

국가 데이터맵 프레임워크의 발전 방향

김학래 | 중앙대학교

배경

데이터 비즈니스의 현실

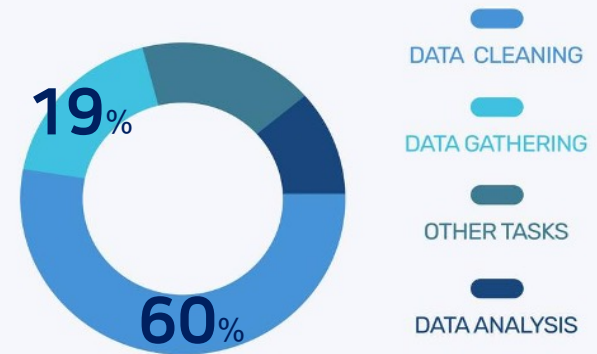
80% 데이터 랭글링 (Wrangling)에 시간 투자

- 조직 데이터를 전반적으로 이해할 수 있는 체계 부족
- 데이터의 탐색과 발견은 데이터 과학자의 개인 역량에 의존
- 전처리 중심의 데이터 분석으로 인한 피로감 증가

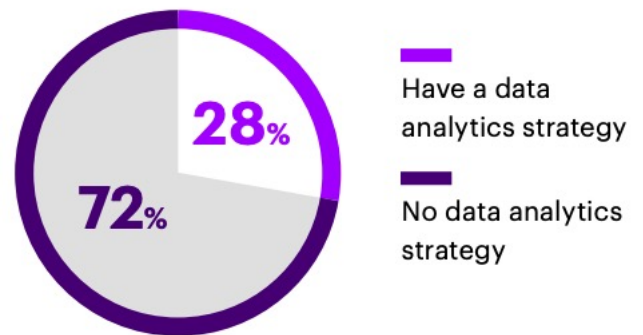
72% 데이터에 대한 전략 부재 (Accenture, 2020)

- 단편적인 데이터 분석으로 데이터 기반 전략적 가치 발굴 미흡
- 데이터 기반 의사결정을 위한 거버넌스 체계 미흡
- 이질적인 데이터의 관리, 처리, 분석을 위한 플랫폼 필요

Data Prep Still Dominates Data Scientists' Time



Do you have a Data & Analytics Strategy?



배경

데이터 거버넌스 갭 (Gap)

데이터로 얻을 수 있는 효과와 실제 얻고 있는 효과의 차이

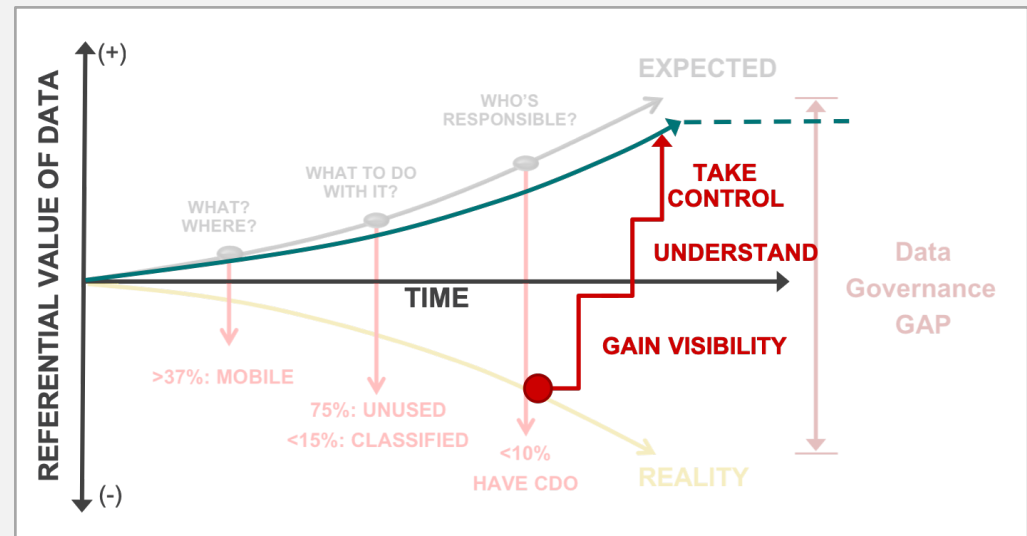
- 어떤 데이터를 어디에 축적하고 있는가?
- 데이터로 어떤 일을 하고 있는가?
- 데이터의 관리 책임자가 있나?

선진 조직 vs 일반 조직

- 데이터 전략 수립: 선진 67% vs 일반 27%
- 데이터 거버넌스 적용: 선진 63% vs 일반 32%
- 데이터 분석 방법론: 선진 63% vs 일반 24%
- 데이터 분석가의 역할: 선진 60% vs 일반 22%

데이터 카탈로그를 통해 거버넌스 갭 해결

- Gain Visibility, Understand, Get Control of the Data



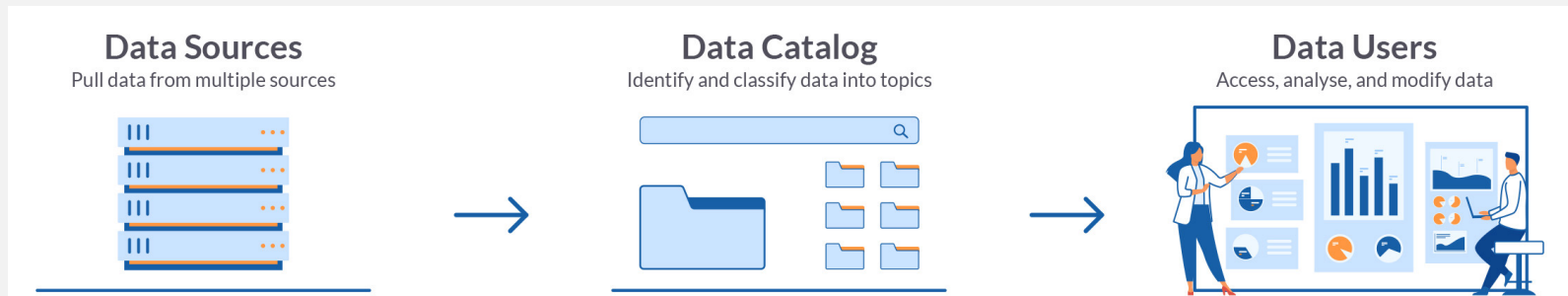
How to close the Data Governance Gap

개념

데이터 카탈로그란?

기업의 모든 데이터 자산을 카탈로그 방식으로 이해할 수 있도록 지원하는 도구 또는 시스템

- 이질적인 데이터에 대한 기술적 세부내용을 사용자들이 간단하고 이해하기 쉬운 방식으로 구축
- 사용자가 사용 가능한 데이터 세트를 식별하고, 이해할 수 있도록 지원
- 데이터 정의, 동의어, 핵심 비즈니스 용어 등 메타데이터에 대한 투명성 제공
- 데이터 소유, 계보, 품질에 대한 통합된 정보 제공
- 검색과 탐색, 다양한 소스에서 메타데이터 수집, 메타데이터 큐레이션, 데이터 분석과 시각화 제공



개념

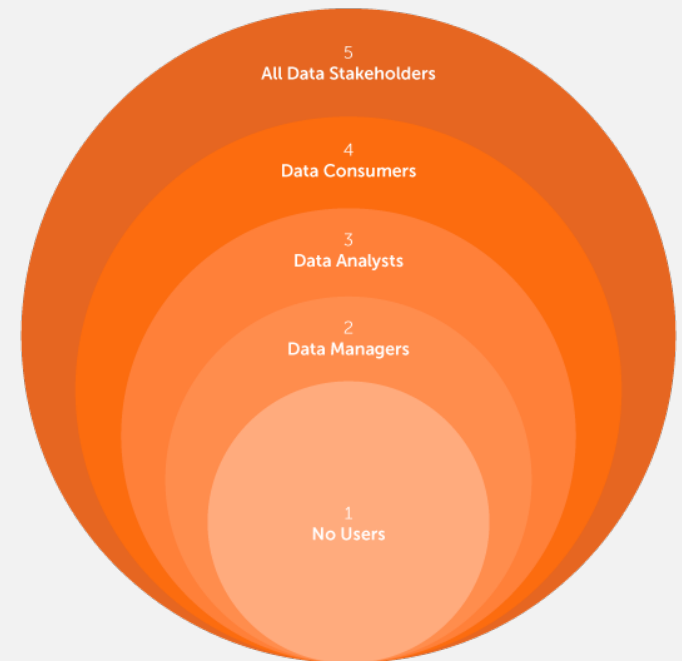
데이터 카탈로그의 주요 기능

조직의 데이터를 통합적으로 관리하기 위한 중앙 집중식 데이터 관리

- 데이터 사이언스, 엔지니어링에서 활용할 수 있는 유연한 검색과 탐색
- 다양한 소스에서 메타데이터 수집을 위한 데이터 파이프라인
- 메타데이터에 대한 기계학습 기반 자동화
- 다양한 방식의 데이터 제공 인터페이스

데이터 카탈로그의 장점

- 올바른 데이터를 검색하는 노력을 줄일 수 있는 통합 뷰 제공
- 기술·비즈니스 메타데이터 강화를 통해 데이터 기반 의사결정 지원
- 데이터 관리 개선을 통해 운영 효율성과 생산성 개선
- 데이터에 대한 소유권 확보와 데이터에 대한 신뢰 향상



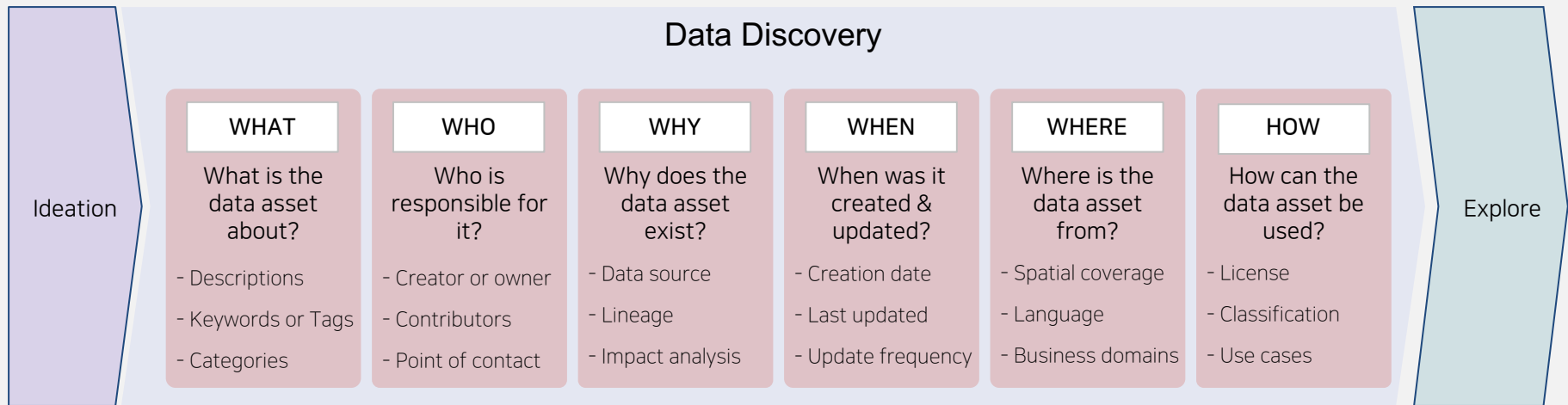
Enterprise Data Catalog Maturity Model

개념

메타데이터 모델과 데이터 탐색 - 5W1H Framework

구조화된 기술 메타데이터 (자동)와 비즈니스 메타데이터 (태그)에 대한 연계 검색

- 서로 다른 데이터 소스에서 수집된 테이블, 컬럼, 소유자, 관리자에 대한 검색
- 데이터의 변경 정보를 포함하는 메타데이터의 품질과 관련된 검색
- 특정 조건에 만족하는 데이터의 탐색과 질의응답 유형의 탐색



개념

메타데이터 유형

어떤 목적을 가고 만들어진 데이터 (constructed data with a purpose)

- 정보의 유형에 따라 메타데이터를 다양하게 구분
- 데이터 관리·탐색을 위해 서로 다른 유형의 메타데이터가 연계 필요 (기계학습/전문가 큐레이션)

Technical Metadata

Schema

Table

Column

Mapping

Profiling Metadata

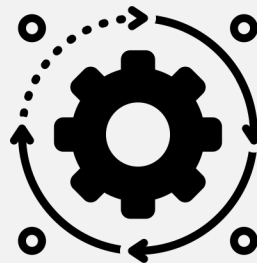
Density

Cardinality

Min, Max, Median

Ranking

Tagging /Linking



Automation
Expert required

Business Metadata

Business terms

Domain model

Ownership

Relationships

Legal Metadata

Sensitive data

GDPR

Security

Classification

시장동향

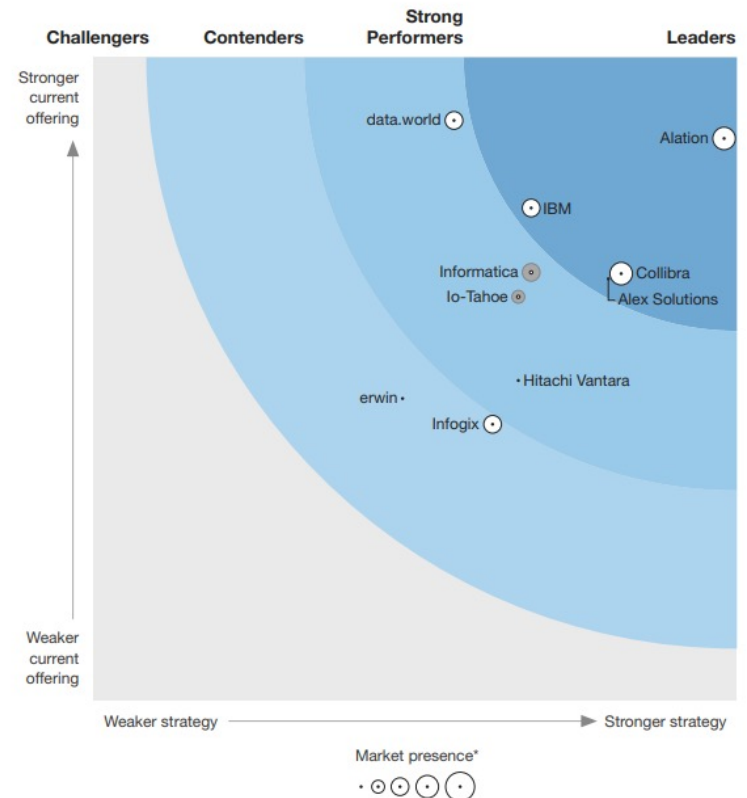
데이터 증가로 급속한 성장

25.7%의 연평균 성장률(CAGR) 전망 (Business Wire, 2021)

- 데이터의 지속적 증가로 글로벌 시장의 급속한 성장
- 미국 시장 770만달러까지 성장할 것으로 예측 (2023년)
- 제조, 헬스케어, 금융 등 아시아는 가장 높은 CAGR 예상
- 대부분의 벤더는 북미 중심으로 시장과 기술을 주도하고 있음

클라우드 기반 서비스를 기본으로 제공

- Cloud vs On-Premise 유형의 Deployment 제공
- 메타데이터 관리 중심 벤더: Alation, Collibra, IBM, data.world
- 구글, 아마존, MS 등 클라우드 벤더 Managed Services 제공
- Airbnb, LinkedIn, Lyft, Netflix, Uber 등 자체 솔루션 구축
- 오픈소스 프로젝트: Amundsen, Datahub, Apache Atlas



Machine Learning Data Catalogs

개요

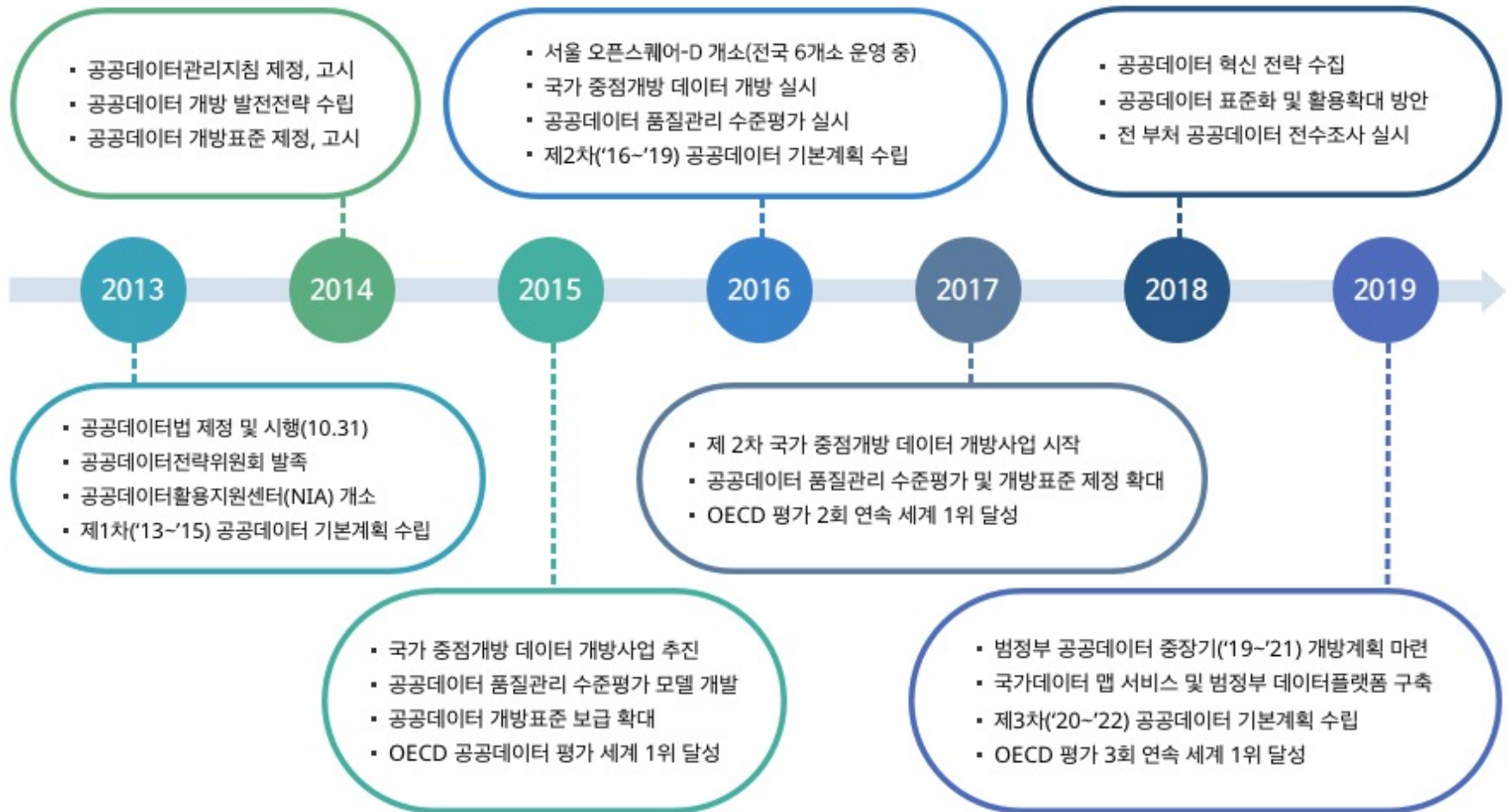
데이터댐 (18조 1,000억 원 투자 / 일자리 38만 9,000개 창출)

분야별 빅데이터 플랫폼을 30개로 확대하고 공공 데이터 14만 2,000개 신속 개방, AI 학습용 데이터 1,300종 구축 등을 추진한다. 5G망 조기구축을 위한 등록면허세 감면·투자세액 공제 등 세제지원을 추진하고 실감기술(VR·AR 등)을 적용한 교육·관광·문화 등 디지털콘텐츠 및 자율차 주행기술 등 5G 융합서비스를 개발한다. 스마트공장 1만 2,000개, 미세먼지 실내정화 등 AI 홈서비스 17종 보급, 생활밀접 분야 'AI+X 7대 프로젝트'도 추진한다. 7대 프로젝트에는 신종감염병 예후·예측, 의료영상 판독·진료, 범죄 예방·대응, 불법복제품 판독, 지역특화산업 품질관리 등이 포함돼 있다.



공공데이터

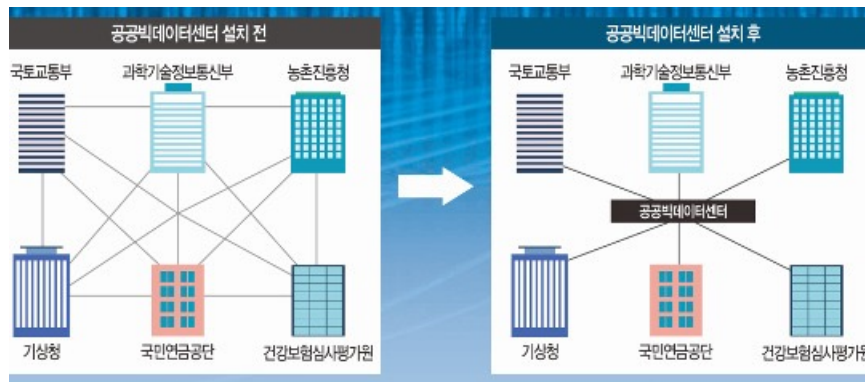
공공데이터 정책 추진경과



공공데이터

데이터기반 행정활성화에 관한 법률

공공기관간 데이터 연계·활용을 촉진하여 데이터 기반의 과학적 행정 시대를 이끄는 제도적 기반 마련하는 목표를 규정한 법률이다. 추진체계는 체계적인 데이터기반행정 수행을 위한 기본 계획 및 시행계획 수립과 데이터기반행정활성화위원회 설치 및 운영 (민간위원장, 관련공공기관 공무원, 민간전문가 등)을 포함하고, 데이터를 체계적으로 관리하기 위해 기관별 메타데이터를 종합하여 데이터관리체계와 공공기관간 데이터의 효율적 연계, 공동활용 및 분석을 위한 데이터통합관리플랫폼 구축을 구축한다.



붙임 데이터기반행정 활성화에 관한 법률안 체계도	
목적	- 데이터 기반 과학적 행정체계 구축하여 행정의 책임성 및 국민 삶의 질 제고 - 공공기관 간 데이터의 공유를 활성화하고 데이터 통합 관리체계 마련
추진 분야	- 주요 정책 수립 등을 위한 신속하고 정확한 국민 의견 수렴 - 비교·분석을 통한 최적화된 대책 마련 및 맞춤형 서비스 제공 - 안전·질병 등 사전에 위협 예측 및 제거방법 제시 - 경제·사회 등 분야에서 미래 수요 충족을 위한 선제적 대응책 마련 - 비용 절감이나 행정처리 절차 개선
추진 체계	데이터기반행정 활성화 위원회 - 주요정책, 기본계획 수립·변경 - 데이터 제공 거부에 대한 조정 - 국가적 차원 데이터 분석 사항 심의 정책기반 마련 - 기본계획(행안부, 3년) - 시행계획(기관별, 매년)
등록 및 제공	데이터 등록 및 제공 - 데이터 통합관리 플랫폼에 데이터 등록 - 기관 간 데이터 요청 및 제공 - 민간데이터 요청(협약) - 제공받은 데이터에 대한 관리
기반 구축	데이터 통합관리 - 기관별 메타데이터 등을 종합하여 '데이터관리체계' 구축 - 데이터의 효율적 제공·연계 및 공동 활용을 위해 '데이터통합관리 플랫폼' 구축 데이터 분석센터 - 기관별 분석센터 설치·운영 - 데이터기반 국가적 중요 의사결정 지원을 위한 정부통합데이터분석센터 설치
	조직 및 인력 - 기관별 데이터기반행정 책임관 지정 - 데이터기반행정 활성화 지원을 위하여 전문기관 지정·운영 평가 - 데이터 제공 및 활용 성과 등 점검·평가 - 데이터기반행정 성과 포상

공공데이터

국가데이터맵

국가 데이터맵은 데이터 소재지와 데이터 간 연관관계를 그래프 형태로 표현한 서비스이다. 국가 데이터맵은 범정부 데이터 플랫폼 구축사업의 결과이며, 700여개 공공기관이 보유한 공공데이터의 소재와 메타데이터 정보를 통합관리한다.

기관메타데이터

정 보 시 스 템	1 기관명	2 정보시스템명	3 관련법령(보유목적)	4 구축년도
	5 운영부서명	6 담당자명		
	9 논리DB명	10 물리DB명		
	13 DBMS정보	14 운영체제정보		
D B	17 데이터용량	18 수집제외사유		
	19 테이블소유자	20 테이블영문명		
	23 테이블설명	24 업무분류체계		
	27 보존기간	28 발생주기		
테 이 블	31 컬럼영문명	32 컬럼한글명		
	35 데이터길이	36 데이터포맷		
	39 FK정보	40 제약조건		
	43 공개/비공개여부			

자동 수집 항목

공공기관의 데이터베이스 표준화 지침, 기준 관리 항목

DATA . GO . KR

로그인 회원가입 사이트맵 ENGLISH

데이터셋 제공신청 활용사례 정보공유 이용안내

인기검색어 1 미세먼지 2 일기예보 3 기온변화 4 수질개선 5 물가동향

미세먼지

기관명 (12건)

- 기상청
- 국립환경과학원
- 서울도시철도공사
- 한국여객선공단
- 한국산업단지공단
- 한국환경공단

업무분류체계 (6건)

- 국토관리
- 교통물류
- 재난안전
- 사회복지
- 사회복지일반
- 환경기성

데이터명 (18건)

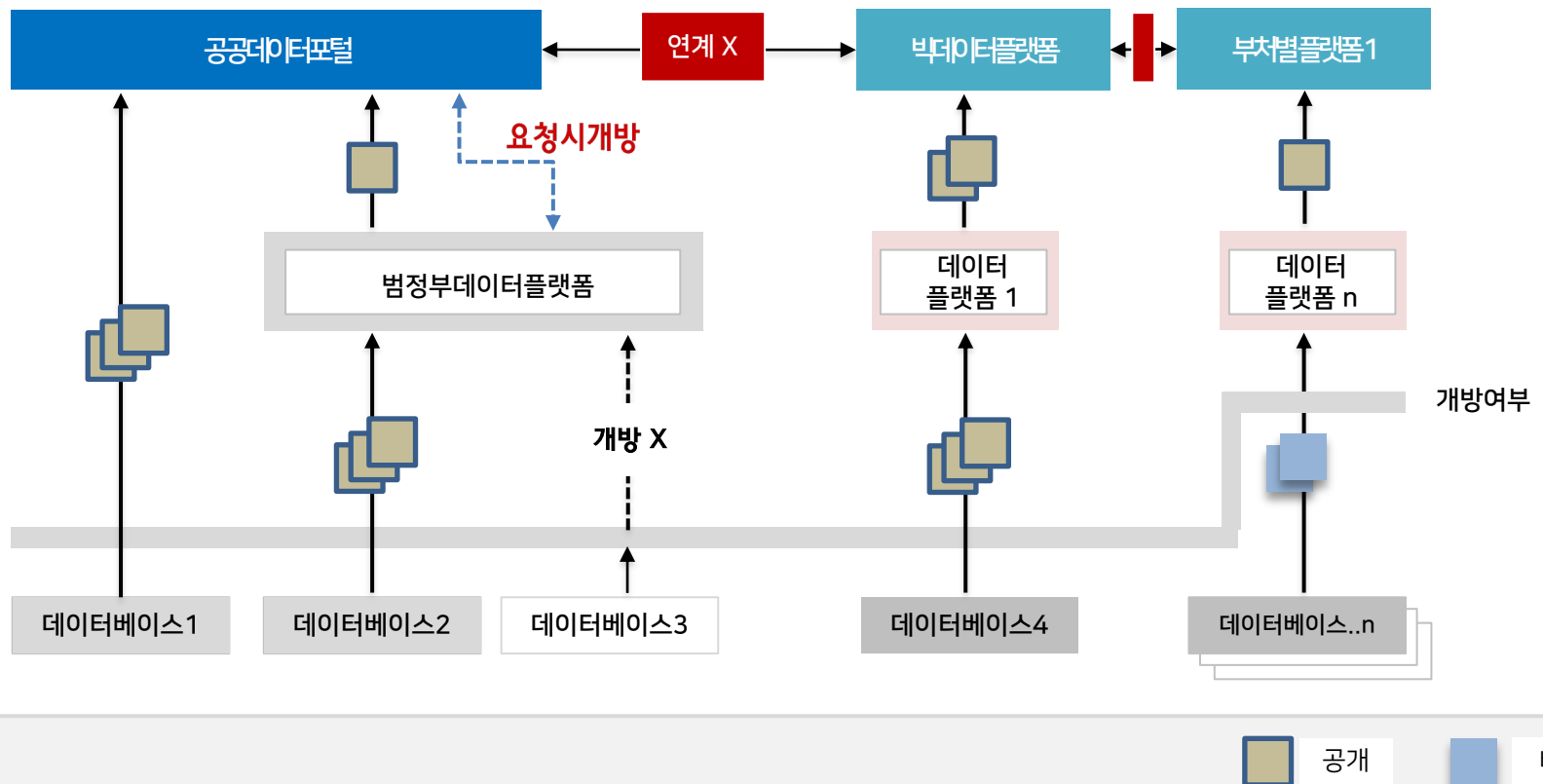
- 미세먼지 정보
- 미세먼지 관련 메타데이터
- 오존황사 발생정보
- 전국산업단지 일반현황
- 황사정보조회서비스
- 자동차 배출가스 및 소음인원 정보 조회

항목명 (7건)

- 속성값
- PM2.5 (mg/m)
- PM10 (mg/m)
- O3(ppm)
- CO(ppm)
- NO2(ppm)
- SO2(ppm)

Reality Check

범정부 데이터 연계와 융합



상호운용성 - 설계

DCAT 기반 데이터 표현

공개표준 데이터의 메타데이터 항목을 추출, 공통항목에 대한 의미 관계를 도출하여 DCAT 어휘를 적용한다. 데이터 모델에 사용한 어휘는 DCAT과 Schema.org에 한정하고 새로운 어휘를 생성하지 않는다 (현재 공개표준 데이터에서 확보할 수 있는 정보의 표현은 두 개의 어휘로 표현할 수 있다)

지식모델 표현을 위한 URI 설계

- <http://knowledge.cau.ac.kr/standard/>
- URI의 식별자: 개별 데이터의 파일 ID (fileID)
- 개별 데이터셋의 메타데이터에 적합한 DCAT 어휘 적용
- 분류체계, 키워드에 대한 모델링 제외 (향후 검토 필요)

지식 그래프 데이터 변환

- 개방표준 데이터의 지식그래프 생성
- 약 14,000개 규모의 트리플 (진술문)으로 표현
- 데이터 규모는 표현력에 따라 차이가 있음

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:dct="http://purl.org/dc/terms/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:dcats="http://www.w3.org/ns/dcat#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">

  <rdf:Description rdf:about="http://knowledge.cau.ac.kr/standard/15013111">
    <rdf:type rdf:resource="http://www.w3.org/ns/dcat#Dataset"/>
    <dcats:description>전국 휴양림 정보(임장료, 숙박정보 등)를 제공합니다. 공공데이터 개방 표준데이터 속성정보(표현형식/단위 등)는
    <dcats:identifier>15013111</dcats:identifier>
    <dct:publisher>자치행정기관</dct:publisher>
    <dcats:modified>2020-08-18</dcats:modified>
    <dcats:keyword>문화관광</dcats:keyword>
    <dct:title>전국휴양림표준데이터</dct:title>
    <dcats:downloadURL rdf:resource="http://data.go.kr/data/15013111/standard.do"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://knowledge.cau.ac.kr/catalog/standard/120">
    <dcats:dataset rdf:resource="http://knowledge.cau.ac.kr/standard/15013111"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://knowledge.cau.ac.kr/standard/15028194">
    <rdf:type rdf:resource="http://www.w3.org/ns/dcat#Dataset"/>
    <dcats:description>전국 도로안전표지 정보(도로안전표지구분, 도로안전표지종별일련번호, 도로안전표지설명 등)를 제공합니다. 공공데이터
    <dcats:identifier>15028194</dcats:identifier>
    <dct:publisher>자치행정기관</dct:publisher>
    <dcats:modified>2020-07-15</dcats:modified>
    <dcats:keyword>교통물류</dcats:keyword>
    <dct:title>전국도로안전표지표준데이터</dct:title>
    <dcats:downloadURL rdf:resource="http://data.go.kr/data/15028194/standard.do"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://knowledge.cau.ac.kr/catalog/standard/120">
    <dcats:dataset rdf:resource="http://knowledge.cau.ac.kr/standard/15028194"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://knowledge.cau.ac.kr/standard/15028204">
    <dcats:identifier>15028204</dcats:identifier>
    <dct:publisher>자치행정기관</dct:publisher>
    <dcats:modified>2020-06-17</dcats:modified>
    <rdf:type rdf:resource="http://www.w3.org/ns/dcat#Dataset"/>
    <dcats:description>전국 자동차정비업체 정보(정비업체종류, 주소 등)를 제공합니다. 공공데이터 개방 표준데이터 속성정보(표현형식/단
    <dcats:keyword>교통물류</dcats:keyword>
    <dct:title>전국자동차정비업체표준데이터</dct:title>
    <dcats:downloadURL rdf:resource="http://data.go.kr/data/15028204/standard.do"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://knowledge.cau.ac.kr/catalog/standard/120">
    <dcats:dataset rdf:resource="http://knowledge.cau.ac.kr/standard/15028204"/>
  </rdf:Description>

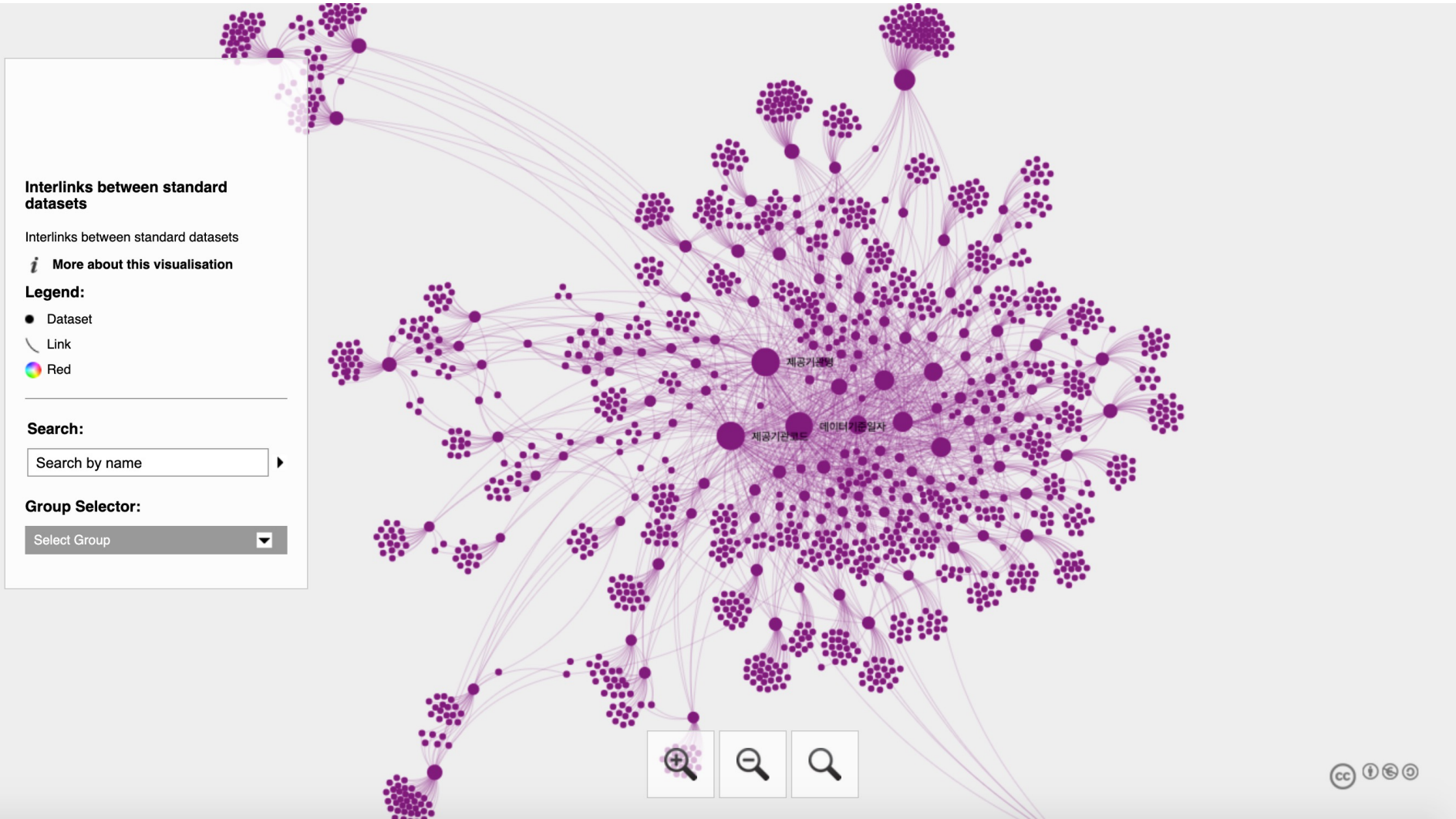
  <rdf:Description rdf:about="http://knowledge.cau.ac.kr/standard/15028208">
    <rdf:type rdf:resource="http://www.w3.org/ns/dcat#Dataset"/>
    <dcats:description>전국 푸드트럭허가구역 정보(허가구역사용료, 허가구역운영사각 등)를 제공합니다. 공공데이터 개방 표준데이터 속성
    <dcats:identifier>15028208</dcats:identifier>
    <dct:publisher>자치행정기관</dct:publisher>
    <dcats:modified>2020-05-20</dcats:modified>
    <dcats:keyword>식품건강</dcats:keyword>
    <dct:title>전국푸드트럭허가구역표준데이터</dct:title>
    <dcats:downloadURL rdf:resource="http://data.go.kr/data/15028208/standard.do"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://knowledge.cau.ac.kr/catalog/standard/120">
    <dcats:dataset rdf:resource="http://knowledge.cau.ac.kr/standard/15028208"/>
  </rdf:Description>
```

상호운용성 - 구현

DCAT 기반 데이터 관계의 시각화

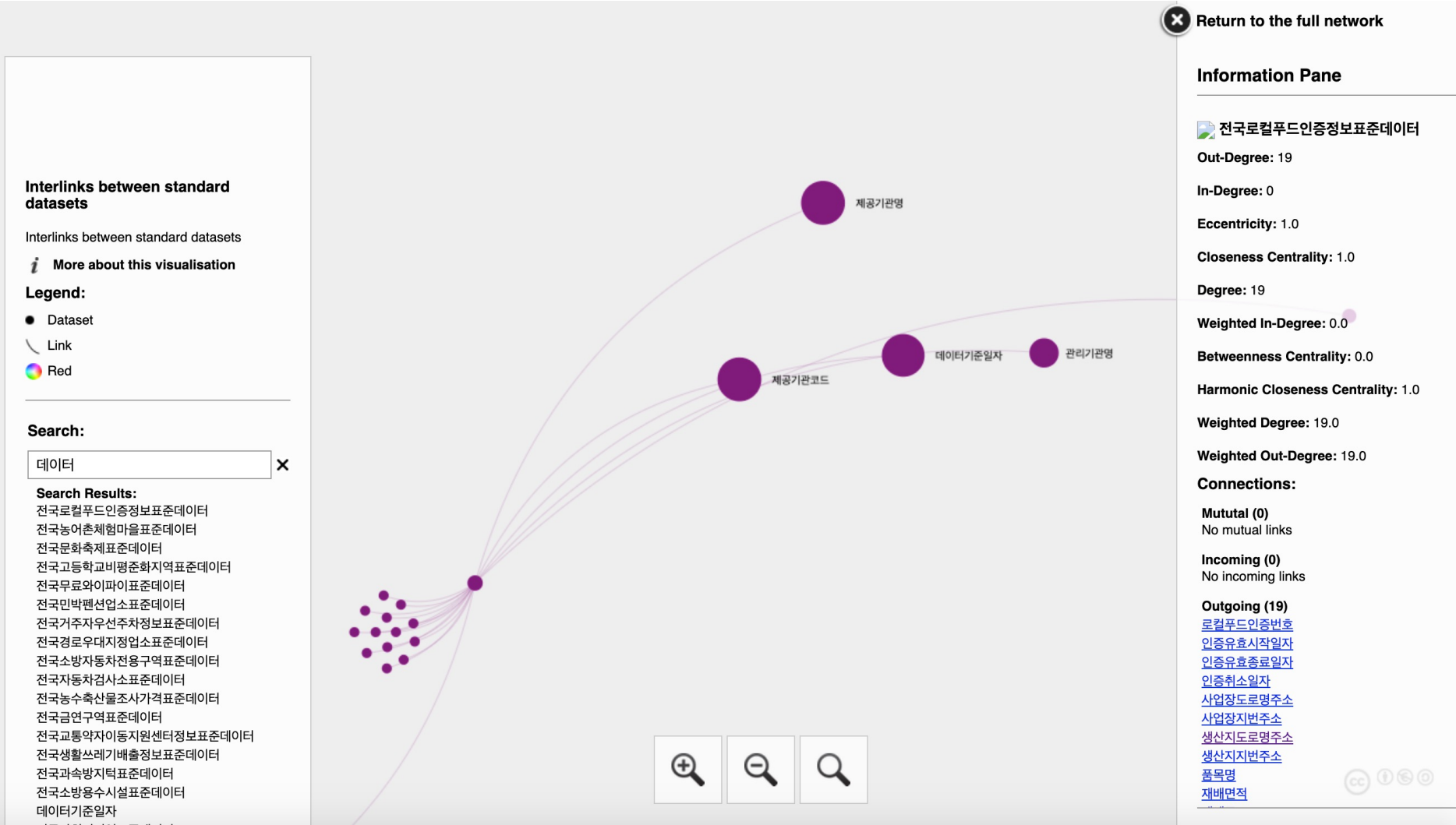
DCAT 데이터에 표현된 데이터의 의미적 관계성을 네트워크 분석을 통해 시각화한다. 네트워크는 [파일명-항목명(컬럼명) 사이의 관계 분석하고 있다. 그래프 데이터는 개별 데이터가 갖고 있는 속성을 기준으로 연결 가능성을 탐색하고, 의사결정하는데 유용하게 활용된다.



상호운용성 - 구현

DCAT 기반 데이터 관계의 시각화

특정한 속성을 갖고 있는 정보를 탐색한다. '데이터'라는 키워드로 검색하면, 해당 키워드가 있는 항목명과 데이터명을 정보를 추출한다.



Reality Check

데이터 서비스 수준의 상호운용성

운영주체, 플랫폼이 다른 환경에서 데이터 서비스의 전반적 현황과 데이터세트의 관계를 파악하는데 한계가 있음

(a) 서로 다른 3개 데이터 포털에서 제공하는 데이터세트의 개수는?

(b) '교통' 관련 데이터 현황은?

	Public data portal	Bigdata platform	Seoul Open data portal
Management	Ministry of the Interior and Safety	Ministry of Science and ICT	Seoul Metropolitan Government
Website	https://www.data.go.kr/	https://www.bigdata-map.kr/	https://data.seoul.go.kr/
Platform	Self-development	Self-development	Self-development
The number of files	37,483	9,638	6,607
The number of APIs	6,526	1,341	4,991
The number of classifications	18	10	12
The number of organisations	953	-	-

문제정의

데이터 서비스 수준의 상호운용성

(a) 서로 다른 3개 데이터 포털에서 제공하는 데이터세트의 개수는?

전제조건:

- (1) 모든 포털이 시스템 수준에서 연계되어 있다.
- (2) 모든 포털이 DCAT으로 데이터를 표현한다.
- (3) 모든 포털이 동일한 수준의 SPARQL 엔드포인트 서비스를 제공한다.

이슈:

- (1) 개별 데이터포털의 데이터 규모가 크거나
- (2) 서로 메타데이터의 규격이 다른 경우
- (3) 실시간 엔드포인트 서비스를 제공하지 않는 경우

문제정의

데이터 서비스 수준의 상호운용성

(b) '교통' 관련 데이터 현황은?

전제조건:

- (1) 모든 포털이 동일한 분류체계를 사용 한다.
- (2) 모든 포털이 동일한 분류체계에 대한 URI를 사용한다.
- (3) 모든 포털이 동일한 수준의 SPARQL 엔드포인트 서비스를 제공한다.

이슈:

- (1) 동일한 분류체계가 존재하지 않는 경우
- (2) 동일한 URI를 제공하지 않는 경우
- (3) 실시간 엔드포인트 서비스를 제공하지 않는 경우

해결방법

기계가 읽을 수 있는 어휘로 데이터 포털 수준의 정보 표현

데이터 서비스의 운영·관리 정보와 데이터세트에 대한 집합적 정보 표현

공공데이터 포털

```
<rdf:Description rdf:about="http://knowledge.cau.ac.kr/data/datamap/public-data-portal">
  <rdfs:label>Datamap of public data portal</rdfs:label>
  <rdf:type rdf:resource="http://knowledge.cau.ac.kr/ns/datamap/DataMap"/>
  <dm:numberOfOrg rdf:datatype="http://www.w3.org/2001/XMLSchema#int">45</dm:numberOfOrg>
  <dm:numberOfDataset rdf:datatype="http://www.w3.org/2001/XMLSchema#int">2345</dm:numberOfDataset>
  <dm:endpoint rdf:resource="http://query.data.go.kr/sparql/">
  <dm:numberOfCatalog rdf:datatype="http://www.w3.org/2001/XMLSchema#int">3</dm:numberOfCatalog>
</rdf:Description>

<rdf:Description rdf:about="http://knowledge.cau.ac.kr/data/datamap/public-data-portal ">
  < dct:subject rdf:resource="http://knowledge.cau.ac.kr/data/subject/transportation"/>
</rdf:Description>
```

빅데이터 플랫폼

```
<rdf:Description rdf:about="http://knowledge.cau.ac.kr/data/datamap/sample-bigdata-map-kr">
  <rdfs:label>Datamap of big data platform</rdfs:label>
  <rdf:type rdf:resource="http://knowledge.cau.ac.kr/ns/datamap/DataMap"/>
  <dm:numberOfOrg rdf:datatype="http://www.w3.org/2001/XMLSchema#int">17</dm:numberOfOrg>
  <dm:numberOfDataset rdf:datatype="http://www.w3.org/2001/XMLSchema#int">9638</dm:numberOfDataset>
  <dm:endpoint rdf:resource="http://query.bigdata-map.kr/sparql/">
  <dm:numberOfCatalog rdf:datatype="http://www.w3.org/2001/XMLSchema#int">1</dm:numberOfCatalog>
</rdf:Description>

<rdf:Description rdf:about="http://knowledge.cau.ac.kr/data/subject/education">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
</rdf:Description>

<rdf:Description rdf:about="http://knowledge.cau.ac.kr/data/datamap/sample-bigdata-map-kr">
  < dct:subject rdf:resource=" http://knowledge.cau.ac.kr/data/subject/transportation"/>
</rdf:Description>
```

해결방법

데이터 포털 수준의 의미적 데이터 표현과 접근

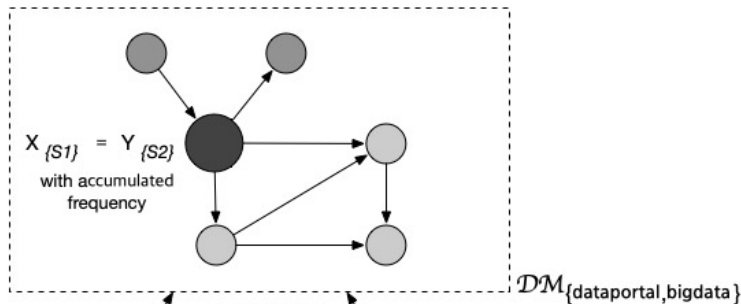
서로 다른 데이터 서비스의 정보를 데이터 맵으로 표현하고 통합할 수 있는 방법 제공

(a) 서로 다른 3개 데이터 포털에서 제공하는 데이터세트의 개수는?

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dm: <http://knowledge.cau.ac.kr/ns/datamap/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dct: <http://purl.org/dc/terms/>
SELECT (SUM(xsd:integer(?org)) as ?orgs)
(SUM(xsd:integer(?dataset)) as ?datasets) (SUM(?catalog)
as ?catalogs)
WHERE
{
    ?s dm:numberOfOrg ?org .
    ?s dm:numberOfDataset ?dataset.
    OPTIONAL {?s dm:numberOfCatalog ?catalog.}
}
```

(b) ‘교통’ 관련 데이터 현황은?

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dm: <http://knowledge.cau.ac.kr/ns/datamap/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dct: <http://purl.org/dc/terms/>
SELECT (SUM(xsd:integer(?dataset)) as ?datasets)
WHERE
{
    ?s dm:numberOfDataset ?dataset.
    ?s dct:subject
    <http://knowledge.cau.ac.kr/data/subject/transportation>.
    OPTIONAL {?s dm:availableType
    <http://knowledge.cau.ac.kr/data/format/json> .}
}
```



Accumulated
by SPARQL
Queries

$DM_{dataportal}$

$DM_{bigdata}$

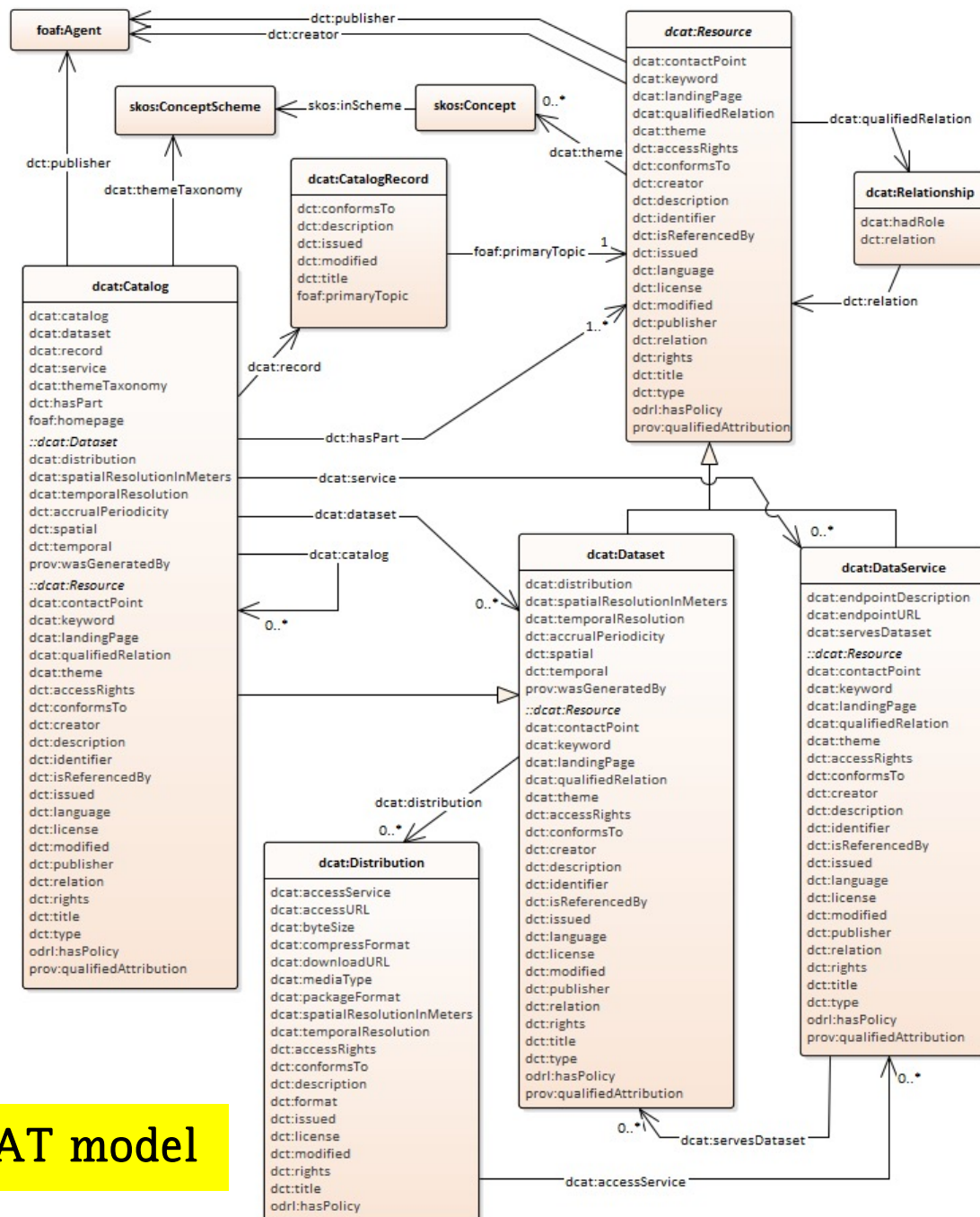
Data Catalog Vocabulary W3C Recommendations

부처·기관의 목적에 따라 파편화된 데이터 정책으로 데이터 연계가 어려움

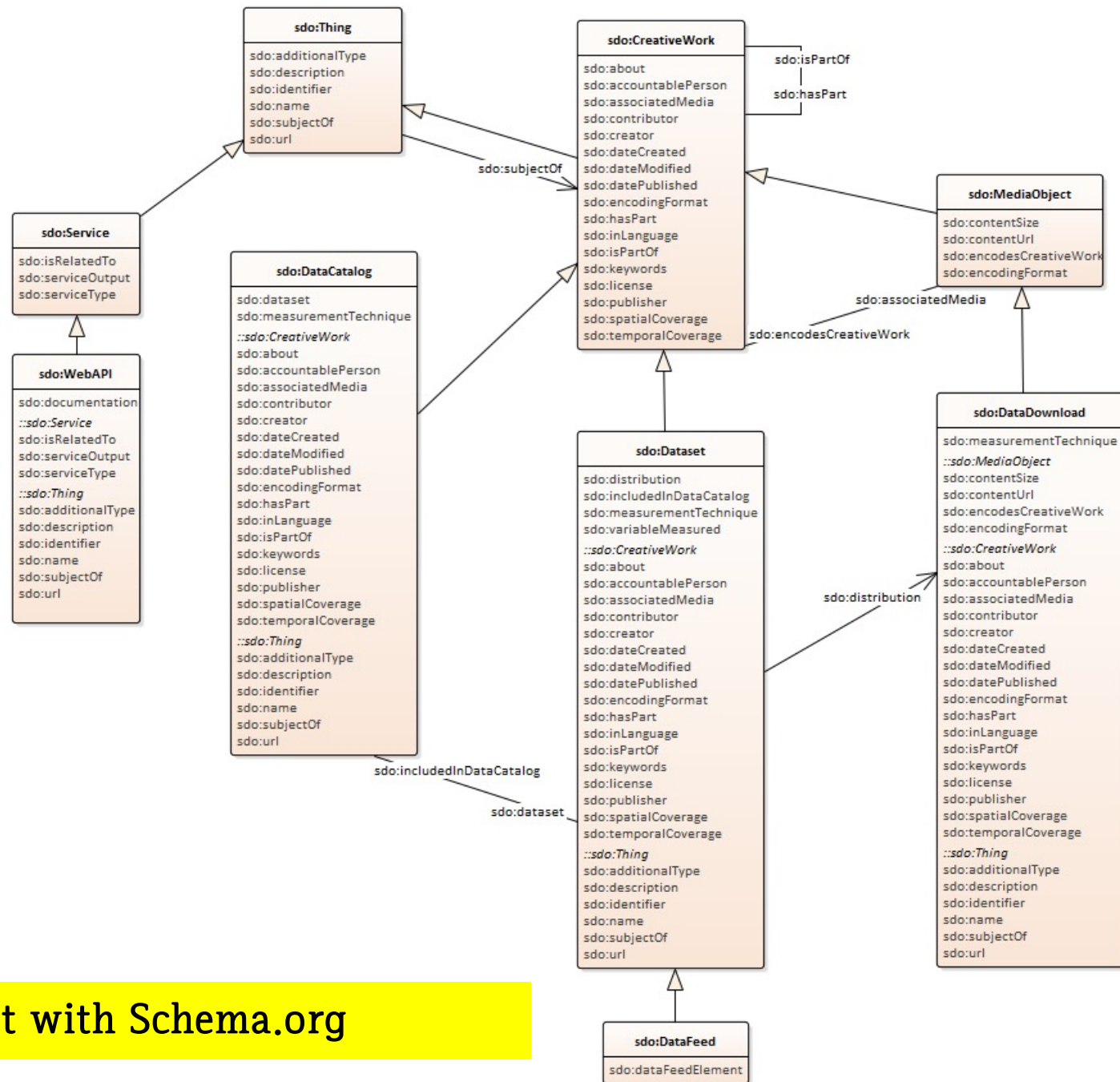
데이터 카탈로그 표현을 위한 어휘

- **dcat:Catalog**: 데이터셋 또는 데이터 서비스의 메타데이터를 표현하기 위한 클래스
- **dcat:Resource**: 카탈로그의 메타데이터를 기술하기 위한 클래스
 - dcat:Dataset, dcat:DataService, dcat:Catalog의 상위 클래스 (parent class)
- **dcat:Dataset**: 데이터셋 (데이터의 집합)을 표현하기 위한 클래스
- **dcat:Distribution**: 데이터셋의 접근 형식을 표현하기 위한 클래스
- **dcat:DataService**: 데이터를 제공하는 서비스를 표현하기 위한 클래스
- **dcat:CatalogRecord**: 카탈로그의 메타데이터 항목을 표현하기 위한 클래스

2020년 2월에 버전 2 발표(<https://www.w3.org/TR/vocab-dcat-2/>)



Overview of DCAAT model



관련 연구

DCAT-AP

부처·기관의 목적에 따라 파편화된 데이터 정책으로 데이터 연계가 어려움

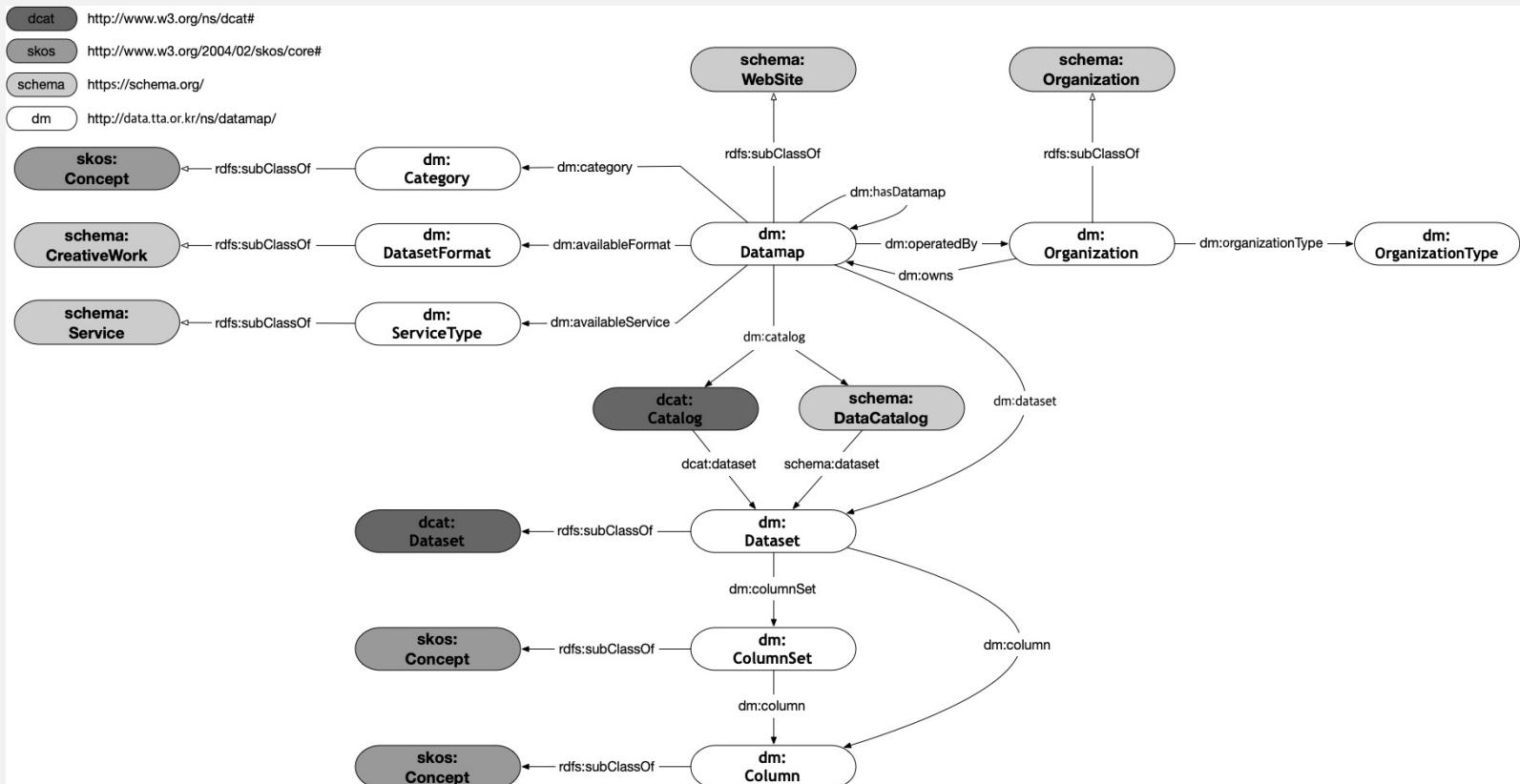
The DCAT Application profile for data portals in Europe

- 유럽의 공공부문 데이터셋을 DCAT 기반으로 기술하기 위한 명세
- **주요 특징**
 - DCAT보다 엄격: 필수/추천 속성 정의
 - 통제 어휘 적용: 카테고리, 언어, 파일 형식, Geonames
 - 확장 어휘: GeoDCAT-AP, StatDCAT-AP, DCAT-AP-JRC, DCAT-AP-IT
- **주요 추진 경과**
 - In February 2015, the ISA Programme of the European Commission launched an activity to revise DCAT-AP.
 - In September 2017, the DCAT-AP version 1.1 was released.
 - In November 2019, the major release took place: DCAT-AP v2.0 is now available.
- **DCAT-AP github:** <https://github.com/SEMICeu/DCAT-AP/tree/master/releases/2.0.0>

표준

데이터맵 어휘 (TTAK.KO-10.1291)

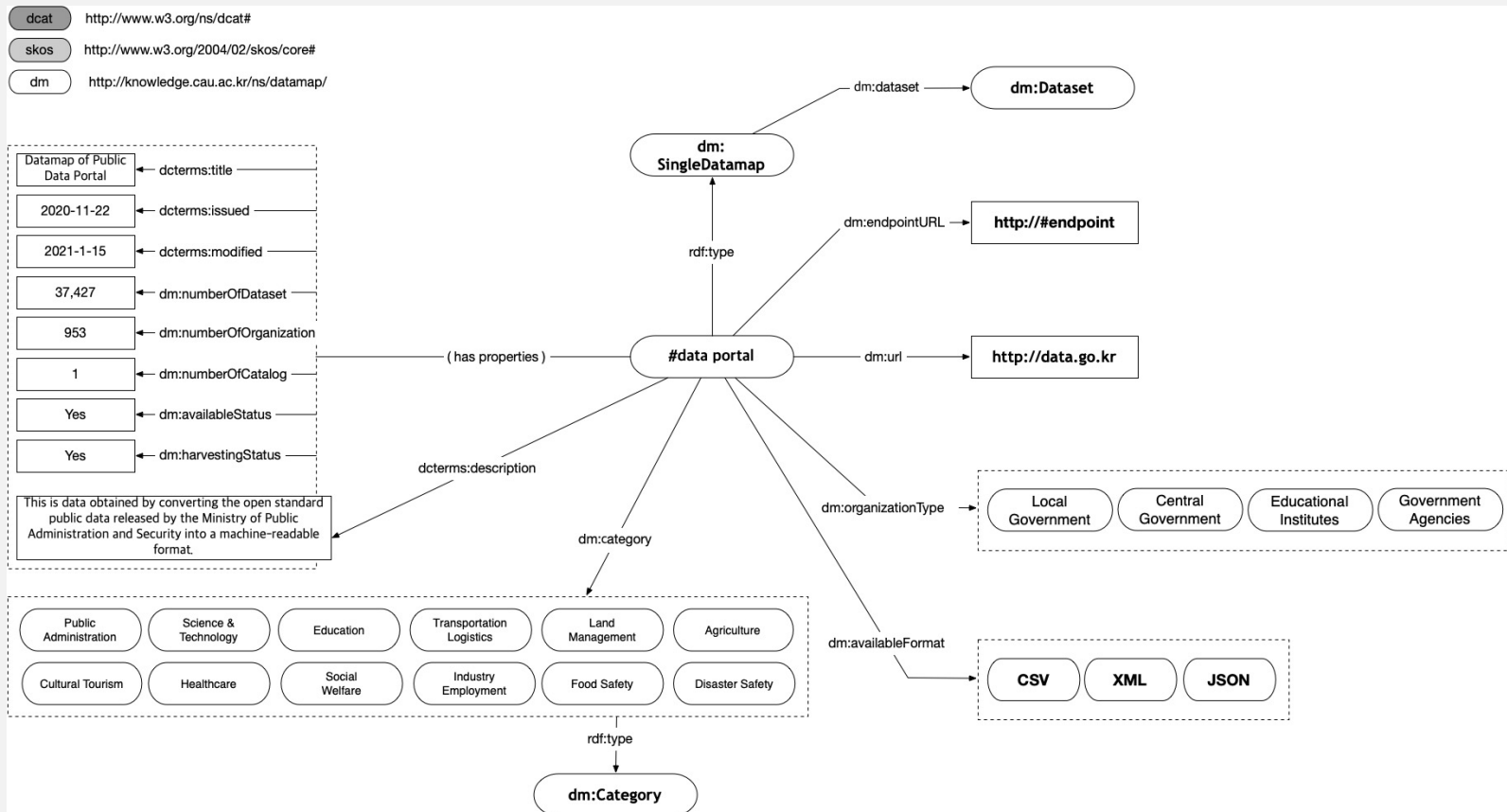
집합적인 데이터 카탈로그의 메타데이터 (DataMap), 데이터 서비스품질, 핵심어휘 (BRM), 데이터셋의 이력정보, 이종의 데이터 서비스의 연계를 표현하기 위한 어휘



표준

데이터맵 어휘 - 공공데이터포털

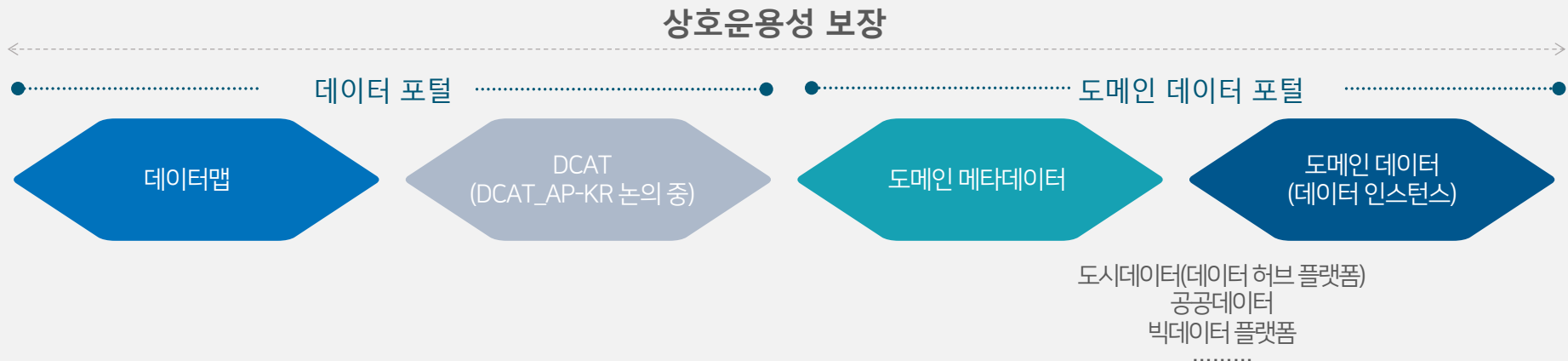
데이터맵 어휘를 적용해서 공공데이터포털의 메타데이터를 표현



표준의 확장

데이터맵 프레임워크 표준 개발 - TTA 단체표준 개발중

서로 다른 도메인 사이의 데이터 상호운용성을 확보하기 위해 다양한 차원의 연계 방안 표준화 필요



- 데이터맵 프레임워크 표준 개발(TTA 단체표준 개발중)
 - 대응 표준작업반: TTA TC010 PG 1004
 - 과제 번호: 2021-P1840
 - 공동제안기관: TTA, 중앙대, 군산대, KISTI

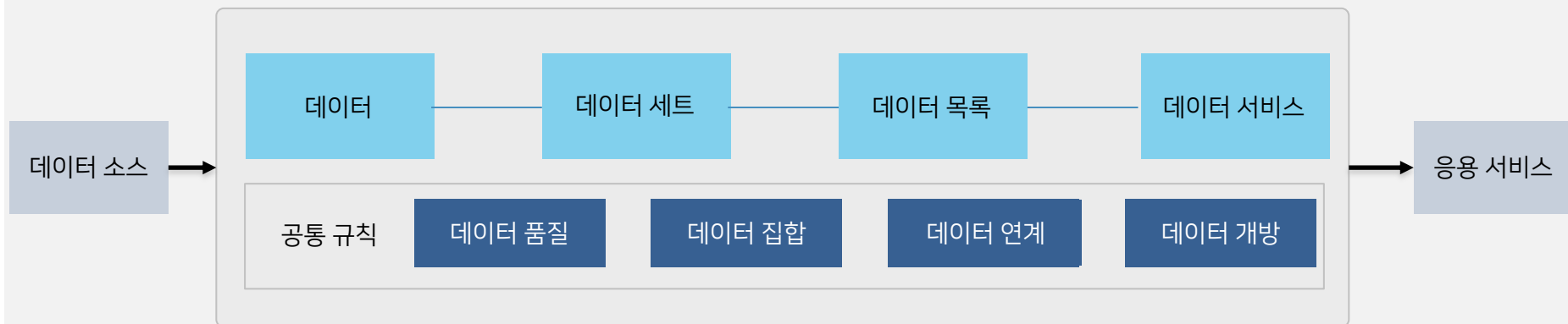
- 행정 분야 데이터 연계를 위한 어휘 모델 개발
- 도로명 주소와 GIS 데이터 연계를 위한 표준 모델 개발
- 문화재 정보 어휘 모델 개발

표준의 확장

데이터맵 프레임워크 표준 개발 - TTA 단체표준 개발중

- (데이터맵 정의) 데이터 서비스(포털)가 운영·관리하는 메타데이터와 이들 사이의 의미적 관계를 표현하기 위한 데이터 구조
- (데이터맵 프레임워크) 데이터맵을 통해 데이터 서비스간 통합·연계를 위한 참조모델, 어휘체계, 서비스 프로파일, 연계 규칙 등의 정의서
- 데이터맵 프레임워크 표준(안) : TTA 단체표준 개발 추진(NP 채택)
 - ✓ 1부: 레퍼런스 아키텍처 (참조모델)
 - ✓ 2부: 데이터 맵 어휘 (TTAK.KO-10.1291)
 - ✓ 3부: DCAT 애플리케이션 프로파일 (DCAT-AP-KR)
 - ✓ 4부: 도메인 메타데이터 연계 규칙

데이터맵 참조 아키텍처 구조



표준의 확장

표준을 넘어 표준화로 ..

규격을 만드는 표준을 넘어 활용할 수 있는 표준 생태계를 만드는 것이 표준화의 궁극적인 목표

TABLE OF CONTENTS

- Abstract
- Status of This Document
- 1. Introduction
- 5. Namespace
- 1. Terminology used in Data Map
- 1. 개념 모델(Conceptual Model)
- 1. Vocabulary overview
 - 1.1 학상 플랫폼의 특성
 - 1.2 Basic example
- 1. Vocabulary specification
 - 1.1 Class: Data Map
 - 1.1.1 Property: has datamap
 - 1.1.2 Property: datamap of
 - 1.1.3 Property: operated by
 - 1.1.4 Property: has catalog
 - 1.1.5 Property: has dataset
 - 1.1.6 Property: has platform
 - 1.1.7 Property: available file format
 - 1.1.8 Property: has column
 - 1.1.9 Property: quality
 - 1.1.10 Property: url
 - 1.1.11 Property: endpoint URL
 - 1.1.12 Property: available Service Type
 - 1.1.13 Property: the size of physical datasets
 - 1.1.14 Property: the number of datasets
 - 1.1.15 Property: the number of organizations
 - 1.1.16 Property: the number of catalogs
 - 1.1.17 Property: the number of views
 - 1.1.18 Property: the number of downloads
 - 1.1.19 Property: the number of columns
 - 1.1.20 Property: the number of distinct columns
 - 1.1.21 Property: the number of tag
 - 1.1.22 Property: the number of registered users
 - 1.1.23 Property: integrating status
 - 1.1.24 Property: available status
 - 1.1.25 Property: harvesting status

Data Map

W3C Editor's Draft 11 October 2021

▼ More details about this document

This version:
<https://some-org.github.io/mySpec/>

Latest published version:
<https://www.w3.org/TR/short/>

Latest editor's draft:
<https://some-org.github.io/mySpec/>

History:
[Commit history](#)

Editors:
[Chaeun Song](#)
[Haram Park](#)

Feedback:
[GitHub some-org/mySpec \(pull requests, new issue, open issues\)](#)

Copyright © 2021 W3C® (MIT, ERCIM, Keio, Beihang). W3C liability trademark and permission document license rules apply.

Abstract

이 표준은 데이터 서비스의 운영과 관리에 대한 메타데이터를 정의하며, 데이터 서비스에 대한 집합적인 정보를 표현하기 위한 RDF 어휘를 정의한다. 데이터별 어휘는 데이터 서비스가 제공하는 관리, 운영, 현황에 대한 수치 정보를 표현하고, 이들과 데이터 서비스의 데이터의 현황을 쉽게 파악하고, 개별적인 데이터세트를 탐색하는데 활용할 수 있다.

Status of This Document

This section describes the status of this document at the time of its publication. Other documents may supersede this document. A list of current W3C publications and the latest revision of this technical report can be found in the [W3C technical reports index](https://www.w3.org/TR/) at <https://www.w3.org/TR/>.

이 문서는 한국정보통신기술협회(TTA)에 의해 검토되었다. 현재 버전은 ITAK-KO-10-1291X에 의해 관리되고 있으며, 최신 개정판은 TTA에서 찾을 수 있다. 데이터별 어휘는 Data Catalog Vocabulary (VOCAB-DCAT) schema.org (SCHEMA-ORG) 등과 같이 기존 어휘를 용이하게 통합하여 구축하였다. 참조한 어휘에 대해서는 reference에서 제공하는 공식 문서를 확인할 수 있다. 이 문서에 대한 의견, 질문이 있으면 hike.cau@gmail.com으로 연락할 수 있다.

[GitHub Issues](#) are preferred for discussion of this specification.

Publication as an Editor's Draft does not imply endorsement by the W3C Membership.

This is a draft document and may be updated, replaced or obsoleted by other documents at any time. It is

Respec을 이용하여 datamap 어휘 명세서 구축

Derived by Respec

koad:Province

<http://data.datahub.kr/administration/administrative-division/def/Province>

Namespace: <http://data.datahub.kr/administration/administrative-division/def/> Recommended prefix: [koad](#).

<code>rdf:type</code>	<code>owl:Class</code>
<code>rdfs:label</code>	<code>lang:en</code> Do <code>lang:en</code> Province <code>lang:zh</code> 道 <code>lang:ko</code> 도
<code>skos:prefLabel</code>	<code>lang:en</code> Do <code>lang:en</code> Province <code>lang:zh</code> 道 <code>lang:ko</code> 도
<code>rdfs:comment</code>	<code>lang:ko</code> 도(道)는 가장 상위의 행정구역 중 하나이다. 도에는 경기도, 강원도, 충청남도, 충청북도, 전라남도, 전라북도, 경상남도, 경상북도, 제주특별자치도가 있다.

Lookup 서버를 이용한 koad:Province 탐색 예시

제안

법정부 데이터 탐색 플랫폼

데이터 카탈로그와 탐색을 지원하는 Discovery-powered Data Analytics

