때는 다음과 같은 사항을 기재해 주십시오.

송철재(2021), 문화·관광·콘텐츠 데이터 공동활용 기반 구축 연구, 한국문화관광연구원

한국문화관광연구원

서울특별시 강서구 금남화로 154

전화 02-2669-9800

팩스 02-2669-9880

[www.kcti.re.kr](http://www.kcti.re.kr)



아래의 DOI 또는 QR코드를 통해  
이 보고서를 무료로 다운로드할 수 있습니다.  
<https://doi.org/10.16937/kcti.rep.2021.e9>



9 788960 358713  
ISBN 978-89-6035-871-3