



# 지자체 현안해결 지원을 위한 데이터셋 연계전략

신수미

복잡한 사회현안의 이해와 해결을 위해 각 분야의 여러 데이터를 다각적으로 분석하고자 하는 요구사항이 증가하고 있으며 공공분야를 비롯해 민간분야 데이터까지 다양하게 연계해서 활용하는 사례도 늘어가고 있다. KISTI는 「AI 기술을 활용한 공공데이터 기반 지역현안 솔루션 개발 및 상용화 사업」을 통해 데이터셋을 효율적으로 연계하여 활용할 수 있는 데이터셋 연계전략을 수립하였다. KISTI의 데이터셋 연계전략은 공공데이터포털의 데이터셋 57,421종의 항목명 200만 여개와 지진, 침수, 미세먼지 연구분야에서 활용된 연계 데이터셋 사례를 분석하여 마련되었으며, 항목명 표준화와 데이터값 표준화를 기반으로 한다. 본 이슈브리프에서는 KISTI에서 수립한 데이터셋 연계전략과 함께 실제 데이터셋 연계를 위해 추진한 항목명 표준화 및 데이터값 표준화 방안과 이를 통한 데이터셋 연계결과를 소개한다. 소개되는 데이터셋 연계전략은 여러 데이터셋을 동시에 활용하고자 하는 모든 분야에서 적용할 수 있는 범용적인 방안이 될 수 있을 것으로 기대된다. 향후 KISTI는 데이터셋 연계의 핵심 전략과 기술을 보다 발전시켜 데이터셋의 공동활용성을 제고하고 이를 통해 데이터셋의 활용과 확산에 기여할 계획이다.

## CONTENTS

### 1. 데이터셋 연계 개요

- 데이터셋 연계의 필요성
- 데이터셋 연계사례
- KISTI의 데이터셋 연계전략

### 2. 항목명 표준화 기반 데이터셋 연계

- 데이터셋 연계방안
- 연계절차 및 결과

### 3. 데이터값 표준화 기반 데이터셋 연계

- 데이터셋 연계방안
- 연계절차 및 결과

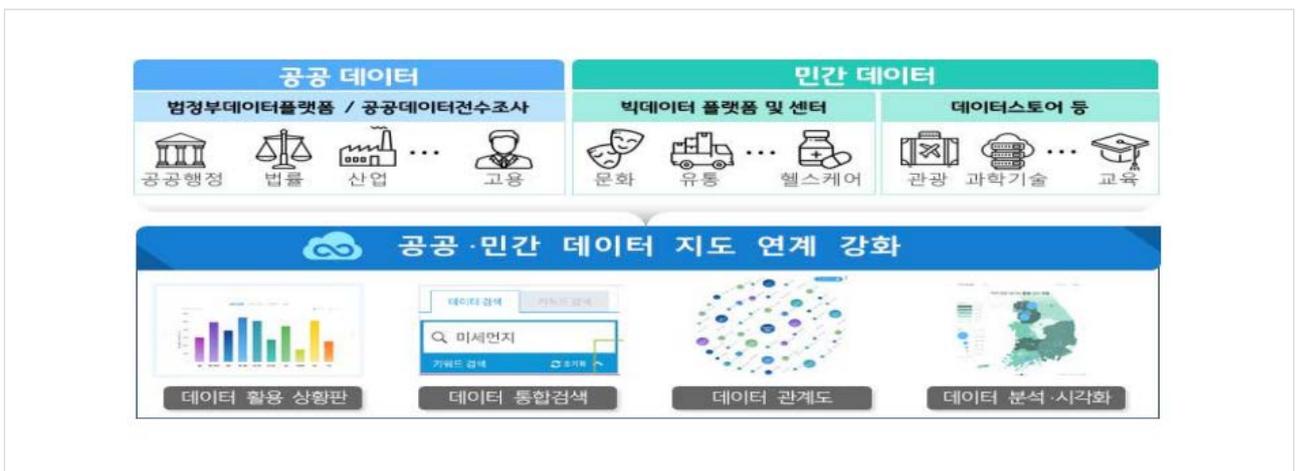
### 4. 맺음말과 제언

# 1. 데이터셋 연계 개요

## ▶ 데이터셋<sup>1)</sup> 연계<sup>2)</sup>의 필요성

- **(데이터의 정확성 향상)** 데이터 연계는 여러 출처의 정보를 결합하여 데이터의 정확성과 완전성을 향상시키는 데 도움이 됨
  - 여러 데이터셋을 연계하여 동시에 활용할 경우 단일 데이터셋만으로는 분명하지 않은 관계와 패턴을 식별할 수 있어 데이터 분석 시 정확하고 안정적인 결과 도출 가능
  - 불완전한 정보가 해소된 새로운 정보의 생산이 가능하며 기존 데이터의 검증 가능성도 향상
- **(부가가치 창출을 위한 데이터 요구 확대)** 경제·사회적 부가가치 창출과 기술 경쟁력 확보를 위해 데이터 융합 및 연계 강조
  - 지자체를 중심으로 자체 보유한 데이터셋과 카드 이용 정보 등 민간데이터셋을 함께 분석하여 지역상권 분석정보나 맞춤형 관광지 추천정보를 제공하는 등 지역경제 활성화를 위해 데이터를 연계해 활용하는 사례 증가
  - 2020년 제정된 「데이터기반행정 활성화에 관한 법률<sup>3)</sup>」에서는 객관적이고 과학적인 행정을 위하여 데이터 연계·제공 및 공동활용에 대한 노력을 공공기관의 책무로 정의
  - 인공지능 국가전략<sup>4)</sup> 인공지능 국가전략(관계부처 합동, 2019)에서는 AI 경쟁력 혁신을 위하여 공공데이터 지도와 민간데이터 지도의 연계가 필요하다고 정의

<그림 1> 공공·민간 데이터 지도 연계 계획



출처) 인공지능 국가전략, 관계부처 합동(2019)

1) 컴퓨터가 처리하거나 분석할 수 있는 형태로 존재하는 관련 정보의 집합체. 한 지역 내의 상이한 지점에서 오랜 기간에 걸쳐 일정한 시간적 간격을 두고 관측한 습도나 온도 수치의 집합이 그 예이다. 학생 기록과 같은 데이터는 기록의 대상이 되는 객체의 특성을 공유하기 때문에 관련을 맺을 수도 있다. (기록학용어사전(2008. 3. 10.) 한국기록학회)

2) 동일한 엔터티에 대한 레코드를 함께 결합하는 프로세스(ons.gov.uk, “Developing standard tools for data linkage”)

3) 법률 제17370호, 약칭 데이터기반행정법

4) 인공지능 국가전략(관계부처 합동, 2019)

- 통신, 금융, 유통·서비스 등 이종산업간 데이터 가공과 활용에 대한 관심이 확대되고 있으나 사용할 수 있는 데이터는 부족한 상황
- **(복합적 원인을 가진 사회현안 증가)** 기후, 인구, 산업화 등 복합적인 원인으로 발생하는 사회현안이 증가하고 있어 이에 대한 다차원적 분석을 위한 데이터가 요구됨
  - 대형화, 복합화되고 있는 재해·재난의 해결에도 융·복합 연구를 통한 문제 해결이 필요한 상황으로 각 전문분야의 데이터나 정보가 타 분야와 상호연동되어야 할 필요성이 증가하고 있음
  - 전염병 상황 등에서 감염병 데이터, 진단 데이터, 병상 및 의료인력 현황 등 여러 정보를 활용하여 대응하여야 하는 사례를 볼 수 있었음
  - 문제 해결에 유용한 정보 및 데이터가 관련 유관기관에서 다양하게 생산되고 있으나 연관 데이터의 획득이나 활용이 어려운 문제가 해결되어야 함

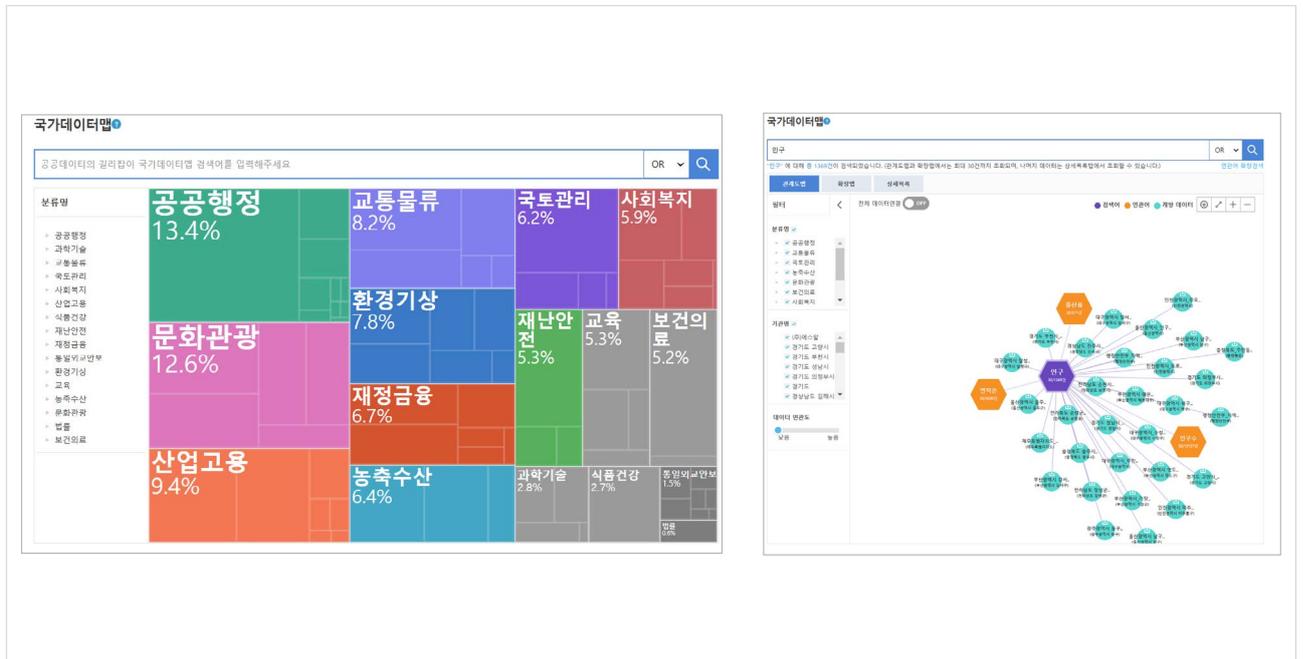
## ▶ 데이터셋 연계사례

- **(민간 공공간 데이터 연계)** 민간기업과 공공기관의 데이터 연계사례는 다음과 같음
  - 대전광역시의 소상공인 경제활동 분석데이터는 공공기관이 보유한 소상공인 데이터와 카드사의 카드매출액 데이터, 통신사의 유동인구 데이터를 연계하여 구축
  - 인천광역시의 불법주정차 단속경로 최적화 모델 개발에 공공데이터인 불법주정차 데이터와 민간데이터인 유동인구를 연계하여 활용
- **(공공부문내 데이터 연계)** 서로 다른 공공기관 간 데이터 연계사례는 다음과 같음
  - 한국산업인력공단의 국가자격정보데이터와 한국고용정보원의 취업정보데이터를 연계하여 맞춤형 일자리 지원 사업에 활용
  - 경상북도의 경우 생활SOC 기초 데이터를 국토지리정보원 이동거리, 통계청의 서비스인구 데이터와 연계하여 SOC시설당 서비스 인구수 정보 데이터셋 생성

● **(공공데이터포털에서의 데이터셋 연계)** 공공데이터포털<sup>5)</sup>의 ‘국가데이터맵’은 데이터셋의 분류<sup>6)</sup>와 제공기관을 기준으로 데이터셋을 연계

- 우리나라는 「공공데이터의 제공 및 이용 활성화에 관한 법률<sup>7)</sup>」에 따라 국가가 보유하고 있는 데이터를 공공데이터 포털에서 공개
- 국가데이터맵에서 데이터셋 연계 기준으로 데이터셋의 분류와 제공기관은 가장 기본적인 메타데이터를 활용한 방식으로 데이터셋의 내용 등을 반영한 정밀한 연관데이터를 보여주는데 한계가 있음

<그림 2> 국가데이터맵



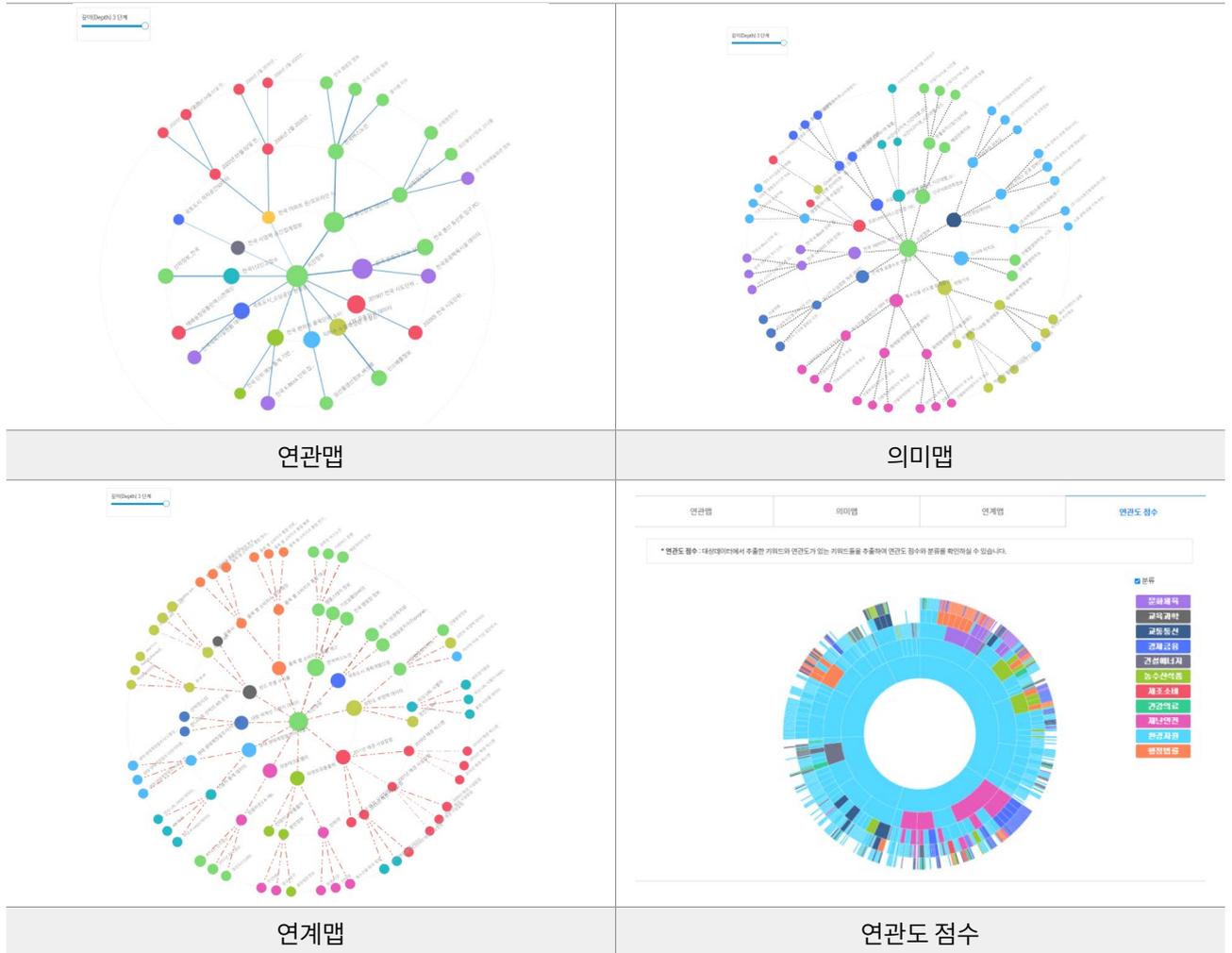
출처) 공공데이터포털, www.data.go.kr

● **(통합 데이터 지도<sup>8)</sup>에서의 데이터셋 연계)** 통합 데이터 지도에서는 데이터셋 그래프라는 명칭으로 연관 데이터셋을 가시화하여 제공한바 있음

- 데이터셋명, 설명, 키워드, 지역, 주제를 분석하여 연관맵, 의미맵, 연계맵, 연관도 점수 4가지 형태의 연관 데이터 그래프를 제공

5) www.data.go.kr, 한국지능정보사회진흥원  
 6) 공공행정, 과학기술, 교통물류 등 16개 분류 사용  
 7) 법률 제17344호, 약칭 공공데이터법  
 8) www.bigdata-map.kr, 한국지능정보사회진흥원

<그림 3> 통합데이터지도의 데이터셋 그래프 예(2021년 기준)



출처) 통합데이터지도, www.bigdata-map.kr

## KISTI의 데이터셋 연계전략

- (공공데이터 포털 데이터셋 분석을 통한 전략)** 2022년 2월 기준 공공데이터 포털에서 공개한 데이터셋 57,421종을 분석하여 「항목명 표준화를 통한 데이터셋 연계전략」 수립
  - 공공데이터 포털에서 공개한 데이터셋 57,421종의 항목명 2,099,994개를 수집 후 비표준화된 항목명 현황과 이의 표준화를 통한 데이터셋 연계 가능성을 분석하여 항목명 표준화 기반의 데이터셋 연계방안을 수립
  - 항목을 내용별로 분류한 결과 공간, 시간, 계측 관련 내용을 포함한 항목이 전체 항목의 67%에 이르러 이에 대한 표준화가 데이터셋 연계에 영향력이 높을 것으로 판단됨

<표 1> 공공데이터포털 데이터셋 항목명 분류

구분	항목분류	정의	분포율
1	공간정보	공간(면,점)의 의미를 가지는 데이터값을 가진 항목	35%
2	시간정보	날짜, 시간 데이터값을 가지는 항목	17%
3	계측정보	측정을 통 계측된 수치 데이터값을 가진 항목	15%
4	코드정보	덤코드, 발전소코드 등의 코드성 데이터값을 가진 항목	5%
5	전화번호	전화번호 값을 가지는 항목	4%
6	임의번호	관리를 위해 생산기관에서 임의 부여한 번호 값을 가지는 항목	4%
7	개인정보	인물명, 휴대전화번호 등과 같은 개인정보 관련 데이터값을 가진 항목	1%
8	기타	1% 이하의 분포를 가진 항목	19%

● **(지역현안 관련 데이터셋 연계 사례 분석을 통한 전략)** 지역현안 해결 방안의 연구를 위해 데이터셋을 연계해서 활용한 사례를 분석하여 「데이터값 표준화를 통한 데이터셋 연계전략」수립

- 도시 파급영향이 큰 재난 유형인 지진, 침수, 미세먼지를 자연현안으로 선정
- 지역현안 해결을 위해 데이터셋을 연계해 활용한 사례를 논문을 중심으로 조사하여 위치 값, 계측 값, 시간 값 등을 현안 해결을 위한 데이터셋 연계 기준으로 선정하는 것을 확인

<표 2> 지역현안 관련 데이터셋 연계 사례

현안구분	활용사례(논문명)	사용 데이터셋	연계항목
지진	지진재난에 대한 스마트 재난안전 관리체계 방안 연구	지진계측데이터, 재난데이터	시간, 공간
	IOT 지진감지를 이용한 재난안전시스템 개발	계측데이터, 센서데이터	시간, 공간
	어플리케이션 기반 건축물 재난 경보 시스템	센터데이터, 건축물데이터	시간, 공간
침수	기상이변·기후변화 대응 농경지 침수예측 서비스 체계 구축 연구 소개	지형데이터, 유역데이터, 기상데이터	시간, 공간
	GIS를 이용한 친수지구 침수범위예측 지원시스템 구축에 관한 연구	지형데이터, 기상데이터	시간, 공간
	클라우드 및 빅데이터 기반 침수 도로 탐색 시스템 개발	지형데이터, 기상데이터, SNS데이터	시간, 공간
미세먼지	미세먼지 자료동화 및 통합예보모형 개발연구	기상데이터, 대기질데이터	시간, 공간
	복합 신경망 구조를 이용한 미세먼지 위험 단계 예측 모델 설계 및 분석	기상데이터, 대기질데이터	시간, 공간
	공공데이터를 활용한 생활환경 정보서비스 제공 앱 개발	기상데이터, 대기질데이터	시간, 공간

- **(2단계 데이터셋 연계전략)** 항목명 단계 연계와 데이터값 단계 연계를 구분한 2단계 연계전략 수립

- 항목명 표준화를 통해 데이터셋을 연계하는 항목명 단계 연계와 데이터값 표준화를 통해 데이터셋을 연계하는 데이터값 단계 연계를 구분

<표 3> 2단계 데이터셋 연계전략

단계 구분	연계기준 및 전략
항목명 단계 연계	· (연계기준) Raw data의 항목명: 항목명의 일치성을 기준으로 연계 · 동일의미 항목명을 일관된 표현을 위해 항목명 표준화 필요
데이터값 단계 연계	· (연계기준) Raw data의 데이터값: 값의 일치성이나 유사성을 기준으로 연계 · 동일 값의 일관된 표현을 위해 공간 값, 시간 값, 코드, 측정 단위 등 표준화 필요

## 2. 항목명 단계 데이터셋 연계

### ▶ 데이터셋 연계방안

- **(표준항목명 및 유사항목명 목록 활용)** 동일의미 항목명 연계를 위하여 항목명 표준화를 위한 표준항목명, 유사항목명 목록 구축

- 표준항목명과 유사항목명의 목록을 구축하여 항목명을 표준화

<표 4> 구축한 표준항목명 및 유사항목명 예

유사항목명	▶	유사항목명
위도		위도좌표
		위도(Latitude)
유사항목명	▶	위도(deg.)
LATITUDE		LATITUDE
		위도
		위도값
		위도 좌표

- 표준항목명은 행정안전부의“공공데이터 공통표준용어”를 활용하되 누락된 항목명은 KISTI에서 자체적으로 구축
- 항목명이 “공공데이터 공통표준용어”에 출현한 용어인지 다른 단어가 추가된 복합항목명인지 분석하여 그에 따른 표준화 방법을 설계

단순 항목명	· ‘도로명주소’ 항목명은 공공데이터 공통표준용어에 등록된 항목명
복합 항목명	· ‘공장도로명주소’ 항목명은 공장과 도로명주소로 이루어진 복합 항목명

- 항목명 표준화는 그 의미를 잃지 않도록 표준화가 진행되어야 하며, 단일명사 표준화와 상호보완적인 관계로 발전해 갈 수 있도록 설계
- 표준항목명과 유사항목명 간의 관계는 1:N 관계를 가지며 유사항목명은 표준항목명을 추적할 수 있도록 저장 관리
- **(데이터셋 연계 절차 자동화)** 데이터셋의 항목명 수집, 항목명 분석, 연계가능성 분석, 데이터셋 연계로 연계 절차를 구분하고 각 단계를 자동화
- 항목명 자동 추출, 자동 추출된 항목명을 표준항목명 및 유사항목명과 매칭 처리, 매칭이 안된 항목명에 대한 처리, 데이터셋 연계 등 자동화

<그림 4> 데이터셋 연계 절차



## 연계절차 및 결과

- **(데이터셋 항목명 수집)** 항목명 단계 연계를 위하여 공공데이터 포털에서 공개한 데이터셋의 항목명을 수집
  - 데이터셋의 항목명 자동 수집기능을 구현하여 57,421종 데이터셋의 항목명 2,099,994개 수집

<표 5> 공공데이터포털 데이터셋 항목명 수집 현황

구분	데이터유형	데이터셋 수	수집 항목명 수
데이터셋 유형별 합계	파일데이터	48,143	1,793,828
	OPEN-API	9,179	303,789
	표준데이터	99	2,377
합 계		57,421	2,099,994

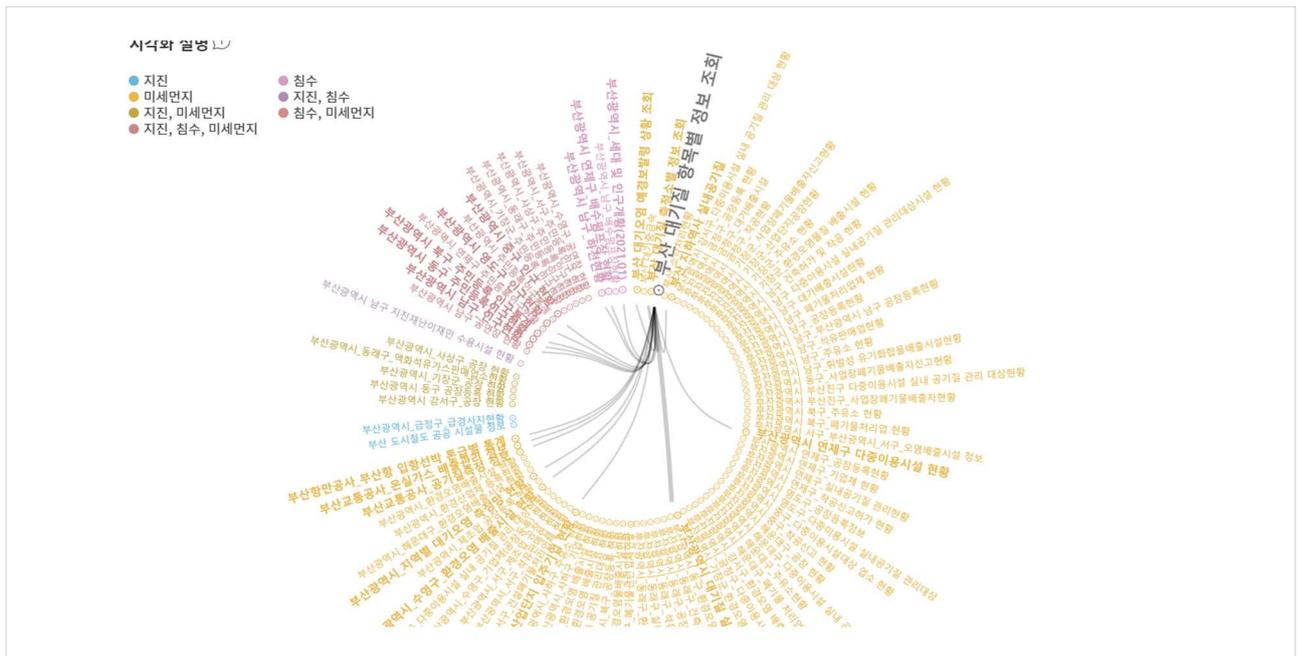
- **(항목명 전처리)** 전처리를 거쳐 동일의미 항목명 분석을 위한 정제 데이터 생성
  - TRIM, 더블이퀄레이션, 특수문자, 캐리지리턴, 더블 스페이스, 구두점 문자 등 제거하여 항목명을 표준화하기 전단계에서 정제 시행
  - 정제된 항목명은 여러 단어를 조합한 용어수준의 항목명을 포함하므로 이를 단어사전을 참조하여 단어 단위로 구분
- **(표준항목명 및 유사항목명 목록 구축)** 2,929건의 표준항목명과 10,043건의 유사항목명 목록 구축
  - 공공데이터 표준용어 및 행정용어표준에 출현한 단어를 표준항목명으로 저장
  - 수집 및 전처리된 항목명을 분석하여 표준항목명과 유사항목명 목록 작성
  - 표준항목명과 유사항목명은 1:N 관계를 가지며 이들 관계는 항목명 관계 테이블에 저장

<표 6> 유형별 표준항목명 및 유사항목명 구축 현황

유형	표준항목명	유사항목명
공간	163	1,323
계측	214	1,115
시간	85	586
개체	2,467	7,019
합계	2,929	10,043

- **(데이터셋 연계)** 데이터셋-항목명의 관계를 이용하여 같은 표준항목명을 갖는 데이터셋을 연계
  - 2장에서 예를 들어 설명한 ‘위도’ 항목명을 표준화할 경우 ‘위도’, ‘LATITUDE’, ‘위도값’, ‘위도 좌표’ 항목을 가진 데이터셋이 연계됨
  - 데이터셋 연계 정보는 데이터셋-항목명, 항목명-표준항목명 관계를 통해 저장 관리
- **(항목명 기반 연계결과 가시화)** 항목명을 기준으로 데이터셋을 연계하고 연계된 항목수에 따라 연결선 굵기를 다르게 가시화

<그림 5> 항목명 단계 연계 결과



### 3. 데이터값 표준화 기반 데이터셋 연계

#### ▶ 데이터셋 연계방안

- **(공공데이터셋 시간 값 표현형식 현황 분석)** 융합연구사업에서 수집한 데이터셋 402종을 분석한 결과 56가지 서로 다른 표현형식을 사용
  - 표현형식이 상이할 경우 값 비교를 통한 데이터셋 연계가 불가능하기 때문에 기존 형식으로 데이터값 표준화 필요

<표 7> 시간 값 표현형식 예

데이터 표현형식	데이터 샘플
YYYY-MM-DD 오후 HH12:MI:SS	2019-11-24 오후 6:07:00
YYYY-MM-DD HH24:MI	2020-07-23 05:00
YYYYMMDDHH24	2020072115
YYYY-MM-DD:HH24	2020-07-06:08
YYYY.MM.DD.HH	2006.09.12.00
YYYYMMDDTHH24	20200721T15

- **(공공데이터셋의 공간 값 보유현황 분석)** 지진, 침수, 미세먼지 주제에 대해 수집한 데이터셋 402종에 대한 값을 분석한 결과 71%의 데이터셋이 공간 값을 보유
  - 공공데이터를 이용해 공간 데이터 유사도의 판별이 가능한지 확인하기 위해 융합연구사업에서 수집한 데이터셋 402종을 분석한 결과 285개 데이터셋이 공간 값을 보유
  - 공간 값 중 과반 이상이 시/군/구까지의 값을 보유

<표 8> 공간 정보를 보유하고 있는 데이터셋의 데이터값 유형별 분포

NO	구축유형 No.	공간 값 수준	데이터셋(개)	백분율(%)
1	001	시/도 + 시/군/구 + 읍/면/동	111	38.9
2	002	시/도 + 시/군/구	61	21.4
3	003	시/도 또는 구축유형001 또는 구축유형002 복합	113	39.6
합계			285	100

- **(시간 및 공간값 유사도 산정방식 정의)** KISTI에서 정의한 값으로 데이터셋 간 동일 공간이나 시간의 정보 보유 비율로 정의하여 함께 분석할 수 있는 시간이나 공간 값을 가지고 있는 데이터셋을 판별하기 위한 척도로 활용
  - 데이터셋들이 보유하고 있는 시간 및 공간 일치도 비율을 유사도로 정의
  - 유사도 계산 대상인 두 개 데이터셋의 전체 시간 및 공간 값을 비교해 같은 값을 가지고 있는 레코드 수를 도출하고 전체 두 개 데이터셋 전체 레코드에서 차지하는 비율을 계산
  - 유사도 산정방식을 KISTI 독자적으로 정의하여 수치화된 유사도 값 계산
- **(공간 값 표준화 전략 수립)** 도로명주소를 기반으로 좌표(UTM-K), 위/경도, 사서함주소, 행정구역, 지번주소 간의 상호 매핑을 통한 표준화 방식 정의
  - 각 데이터셋은 공간 값은 좌표, 위경도, 주소 등 다양한 형태로 관리하고 있어 이를 상호간에 매핑하여 다른 형태의 공간 값으로 변환하되 표준화된 값은 상호 비교가 가능하도록 표준화 방식 수립
  - 공간 값 표준화는 값의 비교와 계산이 가능한 형식을 기준으로 하며 이를 기반으로 실제 데이터셋 간 공간 유사도 도출

## 연계절차 및 결과

- **(시간형식 표준화)** 연도, 날짜 단위의 비교 및 분석이 가능하도록 시간형식 표준화
  - 데이터셋별로 각기 다른 형식으로 표현하고 있는 시간 값을 ISO 19108:2002와 ISO 8601 정의 형식으로 변환
- **(공간 값 표준화)** 공간 값의 비교 및 분석이 가능하도록 공간 값을 표준화
  - 도로명주소를 기반으로 좌표(UTM-K), 위/경도, 사서함주소, 행정구역, 지번주소 등 다양한 형태로 관리되고 있는 공간 값을 일정한 형태로 변환
  - 공간 값 매핑 테이블을 활용하되 공간 값의 상호 비교가 가능한 형태를 선정하여 표준화
- **(유사도 분석)** 데이터셋간 유사도를 시간정보 및 공간정보 일치도를 계산하여 사전 분석 및 저장
  - 표준화된 시간 값과 공간 값을 활용하여 두 데이터셋간 유사도를 각각 계산
  - 데이터셋간 유사도는 값 전체에 대해 분석이 필요한 계산이므로 실시간 계산을 지양하고 주기적으로 계산한 결과를 저장하여 활용

- **(데이터셋 연계)** 데이터셋간 유사도 분포를 분석하여 일정 수준 이상의 유사도를 갖는 데이터셋들을 연계 데이터로 정의

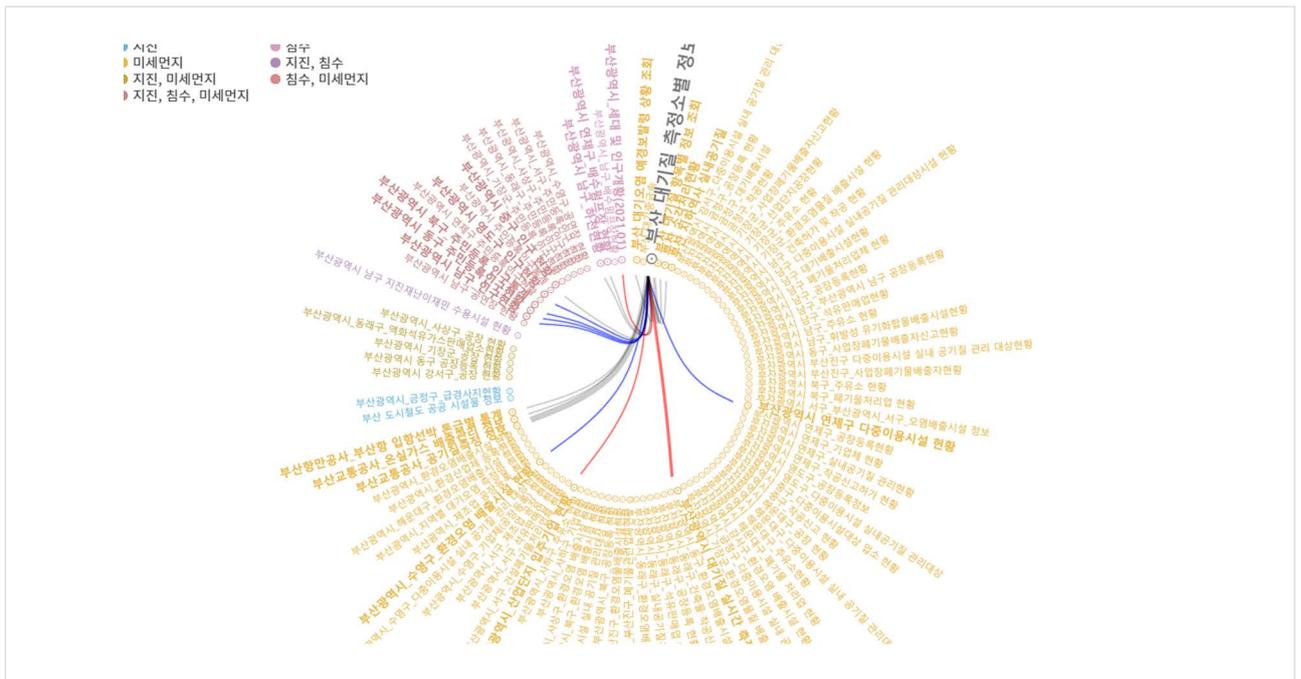
- 유사도 분포를 분석하여 유사도가 일정 수준인 데이터셋을 시간 또는 공간 값 기반 연계 데이터로 정의

- **(유사도 분석결과 가시화)** 데이터셋 연결선을 활용하여 유사도 표현

- 데이터셋 간 공간 유사도 수준을 낮음/중간/높음으로 나누고, 그에 대해 회색/파랑/빨강으로 표시

- 실제 데이터셋에 적용한 결과 “부산 대기질 측정소별 정보”라는 데이터셋과 공간 유사도가 높은 데이터셋으로 “부산광역시 대기질 실시간 측정”이라는 데이터셋이 확인되어 ‘대기질측정소’라는 공간정보 유사도가 높은 데이터셋이 연계되었음을 보여줌

<그림 6> 데이터값 단계 연계 결과



※ 적용한 가시화 모델은 연결선으로 표현할 수 있는 특성이 2개 이하인 한계가 있어 동일항목 수, 공간 유사도 2개 특성만 표현

## 4. 데이터셋 연계를 위한 제언

- **(데이터셋 연계 확대 노력 필요)** 데이터셋 연계에 대해 필요성이 증가되고 있어 데이터 연계 확장을 위한 적극적인 노력이 요구됨
  - 데이터 연계는 여러 출처의 정보를 결합하여 데이터의 정확성과 완전성을 향상
  - 사회·경제적 부가가치 창출과 기술 경쟁력 확보를 위해 데이터 융합 및 연계 강조
  - 기후, 인구, 산업화 등이 복합적인 원인으로 발생하는 사회현안이 증가하고 있어 이에 대한 다차원적 분석을 위한 데이터 연계 필요
- **(데이터셋의 동시 활용 확대를 위한 연계전략 필요)** 데이터셋의 동시 활용을 고려한 데이터셋 연계전략 필요
  - 기존 데이터셋 연계에서 자주 사용되었던 데이터셋의 발행기관, 분야 등 메타데이터를 이용한 방식은 동시에 활용할 데이터셋을 발견하기 위한 데이터셋의 세부 내용을 설명하는데 한계가 있음
  - 데이터셋의 활용 목적은 해당 데이터셋을 구성하는 항목이나 실제 값을 통해 이해가 가능하며 이를 분석하여 연계전략을 수립하는 것이 중요
  - 데이터셋 동시 활용사례 분석을 통해 산업계, 연구계 등 다양한 목적과 분야의 데이터셋 연계 요구 이해 필요
- **(데이터 표준화의 중요성 이해 필요)** 연계 실행 가능성을 높이기 위해 항목명과 값의 표준화 필요
  - ‘위도’, ‘LATITUDE’, ‘위도값’, ‘위도 좌표’ 등의 항목명을 표준화하여 대표항목명인 ‘위도’로 관리할 경우 연계율이 증가하며 또한 ‘위/경도’, ‘주소’, ‘좌표(UTM-K)’를 하나의 표현형식인 ‘위/경도’ 매핑하여 관리할 경우에도 연계율이 증가
  - KISTI의 데이터셋 연계전략은 항목명 단계 전략과 값 단계 전략 두 개로 구분되며 각 전략은 항목명 표준화와 값 표준화를 기반으로 함
  - 항목명 표준화는 표준항목명, 유사항목명의 관리가 요구되며 값 표준화는 값 표현형식, 단위 등의 기준관리 필요

## 참고문헌

- 관계부처 합동(2019), 「인공지능 국가전략」, 2019.12.
- 관계부처 합동(2022), 「2022년 공공데이터 제공 및 이용 활성화 시행계획」, 2022.2.
- 기록학용어사전(2008. 3. 10.) 한국기록학회
- 김원석(2019) 공공데이터를 활용한 생활환경 정보서비스 제공 앱 개발, 한남대학교 대학원
- 김태환(2020) 어플리케이션 기반 건축물 재난 경보 시스템 설계, 경북대학교 대학원
- 박윤수(2020) IOT 지진감지를 이용한 재난안전시스템 개발, 한국산업기술대학교 지식기반기술·에너지대학원
- 전상민, 이경도, 김대하, 황성환, 신용순, 강문성(2017) 기상이변 기후변화 대응 농경지 침수예측 서비스 체계 구축 연구 소개, Rural Resources, pp.53~59
- 심규성(2015) GIS를 이용한 친수지구 침수범위예측 지원시스템 구축에 관한 연구, 아주대학교 대학원
- 송영미(2015) 클라우드 및 빅데이터 기반 침수 도로 탐색 시스템 개발, 부경대학교 대학원
- 윤후영(2017) 미세먼지 자료동화 및 통합예보모형 개발 연구, 안양대학교 대학원
- 이경태(2021) 지진재난에 대한 스마트 재난안전 관리체계 방안 연구, 한양대학교 공학대학원
- 이기혁(2020) 복합 신경망 구조를 이용한 미세먼지 위험 단계 예측 모델 설계 및 분석, 한양대학교 대학원
- <https://www.bigdata-map.kr/>
- <https://www.data.go.kr>

---

**저 자**      **신수미**  
KISTI 국가과학기술데이터본부  
융합서비스센터  
책임연구원  
T. 02-3299-6075  
E. sumi@kisti.re.kr

# KISTI      제54호 ISSUE BRIEF

**발 행 일**    2023. 05. 08.

**발 행 인**    김재수

**편 집 위 원**    조민수, 최희석, 최장원, 정한민, 함재균,  
이준영, 이상환, 곽영

**발 행 처**    34141 대전광역시 유성구 대학로 245  
한국과학기술정보연구원 정책연구실  
<https://www.kisti.re.kr>

**I S S N**    2635-5728