

TTA Standard

정보통신단체표준(국문표준)

TTAK.KO-10.1291-Part2

제정일: 2022년 12월 07일

데이터맵 - 제2부: 데이터맵 기술을
위한 가이드라인

Datamap - Part 2: Guideline for Describing
Datamap Vocabulary

표준초안 검토 위원회 빅데이터 프로젝트그룹(PG1004)

표준안 심의 위원회 지능정보기반 기술위원회(TC10)

	성명	소속	직위	위원회 및 직위
표준(과제) 제안	김학래	중앙대학교	교수	PG1004 위원
	송채은	중앙대학교	연구원	
	박하람	중앙대학교	연구원	
표준 초안 에디터	김학래	중앙대학교	교수	PG1004 위원
	김장원	군산대학교	교수	PG1004 위원
	전속현	TTA	책임연구원	PG1004 위원
사무국 담당	김영성	TTA	선임	-

본 문서에 대한 저작권은 TTA에 있으며, TTA와 사전 협의 없이 이 문서의 전체 또는 일부를 상업적 목적으로 복제 또는 배포해서는 안 됩니다.

본 표준 발간 이전에 접수된 지식재산권 협약서 정보는 본 표준의 '부록(지식재산권 협약서 정보)'에 명시하고 있으며, 이후 접수된 지식재산권 협약서는 TTA 웹사이트에서 확인할 수 있습니다. 준용표준인 경우 해당 표준화기구 또는 단체의 웹사이트에서 이를 확인해야 합니다.

본 표준과 관련하여 접수된 협약서 외의 지식재산권이 존재할 수 있습니다.

발행인 : 한국정보통신기술협회 회장

발행처 : 한국정보통신기술협회

13591, 경기도 성남시 분당구 분당로 47

Tel : 031-724-0114, Fax : 031-724-0109

발행일 : 2022. 12. 07.

서 문

1 표준의 목적

이 표준의 목적은 정부와 기업에서 운영하는 데이터 포털 또는 플랫폼과 같은 데이터 서비스의 대규모 데이터 세트 관리를 위해 데이터맵 어휘를 사용하여 데이터맵을 기술하기 위한 기능적 요소를 설명하고, 데이터맵 어휘를 적용하는 데 필요한 자세한 방법을 소개한다.

2 주요 내용 요약

이 표준은 데이터맵 어휘의 적용 범위, 데이터맵의 기술 방법, 데이터맵 사이의 연계와 통합, 도메인별 유스케이스를 정의한다.

3 인용 표준과의 비교

해당 사항 없음

Preface

1 Purpose

Data services such as data portals or platforms operated by governments and enterprises provide large and diverse datasets. This standard describes the functional elements for constructing datamaps using the datamap vocabulary, and introduces detailed methods required to apply the individual vocabulary of the datamap.

2 Summary

This standard describes the scope of the Datamap vocabulary, methods for constructing a set of datamaps, interlinking and integrations between heterogeneous Datamaps, and use cases for each domain.

3 Relationship to Reference Standards

None.

목 차

1	적용 범위	1
2	인용 표준	1
3	용어 정의	1
4	약어	2
5	데이터맵 어휘의 개요	2
5.1	배경	2
6	데이터맵 기술	3
6.1	단일 데이터맵으로 기술하기	3
6.2	주제별 데이터맵으로 기술하기	4
6.3	데이터맵을 연계하기	5
6.4	분산되어있는 데이터 카탈로그를 데이터맵으로 기술하기	7
6.5	기관별로 데이터맵 기술하기	8
7	데이터맵의 통계적 수치	9
7.1	데이터맵의 통계적 수치 기술하기	10
7.2	SPARQL 질의문으로 기술하기	13
부록 I	데이터맵 기술 사례	15
부록 II-1	지식재산권 협약서 정보	20
II-2	시험인증 관련 사항	21
II-3	본 표준의 연계(family) 표준	22
II-4	참고 문헌	23
II-5	영문표준 해설서	24
II-6	표준의 이력	25

데이터맵 어휘 기술을 위한 가이드라인 (Guideline for Describing Datamap Vocabulary)

1 적용 범위

데이터맵 어휘는 개별적인 데이터 서비스의 운영·관리적 정보를 기계가 읽을 수 있는 메타데이터로 표현하기 위한 목적을 갖고 있다. 데이터 서비스에 따라 데이터맵을 구성하는 방식이 다양할 수 있다. 이 표준은 데이터맵을 기술하기 위해 데이터맵 어휘에 포함된 구성요소를 상세하게 소개한다. 한편, 데이터맵 어휘에서 사용하는 기존 어휘를 소개하고, 데이터맵 어휘와 연계하는 방법을 사례를 통해 설명한다.

2 인용 표준

해당 사항 없음

3 용어 정의

3.1 SPARQL 엔드포인트

SPARQL 1.1 권고안을 기반으로 데이터와 데이터 목록에 대한 연합 질의 (federated query)를 지원하는 서비스

3.2 데이터맵(Datamap)

데이터 서비스가 운영×관리하는 메타데이터와 이들 사이의 의미적 관계를 표현하기 위한 데이터 구조

3.3 데이터맵 어휘(Datamap Vocabulary)

데이터맵의 구성 요소와 그 관계를 의미적으로 표현하기 위한 어휘

3.4 데이터 서비스(Data Service)

데이터와 데이터 목록에 접근할 수 있는 엔드포인트 서비스 제공 (예: 데이터 포털)

3.5 데이터 목록(Data Catalog)

특정한 데이터 서비스가 보유한 모든 데이터에 대한 메타데이터

3.6 데이터 세트(Dataset)

하나의 행에 하나의 관측치가 포함되도록 구성된 데이터의 집합

3.7 데이터 포털(Data Portal)

사용자가 데이터 세트에 접근할 수 있도록 지원하는 온라인 데이터 플랫폼

3.8 메타데이터(Metadata)

데이터에 대한 데이터. 다른 데이터를 기술하기 위해 사용되는 데이터

4 약어

DCAT	Data Catalog Vocabulary
DM	Datamap
IRI	Internationalized Resource Identifier
OWL	Web Ontology Language
RDF	Resource Description Framework
SPARQL	SPARQL Protocol and RDF Query Language
URI	Uniform Resource Identifier

5 데이터맵 어휘의 개요

5.1 배경

전 세계적으로 데이터의 급격한 증가로 데이터를 체계적으로 관리하기 위한 데이터 거버넌스와 기술적 인프라스트럭처의 필요성이 증가하고 있다. 데이터 포털, 데이터 플랫폼, 또는 데이터 허브는 데이터 세트를 제공하고, 사용자가 데이터 세트를 탐색하고 다운로드할 수 있는 기능을 제공한다. 일반적으로 이런 종류의 웹사이트를 ‘데이터 서비스’로 정의한다. 데이터 서비스는 데이터 세트와 다양한 수준의 메타데이터를 함께 포함하고 있다. 예를 들어, 하나의 도메인 데이터 세트는 데이터의 제공기관, 생성날짜, 분류체계와 같은 메타데이터를 갖고 있고, 유사한 성격이나 분류를 갖는 데이터 세트는 컬렉션으로 묶어서 제공할 수 있다.

한편, 개별적인 데이터 세트와 관계없이 데이터 서비스 자체에 대한 메타데이터가 필요하다. 데이터 서비스의 관리기관, 보유 데이터 세트의 규모, 데이터 세트의 제공기관, 제공하는 데이터 유형은 데이터 서비스에 대한 고유한 메타데이터다. 일반적으로 데이터 서비스는 데이터 세트에 대한 메타데이터를 기술하기 위한 공통 항목을 제공한다. 반면, 데이터 서비스에 대한 메타데이터는 현존하는 표준을 적용하는 데 한계가 있다. 데이터 포털과 같은 데이터 서비스가 확산되면서 웹 수준에서 이종의 데이터를 서로 연계하기 위한 다양한 시도가 있다. 그러나 데이터 포털 간 데이터의 연계를 위해 시스템에 대한 논의와 별개로 데이터 포털의 구성요소를 구체화하고 이를 위한 표준 기반의 접근이 요구되고 있다.

6 데이터맵 기술

일반적으로 데이터 서비스는 데이터 세트, 제공기관, 사용자 등 다양한 정보 개체를 포함

하고 있다. 정보 개체는 데이터 서비스에서 제공하는 기능에 따라 다양한 방식으로 구현된다. 예를 들어, 데이터 세트는 특정한 분류체계를 메타데이터로 표현하거나, 별도의 주제영역을 정의하여 컬렉션으로 관리할 수 있다. 이와 유사하게 제공기관은 데이터 세트에 대한 메타데이터로 정의하거나, 기관을 정보 개체로 표현해서 데이터 세트와 관계를 구체적으로 표현할 수 있다.

데이터맵 어휘는 데이터 서비스에 대한 메타데이터를 표현하기 위한 목적이며, 개별 데이터 서비스의 기능과 특징에 따라 서로 다른 수준에서 데이터맵을 구성할 수 있다. 데이터맵을 구성하는 모든 정보 개체는 URI 형식의 식별자로 정의하기 때문에, 서로 다른 유형의 데이터맵에 있는 모든 정보는 의미적으로 연결된다. 한편, 서로 다른 목적으로 구성된 데이터맵은 하나의 데이터맵으로 통합할 수 있다.

6.1 단일 데이터맵으로 기술하기

데이터 서비스에서 제공하는 모든 정보는 하나의 데이터맵 (dm:Datamap) 클래스로 표현한다. 데이터맵을 구성하는 가장 기본적인 방법이며, 모든 정보 개체는 하나의 데이터맵 클래스와 연결된다. 이때, 수량을 표현하는 속성값은 데이터 서비스에 있는 해당 정보 개체의 값을 표현한다. 예를 들어, 공공데이터 포털(<http://data.go.kr>)은 하나의 데이터맵 (:datamap1)으로 표현할 수 있다. 이때, 공공데이터 포털에 있는 데이터 카탈로그와 데이터 세트는 하나의 데이터맵에 포함된다. 공공데이터 포털을 하나의 데이터맵으로 표현한 예시는 다음과 같다.

```

:datamap1 rdf:type dm:Datamap ;
          rdfs:label "공공데이터 포털"@ko ;
          rdfs:label "Public Data Portal"@en ;
          dm:datamapStat :datamapStat1 .

:datamapStat1 rdf:type dm:DatamapStat ;
              dm:numberOfCatalog "137"^^xsd:NonNegativeInteger ;
              dm:numberOfDataset "68,967"^^xsd:NonNegativeInteger .

:datamapStat-1 rdf:type dm:DatamapStat ;
               dm:numberOfCatalog "2"^^xsd:NonNegativeInteger ;
               dm:numberOfDataset "5"^^xsd:NonNegativeInteger .
    
```

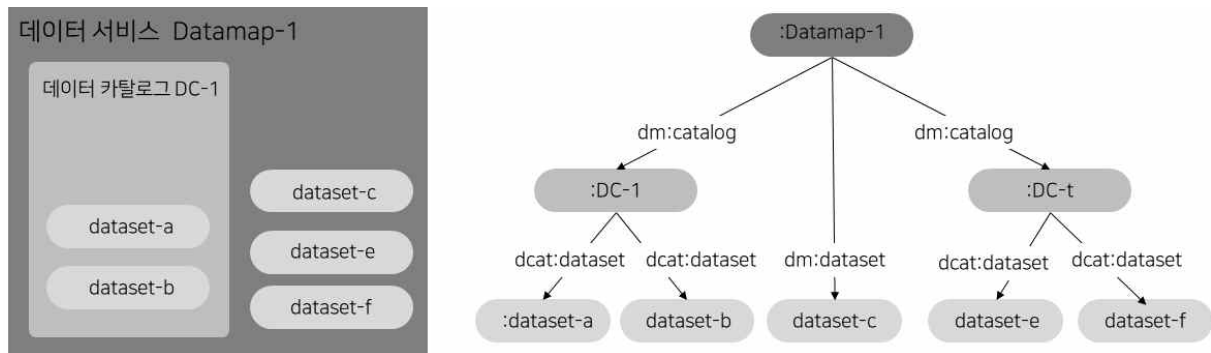
6.2 주제별 데이터맵으로 기술하기

일반적으로 데이터 서비스는 여러 개의 주제를 갖고 있고, 개별 주제는 다시 여러 개의 데이터 세트를 갖고 있다. 이런 유형을 데이터 컬렉션 또는 데이터 카탈로그라고 한다.

DCAT은 데이터 카탈로그를 기술하기 위해 dcat:Catalog를 정의하고 있다. 명시적으로 기술된 하나의 데이터 카탈로그가 있으면, 해당 데이터 카탈로그 (dcat:Catalog)를 dm:catalog 속성을 이용해서 데이터맵과 연결한다. 데이터맵의 카탈로그와 데이터 세트를 기술하기 위해 다음의 유형을 고려할 수 있다 (그림 6-1 참고).

1) 명시적으로 데이터 카탈로그가 존재하지 않고 데이터 세트만 있으면, 임의의 카탈로그를 생성하고 해당 데이터 세트는 dcat:dataset 속성으로 연결한다. 그림 6-1에서 “:DC-t”는 데이터 서비스에 해당 카탈로그가 존재하지 않지만, 데이터 세트 dataset-c, dataset-e, dataset-f를 포함하기 위해 가상으로 생성된 인스턴스다.

2) 데이터맵과 데이터 세트는 가상의 데이터 카탈로그를 정의하지 않고 직접 연결한다. 예를 들어, 데이터 세트 dataset-c, dataset-e, dataset-f는 dm:dataset 속성으로 데이터맵 Datamap-1과 직접 연결한다.



(그림 6-1) 데이터 서비스의 카탈로그와 데이터 세트

```

:Datamap-1 rdf:type dm:Datamap ;
    rdfs:label "Datamap-1"@ko ;
    dm:datamapStat :datamapStat1 ;
    dm:catalog :DC-1, :DC-t ;
    dm:dataset :dataset-c ;
    dm:catalog :DC-1, :DC-t .

:DC-1 rdf:type dcat:Catalog ;
    rdfs:label "DC-1" ;
    dct:description "데이터 서비스에 명시적으로 존재하는 카탈로그" ;
    dcat:dataset :dataset-a, :dataset-b .

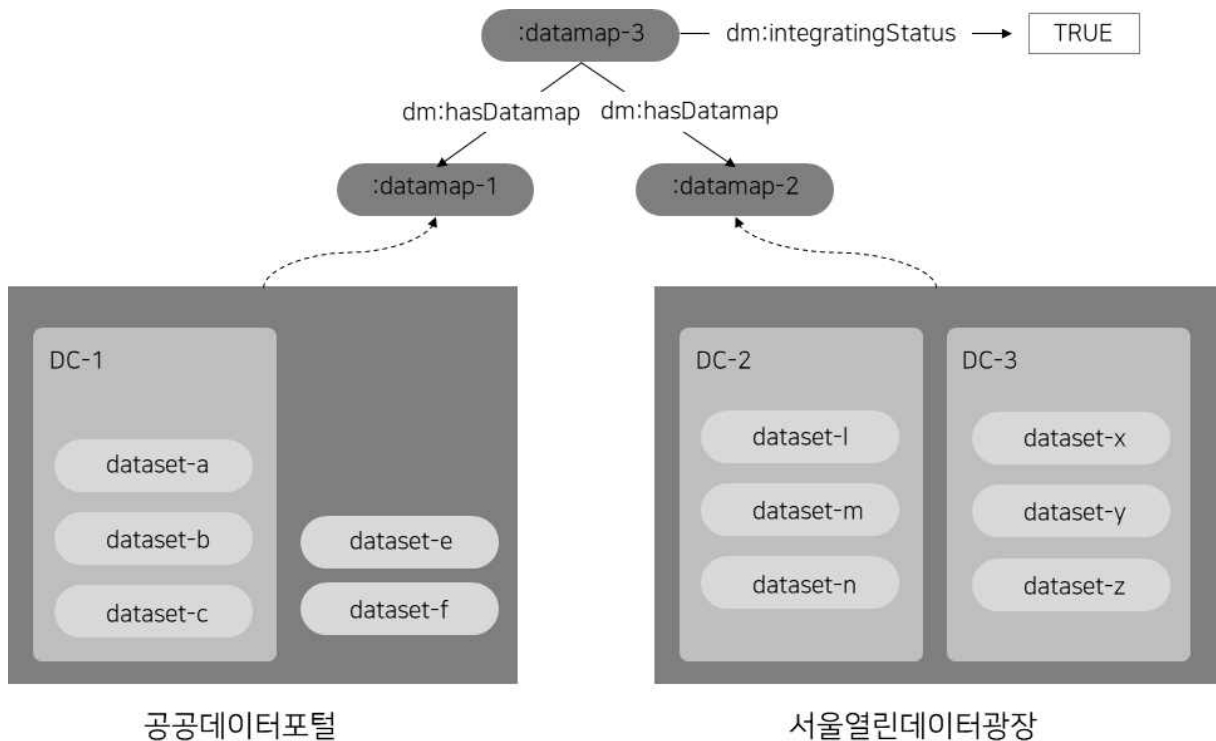
:DC-t rdf:type dcat:Catalog ;
    rdfs:label "DC-t" ;
    dct:description "임의로 생성한 카탈로그" ;
    
```

dcat:dataset :dataset-d, :dataset-f .

6.3 데이터맵을 연계하기

단일 데이터 서비스는 하나 이상의 데이터 카탈로그를 가질 수 있다. 만약, 데이터 카탈로그가 중요한 목적이 있거나, 보유하는 데이터 세트가 많으면, 데이터맵을 데이터 카탈로그를 기준으로 개별적으로 생성할 수 있다. 이 경우, 개별 데이터맵은 dm:hasDatamap 속성으로 통합된 데이터맵으로 연결되고, 통합된 데이터맵은 dm:dm:integratingStatus 속성에 “integrated” 값을 부여해 개별적으로 구성된 데이터맵과 구분된다. 통합된 데이터맵의 수치를 표현하는 속성은 개별 데이터맵의 값을 합산해서 기술한다.

서로 다른 데이터 서비스에서 기술한 데이터맵은 상위 수준의 새로운 데이터맵으로 통합될 수 있다. 새롭게 구성된 데이터맵은 통합되는 데이터맵의 모든 구성요소를 포함하고, 수치 정보는 합산해서 표현한다. 예를 들어, 공공데이터 포털(<http://data.go.kr>)과 서울열린데이터광장(<http://data.seoul.go.kr>)의 정보가 각각 데이터맵(dm:Datamap)으로 표현되어 있다면, 데이터맵을 서로 연계하여 새로운 데이터맵(dm:Datamap)을 만들 수 있다. 통합된 새로운 데이터맵(:datamap3)은 dm:integratingStatus 속성이 “True”이다. 개별 데이터맵 :datamap1과 :datamap2는 dm:hasDatamap 속성으로 :datamap3과 연결된다.



(그림 6-2) 분산되어 있는 데이터맵을 통합한 데이터맵

```

:datamap-1 rdf:type dm:Datamap ;
    rdfs:label "공공데이터 포털"@ko ;
    rdfs:label "Public Data Portal"@en ;
    schema:mainEntityOfPage <http://data.go.kr> ;
    dm:catalog :DC-1 ;
    dm:dataset :dataset-e, :dataset-f ;
    dm:datamapStat :datamapStat1 .

:DC-1 rdf:type dcat:Catalog ;
    dcat:dataset :dataset-a, :dataset-b, :dataset-c .

:datamapStat-1 rdf:type dm:DatamapStat ;
    dm:numberOfCatalog "1"^^xsd:NonNegativeInteger ;
    dm:numberOfDataset "5"^^xsd:NonNegativeInteger .

:datamap-2 rdf:type dm:Datamap ;
    rdfs:label "서울열린데이터광장"@ko ;
    rdfs:label "Public Data Portal"@en ;
    schema:mainEntityOfPage <http://data.seoul.go.kr> ;
    dm:catalog :DC-2, :DC-3 ;
    dm:datamapStat :datamapStat2 .

:DC-2 rdf:type dcat:Catalog ;
    dcat:dataset :dataset-l, :dataset-m, :dataset-n .

:DC-3 rdf:type dcat:Catalog ;
    dcat:dataset :dataset-x, :dataset-y, :dataset-z.

:datamapStat-2 rdf:type dm:DatamapStat ;
    dm:numberOfCatalog "2"^^xsd:NonNegativeInteger ;
    dm:numberOfDataset "6"^^xsd:NonNegativeInteger .

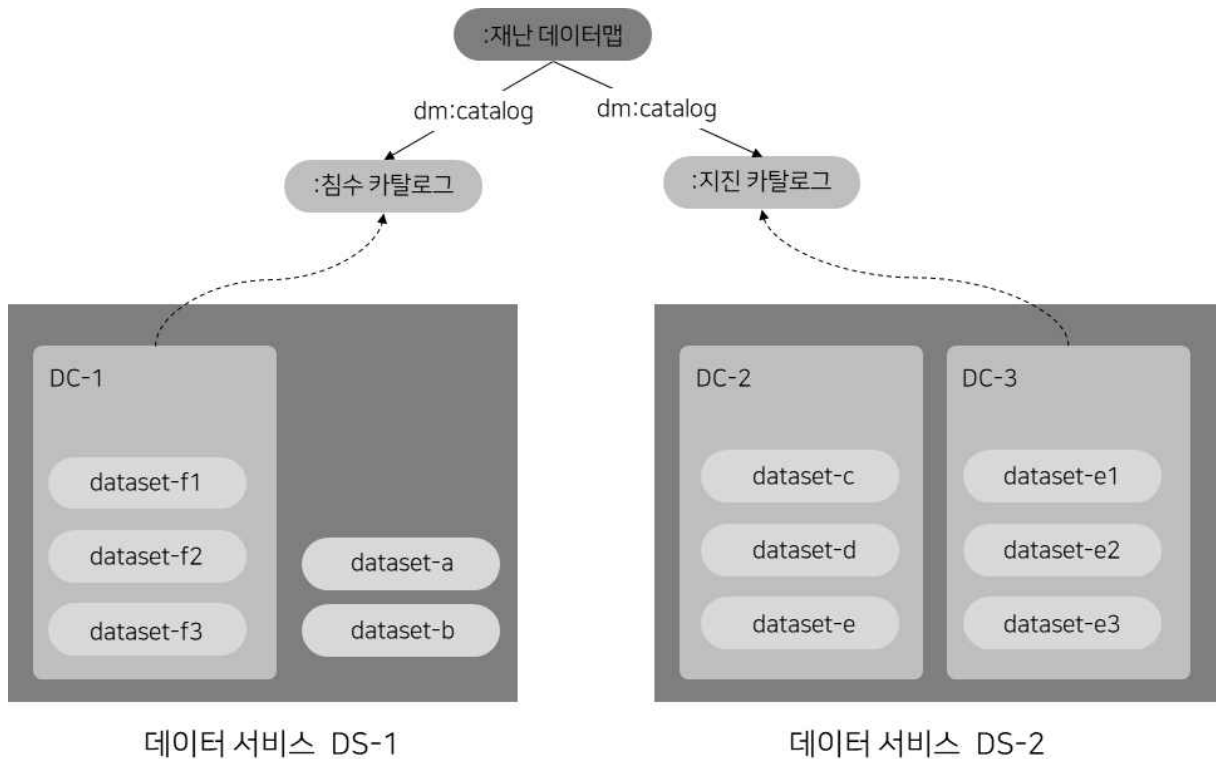
:datamap3 rdf:type dm:Datamap ;
    dm:integratingStatus true ;
    dm:hasDatamap :datamap1, :datamap2 ;
    rdfs:label "통합된 데이터맵"@ko ;
    rdfs:label "Integration of datamaps"@en ;
    dm:datamapStat :datamapStat3 .
    
```

```

:datamapStat3 rdf:type dm:DatamapStat ;
              dm:numberOfCatalog "3"^^xsd:NonNegativeInteger ;
              dm:numberOfDataset "11"^^xsd:NonNegativeInteger .
    
```

6.4 분산되어 있는 데이터 카탈로그를 데이터맵으로 기술하기

DCAT으로 기술된 데이터 카탈로그가 서로 다른 데이터 서비스에 있는 경우, 가상의 데이터맵을 생성하여 연결할 수 있다. 예를 들어, 침수, 지진에 대한 데이터 카탈로그 (dcat:Catalog)가 서로 다른 데이터 서비스에 있다고 가정해 보자. 새롭게 생성하는 데이터맵은 dm:catalog 속성으로 분산적으로 존재하는 데이터 카탈로그를 연결하고, 필요한 속성을 기술할 수 있다. DCAT의 의미 수준을 만족하는 데이터 카탈로그는 추가적인 작업 없이 데이터맵으로 연계할 수 있다.



(그림 6-3) 분산되어 있는 카탈로그를 통합한 데이터맵

```

:datamap-4 rdf:type dm:Datamap ;
           dm:catalog :DC-1, :DC-3 ;
           rdfs:label "재난 데이터맵"@ko ;
           rdfs:label "Datamap of Disaster"@en ;
           dm:datamapStat :datamapStat4 .
    
```

```

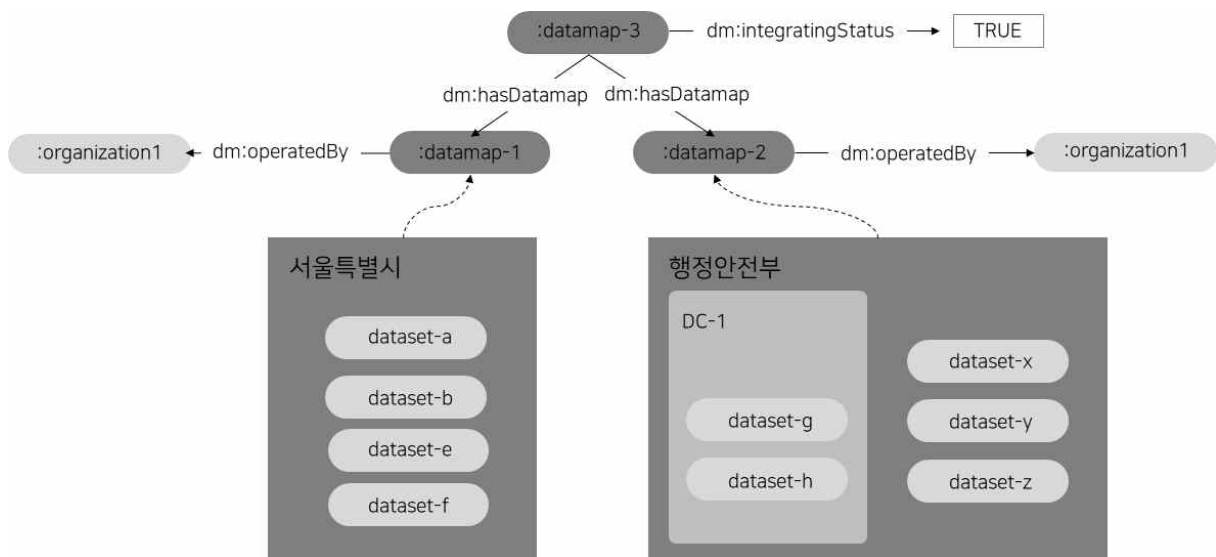
:datamapStat-4 rdf:type dm:DatamapStat ;
    dm:numberOfCatalog "2"^^xsd:NonNegativeInteger ;
    dm:numberOfDataset "6"^^xsd:NonNegativeInteger .

:DC-1 rdf:type dcat:Catalog ;
    rdfs:label "침수 카탈로그"@ko ;
    rdfs:label "Catalog of Flooding"@en ;
    dcat:dataset :dataset-f1, :dataset-f2, :dataset-f3 ;
    dm:catalogStat :catalogStat1 .

:DC-3 rdf:type dcat:Catalog ;
    rdfs:label "지진 카탈로그"@ko ;
    rdfs:label "Catalog of Earthquake"@en ;
    dcat:dataset :dataset-e1, :dataset-e2, :dataset-e3 .
    
```

6.5 기관별로 데이터맵 기술하기

여러 기관이 데이터 세트를 제공하는 데이터 서비스는 기관 중심으로 데이터맵을 기술할 수 있다. 기관에 대한 데이터맵은 해당 기관이 제공하는 데이터 세트와 관련 메타데이터의 관계를 표현한다. 데이터 서비스에 대한 데이터맵은 기관별로 구성된 데이터맵을 포함한다.



(그림 6-4) 기관별 데이터맵을 통합한 데이터맵

```

:datamap-1 rdf:type dm:Datamap ;
    rdfs:label "서울특별시 데이터맵"@ko ;
    dm:operatedBy :organiation1 ;
    dm:dataset :dataset-a, :dataset-b, :dataset-c ;
    dm:datamapStat :datamapStat1 .

:organization1 rdf:type foaf:Organization ;
    rdfs:label "서울특별시" .

:datamapStat-1 rdf:type dm:DatamapStat ;
    dm:numberOfDataset "4"^^xsd:NonNegativeInteger .

:datamap-2 rdf:type dm:Datamap ;
    rdfs:label "행정안전부 데이터맵"@ko ;
    dm:operatedBy :organiation2 ;
    dm:catalog :DC-1 ;
    dm:dataset :dataset-x, :dataset-y, :dataset-z ;
    dm:datamapStat :datamapStat2 .

:organization2 rdf:type foaf:Organization ;
    rdfs:label "행정안전부" .

:DC-1 rdf:type dcat:Catalog ;
    dcat:dataset :dataset-g, :dataset-h .

:datamapStat-2 rdf:type dm:DatamapStat ;
    dm:numberOfCatalog "1"^^xsd:NonNegativeInteger ;
    dm:numberOfDataset "2"^^xsd:NonNegativeInteger .

:datamap-3 rdf:type dm:Datamap ;
    dm:integratingStatus true ;
    dm:hasDatamap :datamap-1, :datamap-2 ;
    rdfs:label "통합된 데이터맵"@ko ;
    rdfs:label "Integration of datamaps"@en ;
    dm:datamapStat :datamapStat3 .

:datamapStat3 rdf:type dm:DatamapStat ;
    dm:numberOfCatalog "1"^^xsd:NonNegativeInteger ;
    dm:numberOfDataset "9"^^xsd:NonNegativeInteger .

```

7 데이터맵의 통계적 수치

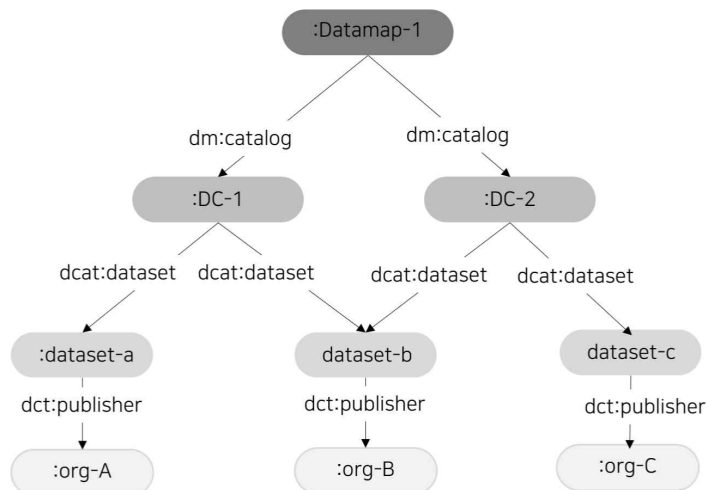
데이터맵은 데이터 포털과 같은 데이터 서비스가 관리하는 메타데이터를 표현한다. 데이터 서비스가 관리하는 메타데이터 중 수치를 파악할 수 있는 속성을 데이터맵 어휘는 dm:DatamapStat과 dm:CatalogStat으로 표현한다. 데이터맵 또는 데이터 카탈로그가 보유한 데이터 카탈로그, 데이터 세트, 컬럼, 데이터 제공기관, 데이터 세트의 메타데이터, 데이터 세트의 키워드를 집계한 값을 기술한다.

<표 7-1> 데이터맵의 클래스와 속성

클래스	속성
dm:DatamapStat	dm:numberOfCatalog, dm:numberOfDataset, dm:numberOfColumn, dm:numberOfOrganization, dm:numberOfMetadata, dm:numberOfTag
dm:CatalogStat	dm:numberOfDataset, dm:numberOfColumn, dm:numberOfOrganization, dm:numberOfMetadata, dm:numberOfTag

7.1 데이터맵의 통계적 수치 기술하기

데이터맵이 보유한 데이터 카탈로그와 데이터 세트의 수량 정보와 데이터 세트를 관리하는 기관의 수량을 집계하여 dm:DatamapStat으로 기술한다. 개별 데이터 카탈로그가 보유한 데이터 세트와 관리기관의 수량은 dm:CatalogStat으로 기술해 카탈로그별 메타데이터 현황을 파악할 수 있다. 데이터맵의 통계적 수치는 기본적으로 중복을 제거한 총 수량을 의미한다.



(그림 7-1) 단일 데이터맵의 메타데이터

```

:Datamap-1 rdf:type dm:Datamap ;
    rdfs:label "Datamap-1"@ko ;
    dm:datamapStat :datamapStat1 ;
    dm:catalog :DC-1, :DC-2 ;
    dm:dataset :dataset-a :dataset-b :dataset-c .

Catalog-1 rdf:type dcat:Catalog ;
    dm:catalogStat :CatalogStat-1 .

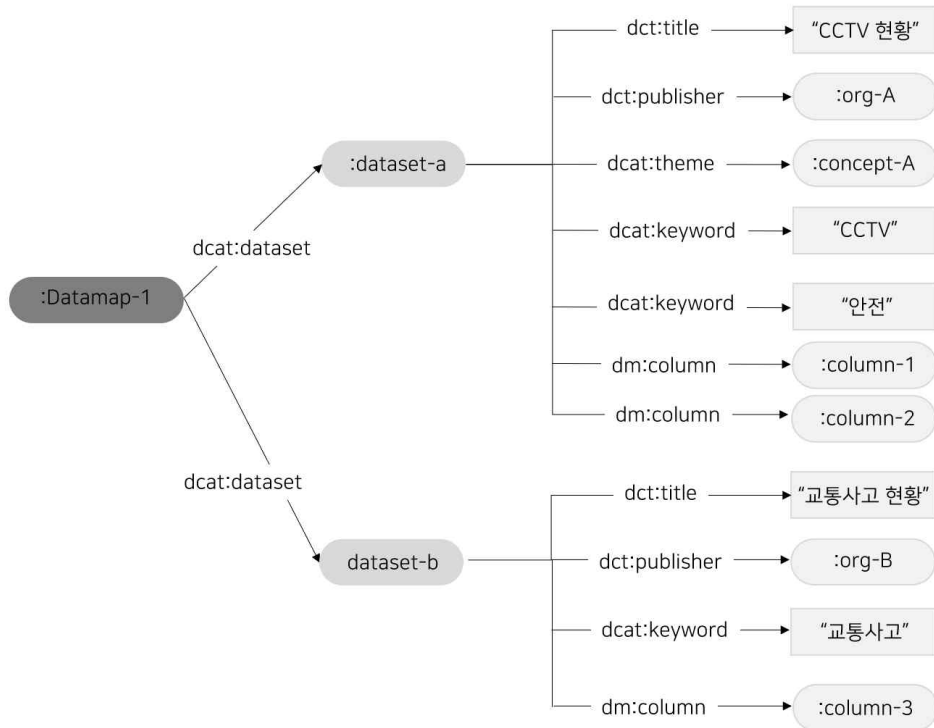
Catalog-2 rdf:type dcat:Catalog ;
    dm:catalogStat :CatalogStat-2 .

:datamapStat-1 rdf:type dm:DatamapStat ;
    dm:numberOfCatalog "2"^^xsd:NonNegativeInteger ;
    dm:numberOfDataset "3"^^xsd:NonNegativeInteger ;
    dm:numberOfOrganization "3"xsd:NonNegativeInteger .

:CatalogStat-1 rdf:type dm:CatalogStat ;
    dm:numberOfDataset "2"^^xsd:NonNegativeInteger ;
    dm:numberOfOrganization "2"xsd:NonNegativeInteger .

:CatalogStat-2 rdf:type dm:CatalogStat ;
    dm:numberOfDataset "2"^^xsd:NonNegativeInteger ;
    dm:numberOfOrganization "2"xsd:NonNegativeInteger .
    
```

dm:numberOfOrganization, dm:numberOfColumn, dm:numberOfTag 속성은 데이터맵이 보유한 개별 데이터 세트의 dct:publisher, dm:column, dcat:keyword의 값을 중복없이 집계한다. 데이터 세트를 표현하는 메타데이터 항목의 수량은 dm:numberOfMetatdata으로 표현하고, 이 속성을 통해 데이터 서비스가 데이터 세트에 대한 정보를 얼마나 제공하는지 알 수 있다.



(그림 7-2) 데이터맵과 데이터 세트의 메타데이터

```
:Datamap-1 rdf:type dm:Datamap ;
  rdfs:label "Datamap-1"@ko ;
  dm:datamapStat :datamapStat1 ;
  dm:dataset :dataset-a :dataset-b .
```

```
:dataset-a rdf:type dcat:Dataset ;
  dct:title "CCTV 현황" ;
  dct:publisher :org-A ;
  dcat:theme :concept-A ;
  dcat:keyword "CCTV", "안전" ;
  dm:column :column-1, :column-2 .
```

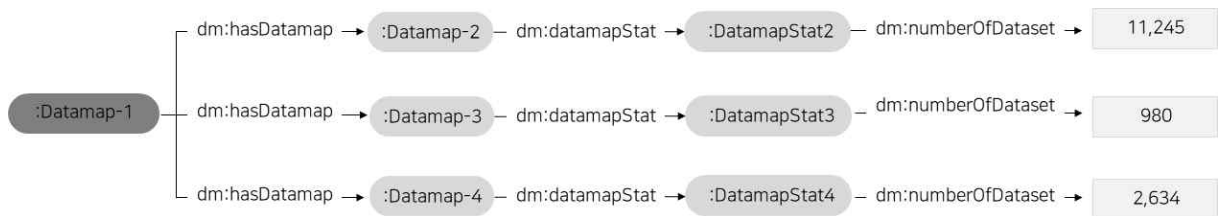
```
:dataset-b rdf:type dcat:Dataset ;
  dct:title "교통사고 현황" ;
  dct:publisher :org-B ;
  dcat:theme :concept-A ;
  dcat:keyword "교통사고" ;
  dm:column :column-3 .
```

```

:datamapStat-1 rdf:type dm:DatamapStat ;
                dm:numberOfDataset "2"^^xsd:NonNegativeInteger ;
                dm:numberOfOrganization "2"xsd:NonNegativeInteger ;
                dm:numberOfTag "3"xsd:NonNegativeInteger ;
                dm:numberOfColumn "3"xsd:NonNegativeInteger .
    
```

7.2 SPARQL 질의문으로 기술하기

데이터맵의 통계적 수치는 데이터맵이 통합되거나 추가되면 주기적으로 변경해야한다. RDF 데이터는 SPARQL 질의를 통해 원하는 값을 추출할 수 있고, 저장된 데이터를 변경할 수도 있다. INSERT를 사용하면, 데이터맵의 데이터 현황을 업데이트할 수 있다. 복수의 데이터맵을 통합하여 새로운 데이터맵을 기술할 경우, dm:DatamapStat을 통해 통계적 수치를 간편히 계산할 수 있다. 중복값을 고려하지 않고 데이터 카탈로그, 데이터 세트, 컬럼, 제공기관 등의 수량을 덧셈한 총합을 새로운 데이터맵의 dm:DatamapStat에 부여하는 방법과 중복을 제거한 고유한 수량의 총합을 집계하는 경우로 구분할 수 있다.



(그림 7-3) 통합된 데이터맵과 데이터 세트 수량

```

PREFIX dm: <http://vocab.datahub.kr/def/datamap/>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dcat: <http://www.w3.org/ns/dcat#>

INSERT {?dmstat1 dm:numberOfDataset ?sum }
WHERE
{
  {
    SELECT (SUM(?dsCount) AS ?sum)
    {
      :datamap1 a dm:Datamap ;
    }
  }
}
    
```

```

        dm:datamapStat ?dmstat1 ;
        dm:hasDatamap ?subDatamap.
    ?subDatamap a dm:Datamap ;
        dm:datamapStat ?subStat .
    ?subStat a dm:DatamapStat ;
        dm:numberOfDataset ?dsCount .
}
}
}

```

중복없는 데이터 세트의 수량을 계산하고 싶다면, 데이터 세트의 URI를 기준으로 집계한다. 아래 질의문은 dm:numberOfDataset의 총합이 아닌, 데이터맵이 보유한 데이터 세트의 고유한 URI의 총 수량을 계산한다.

```

PREFIX dm: <http://vocab.datahub.kr/def/datamap/>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dcat: <http://www.w3.org/ns/dcat#>

INSERT { ?dmstat1 dm:numberOfDataset ?sum }
WHERE
{
  {
    SELECT (COUNT(distinct ?subDataset) AS ?sum)
    {
      :datamap1 a dm:Datamap ;
        dm:datamapStat ?dmstat1 ;
        dm:hasDatamap ?subDatamap.
      ?subDatamap a dm:Datamap ;
        dm:dataset ?subDataset .
    }
  }
}
}
}

```

부 록 I

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

데이터맵 기술 사례

1.1 기술 대상 선정

데이터맵 어휘는 데이터 서비스의 메타데이터를 표현하기 위한 어휘다. 데이터맵은 하나의 데이터 서비스를 대상으로 기술할 수 있지만, 데이터 서비스를 구성하는 정보 개체에 따라 다양하게 기술할 수 있다. <표 1.1-1>은 데이터맵을 기술할 때 포함되어야 하는 항목과 내용을 설명한다.

<표 1.1-1> 데이터맵 메타데이터 작성 항목과 설명

항목	내용
메인페이지 URL	데이터 서비스를 제공하는 웹사이트 주소
데이터맵 URI	데이터맵에 접근할 수 있는 URI
관련 데이터맵 URI	하나 이상의 데이터맵이 연계되어 구성된 복수의 데이터맵
관리기관	데이터맵의 관리를 담당하는 기관이나 사람
제공기관	데이터 서비스를 제공하는 기관이나 사람
생성날짜	데이터맵을 최초로 생성한 날짜 (YYYY-MM-DD로 작성)
갱신날짜	데이터맵을 갱신한 최종 날짜 (YYYY-MM-DD로 작성)
데이터 세트	데이터 서비스에서 제공하는 파일 유형의 데이터 세트 개수 (정수형의 숫자로 표시)
API	데이터 서비스에서 제공하는 API 유형의 데이터 세트 개수 (정수형의 숫자로 표시)
데이터 카탈로그	데이터 서비스에서 제공하는 카탈로그 개수 (숫자로 표시)
라이선스 또는 이용약관	라이선스를 설명하는 URL (URL이 없으면, 라이선스를 텍스트로 기재)
분류체계 방식	데이터 세트를 구분하는 분류 항목 (자체 결합 외부 중 선택)
분류체계 URL	분류체계에 접근할 수 있는 URL
사용어휘	데이터맵을 기술하는데 사용한 어휘 (예: RDF)

1.2 데이터맵 기술 사례 - 공공데이터 포털

공공데이터 포털은 공공기관이 생성 또는 취득하여 관리하고 있는 공공데이터를 한 곳에서 제공하는 데이터 서비스다. 공공데이터 포털은 국민이 쉽고 편리하게 공공데이터를 이용할 수 있도록 파일데이터, 오픈API, 시각화 등 다양한 방식으로 제공하고 있으며, 누구나 쉽고 편리한 검색을 통해 원하는 공공데이터를 빠르고 정확하게 찾을 수 있다. 2022년 2월 현재, 데이터 개방기관은 990개이고, 파일데이터 51,114건, 오픈 API 8,977

건, 표준데이터 세트 8,735건이 제공되고 있다 (2022년 5월 기준).

a) 데이터 서비스의 메타데이터 수집

데이터맵(dm:Datamap) 기술을 위해 데이터 서비스의 메타데이터를 수집한다. 예를 들어, 공공데이터 포털(data.go.kr)을 데이터맵으로 표현하기 위해 수집한 정보는 <표 1.2-1>과 같다 (상세 내용은 변경될 수 있다).

<표 1.2-1> 공공데이터 포털의 메타데이터 현황

항목	내용
메인페이지 URL	http://data.go.kr
데이터맵 URI	http://data.datahub.kr/id/datamap/data-portal-dm
관련 데이터맵 URI	NA
관리기관	중앙대학교 HIKE 연구실
제공기관	행정안전부
생성날짜	2022-03-01
갱신날짜	2022-05-01
데이터 세트	68,897
API	8,977
데이터 카탈로그 수	137
라이선스 또는 이용약관	https://www.data.go.kr/ugs/selectPortalPolicyView.do#use_s tplat
분류체계 방식	자체
분류체계 URL	http://data.datahub.kr/datamap/data-portal-dm/category/
사용어휘	RDF, Datamap, DCAT, Dublin Core Metadata

b) 어휘 재사용

데이터맵 어휘는 기존 RDF 어휘를 재사용해서 기술할 수 있다. 자원에 대한 일반적인 기술 (description)은 추가 어휘를 정의하지 않고 RDF(Resource Description Framework), DCMI(Dublin Core Metadata Initiative)의 메타데이터를 재사용한다. 특히, 데이터 세트와 관련된 어휘는 DCAT-AP-KR 어휘를 사용하는 것을 권장한다. 데이터맵에서 사용하고 있는 어휘는 다음과 같다.

<표 1.2-2> RDF 어휘와 네임스페이스

접두사(Prefix)	네임스페이스(Namespace)
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#

dcat	http://www.w3.org/ns/dcat#
dcatkr	http://vocab.datahub.kr/def/dcat-ap-kr/
dct	http://purl.org/dc/terms/
skos	http://www.w3.org/2004/02/skos/core#
foaf	http://xmlns.com/foaf/0.1/
vcard	http://www.w3.org/2006/vcard/ns#
xsd	http://www.w3.org/2001/XMLSchema#

c) URI 설계

URI(Uniform Resource Identifier)는 자원(resource)을 식별하기 위한 핵심 요소이며, 데이터맵의 구성요소는 일관성 있는 체계로 정의해야 한다. 데이터맵을 위한 URI 체계와 구성요소는 다음과 같다.

<표 1.2-3> 데이터맵 URI 체계

http://{domain}/{datamap}/{id}

- URI 구성요소는 영어 또는 숫자 형식으로 표현하는 것을 권장한다. 개별 구성요소의 이름은 단일 또는 복합 명사로 기술하며, 소문자로 표기한다. 복합 명사는 대시 (i.e. '-') 기호로 연결해서 표기한다 (예: public-administrative).
- {domain}은 데이터맵을 기술하는 기관에서 사용하는 도메인 또는 서브도메인을 사용한다. 실제 운영되고 있는 도메인 또는 서브 도메인을 사용할 것을 권장한다.
- 데이터맵의 유형 {datamap}은 'datamap'으로 표기한다.
- {id}는 데이터맵으로 기술되는 자원을 식별할 수 있는 값을 사용한다. {id} 값은 데이터맵을 기술하는 기관에서 일관성 있게 사용하는 식별자로 표현할 것을 권장한다. 예를 들어, 식별자는 기관명의 영문약자와 기관코드를 연결한 표시 형식을 권장한다 (예: 'cau-53099000').
- 동일한 기관에서 여러 개의 데이터맵을 기술하는 경우, 일련번호를 적용할 수 있다 (예: 'cau-53099000-1').
- {collection} 또는 {sub-collection}은 {domain}이 보유하는 데이터 컬렉션을 구분하기 위한 경로로서 선택적으로 사용할 수 있다. {collection}은 {domain} 뒤에 표기하고, {sub-collection}은 {collection} 뒤에 표기한다. 다양한 주제의 데이터를 URI로 표현하는 경우, {collection}을 사용하여 경로를 구분할 수 있다. 예를 들어, 데이터맵으로 표현하는 데이터 세트 주제가 공공행정, 교육, 환경으로 구분될 경우
 http://{domain}/administration/, http://{domain}/education/,
 http://{domain}/environment/ 로 표현할 수 있다. 하위 분류는
 http://{domain}/environment/earthquake/ 와 같이 표현할 수 있다.

d) RDF 변환

데이터맵 어휘는 RDFS(RDF Schema)로 설계되었고, 모든 데이터맵은 기본적으로 RDF 로 기술해야 한다. 데이터맵 어휘 외 다른 RDF 어휘를 사용한 경우, 반드시 해당 어휘를 명시적으로 설명해야 한다. 그래프 데이터를 직렬화 방식이 다양하지만, 데이터맵의 배포 형식은 RDF/XML, Turtle, JSON-LD, SPARQL 엔드포인트 방식을 권장한다. 파일 형식으로 데이터맵을 배포하려면 유효성 검증을 완료하고 배포해야 한다 (예: W3C RDF Validator: <https://www.w3.org/RDF/Validator/>).

SPARQL 엔드포인트를 제공하는 경우, Endpoint URL을 명시하고, 예제 질의문을 함께 제공해야 한다. 한편, 데이터맵의 구성요소를 설명하기 위한 메타데이터와 다이어그램을 함께 제공하는 것을 고려할 수 있다.

<표 1.2-3> 권장하는 직렬화 방식

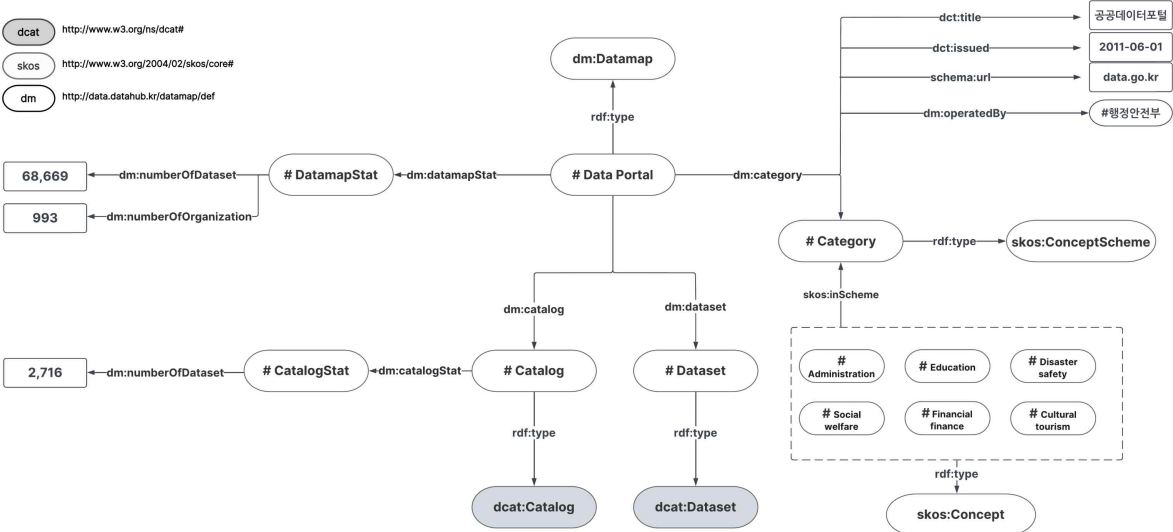
구분	내용
RDF/XML	https://www.w3.org/TR/rdf11-concepts/
Turtle	http://www.w3.org/TR/2014/REC-turtle-20140225/
JSON-LD	https://www.w3.org/TR/json-ld/
SPARQL	https://www.w3.org/TR/2013/REC-sparql11-query-20130321/

예를 들어 ‘공공데이터 포털’은 하나의 데이터맵(dm:Datamap)으로 표현할 수 있고, <표 1.2-1>의 정보를 데이터맵 어휘와 기존 표준어휘로 나타내면 아래와 같다.

```

:/datamap/datamap1 rdf:type dm:Datamap ;
    rdfs:label "공공데이터 포털"@ko ;
    rdfs:label "Public Data Portal"@en ;
    schema:url "data.go.kr" ;
    dct:issued "2013-10-01" ;
    schema:usagelInfo
"https://www.data.go.kr/ugs/selectPortalPolicyView.do#use_stplat" ;
    dm:operatedBy :/datamap/organization1 ;
    dm:datamapStat :/datamap/datamapStat1 ;
:/datamap/organization1 foaf:Organization ;
    rdfs:label "행정안전부" .
:/datamap/datamapStat1 rdf:type dm:DatamapStat ;
    dm:numberOfCatalog "137"^^xsd:NonNegativeInteger ;
    dm:numberOfDataset "68,967"^^xsd:NonNegativeInteger .
    
```

공공데이터 포털의 데이터맵은 (그림 1.2-1)과 같은 그래프 형식으로 시각화할 수 있다.



(그림 1.2-1) 공공데이터 포털에 대한 데이터맵

부 록 II-1

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

지식재산권 협약서 정보

아래에 기재된 지식재산권 협약서 이외에도 본 표준이 발간된 후 접수된 협약서가 있을 수 있으니, TTA 웹사이트에서 확인하시기 바랍니다.

해당 사항 없음

부 록 II-2

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

시험인증 관련 사항

II-2.1 시험인증 대상 여부

해당 사항 없음

II-2.2 시험표준 제정 현황

해당 사항 없음

부 록 II-3

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

본 표준의 연계(family) 표준

II-3.1 TTA.KO-10.1291-Part1, 데이터맵 - 제1부: 데이터맵 어휘

부 록 II-4

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

참고 문헌

아래 기재된 참고 문헌의 발간일이 기재된 경우, 해당 표준(문서)의 해당 버전에 대해서만 유효하며, 연도를 표시하지 않은 경우에는 해당 표준(권고)의 최신 버전을 따른다.

해당 사항 없음.

부 록 II-5

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

영문표준 해설서

해당 사항 없음

부 록 II-6

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

표준의 이력

판수	채택일	표준번호	내용	담당 위원회
제1판	2022.12.07.	제정 TTAK.KO-10.1291-Part2	-	빅데이터 프로젝트그룹 (PG1004)