

주소지식모델 - 제3부: 웹 URI 체계

Address Knowledge Model - Part 3: Web
URI System

표준초안 검토 위원회 빅데이터 프로젝트그룹(PG1004)

표준안 심의 위원회 지능정보기반 기술위원회(TC10)

	성명	소 속	직위	위원회 및 직위
표준(과제) 제안	김학래	중앙대학교	교수	PG1004 특별위원
	박하람	중앙대학교	연구원	
	송채은	중앙대학교	연구원	
	김장원	군산대학교	교수	PG1004 특별위원
	고종신	행정안전부	서기관	
	박재호	행정안전부	사무관	
표준 초안 에디터	김학래	중앙대학교	교수	PG1004 특별위원
	김장원	군산대학교	교수	PG1004 특별위원
	고종신	행정안전부	서기관	PG1004 특별위원
	박재호	행정안전부	사무관	PG1004 특별위원
사무국 담당	이인송	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 확약서 정보는 본 표준의 '부록(지식재산권 확약서 정보)'에 명시하고 있으며, 이후 접수된 지식재산권 확약서는 TTA 웹사이트에서 확인할 수 있습니다. 준용표준인 경우 해당 표준화기구 또는 단체의 웹사이트에서 이를 확인해야 합니다.

본 표준과 관련하여 접수된 확약서 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 : 2023. 12. 06.

서 문

1 표준의 목적

본 표준의 목적은 한국의 주소체계를 주소그래프로 표현하기 위한 URI 설계원칙을 정의하는 것이다. 주소그래프는 주소어휘를 활용하여 다양한 주소구성요소와 그 관계를 표현한다. 그러므로 이 표준은 주소그래프의 어휘, 개체에 대한 웹 URI 체계, 설계원칙, 표기법과 URI 설계 패턴을 기술한다.

2 주요 내용 요약

이 표준은 한국의 주소체계를 기반으로 표현되는 주소 개체들의 식별을 위해서 월드와이드웹에서 사용되는 식별 시스템을 도입하여, 주소와 관련된 정보자원을 위한 웹 URI의 설계 원칙과 패턴을 정의한다. 한국의 주소체계는 주소참조체계, 국가주소정보, 주소지능정보로 구분되며, 각 유형에 따라 주소의 부여 방식과 활용 대상이 다르다. 이 표준은 각 유형별 세부 요소의 식별 방법을 상세히 설명한다.

3 인용 표준과의 비교

해당 사항 없음.

Preface

1 Purpose

The standard aims to establish a web URI system encompassing the address referencing system, national address information, and intelligence information constituting Korea's addressing system. It defines the design principles, notations, and patterns of web URIs based on the types of resources within the address system to identify the resources that make up an address standard.

2 Summary

The standard introduces the identification system used in the World Wide Web to identify address entities based on Korea's addressing system. It defines the design principles and patterns of web URIs for address information. Korea's addressing system consists of the address referencing system, national address information, and address intelligence information. Each type has different allocation methods and utilization targets for addresses. This standard provides detailed explanations on identifying the specific elements for each type. Therefore, by utilizing the identification system defined for the address elements, the identified address web URI system can be applied to integrate and use various knowledge graphs standards.

3 Relationship to Reference Standards

None

목 차

1	적용 범위	1
2	인용 표준	1
3	용어 정의	1
4	약어	1
5	주소정보를 위한 웹 URI 설계원칙	2
5.1	HTTP URI 사용	2
5.2	기계가 읽을 수 있는 형식	2
5.3	변경되지 않는 정보로의 URI 구성	2
5.4	FAIR 데이터 원칙의 준수	3
6	주소의 구성요소 표기법	3
6.1	왼쪽 구성요소의 표기	3
6.2	오른쪽 구성요소의 표기	4
7	URI 설계 패턴	5
7.1	공통사항	5
7.2	도메인 구조	8
7.3	유형별 URI 설계	8
7.4	URI 확인	11
8	사람과 기계가 읽을 수 있는 형식	13
부록 I	웹 URI	15
II-1	FAIR 데이터 원칙	18
III-1	지식재산권 협약서 정보	20
III-2	시험인증 관련 사항	21
III-3	본 표준의 연계(family) 표준	22
III-4	참고 문헌	23
III-5	영문표준 해설서	24
III-6	표준의 이력	25

주소지식모델 - 제3부: 웹 URI 체계 (Address Knowledge Model - Part 3: Web URI System)

1 적용 범위

이 표준은 한국의 주소체계를 주소그래프로 표현하기 위한 URI 설계원칙을 정의한다. 주소그래프는 주소어휘를 활용하여 주소구성요소와 관계를 나타낸다. 이 표준은 주소그래프의 어휘, 웹 URI 체계, 설계원칙, 표기법, URI 설계 패턴을 설명한다. 한편, 한국의 주소체계를 기반으로 주소 개체를 식별하기 위해 월드와이드웹에서 사용되는 식별 시스템을 도입하며, 주소정보를 위한 웹 URI의 설계 원칙과 패턴을 정의한다. 한국의 주소체계는 주소참조체계, 국가주소정보, 주소지능정보로 구분되며, 각 유형에 따라 주소의 부여 방식과 활용 대상이 다르다. 이 표준은 각 유형별 세부 요소의 식별 방법을 상세히 설명한다.

2 인용 표준

해당 사항 없음.

3 용어 정의

TTAx.xx-xx.xxxx.Part1의 3 용어 정의를 준용한다.

4 약어

DCAT	Data Catalog Vocabulary
DCT	DCMI Metadata Terms
DNS	Domain Name System
FAIR	Findable, Accessible, Interoperable, Reusable
HTTP	HyperText Transfer Protocol
HTTPS	Hyper Text Transfer Protocol Secure
IRI	Internationalized Resource Identifier
PSI	Public Sector Information
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URN	Uniform Resource Name

5 주소정보를 위한 웹 URI 설계원칙



(그림 5-1) 주소지식모델 패밀리 표준

본 표준은 주소지식모델 패밀리 표준의 제3부 웹 URI 체계에 해당한다. 웹은 URI를 사용하여 자원을 식별한다. 그러므로 자원을 식별하는 URI는 웹에 존재하는 모든 자원을 연결하는 대규모 네트워크 효과를 만든다. 특히 주소정보는 정부 내부의 정보시스템과 더불어 민간의 다양한 산업영역에서 활용되고 있으므로 주소정보와 관련된 자원은 URI로 식별자를 제공해야 한다. 주소정보를 위한 URI 설계원칙은 다음을 포함한다.

5.1 HTTP URI 사용

데이터의 연결과 탐색 등 월드와이드웹의 장점을 활용하기 위해 주소정보와 관련된 모든 자원의 식별자로 HTTP URI를 제공해야 한다. LD원칙(Heath & Bizer, 2011)에 명시되었듯이, HTTP URI는 식별되는 자원의 상세한 정보에 접근하기 위해 URI를 조회 또는 역참조할 수 있게 한다.

5.2 기계가 읽을 수 있는 형식

HTTP URI 역참조를 위해서 데이터 게시자가 제공하는 자원은 사람이 읽을 수 있는 HTML 방식 또는 기계가 읽을 수 있는 RDF/XML 방식으로 표현해야 한다. 기계가 읽을 수 있는 표현을 하나 이상 제공하면 데이터 사용자는 환경에 맞는 자원의 설명을 선택해서 활용할 수 있다.

5.3 변경되지 않는 정보로의 URI 구성

세션 토큰, 웹페이지의 상태 정보처럼 쉽게 변경되거나, 변경될 것으로 예상되는 정보는 URI에 포함하지 않는다. 따라서, 링크드 데이터 형태로 제공되는 데이터에서 규칙적이고 구조화된 형식의 URI는 사용자로부터 URI로 식별되는 정보에 대해서 신뢰성을 높여줄 수 있는 핵심 요소이다.

5.4 FAIR 데이터 원칙의 준수

링크드 데이터는 본질적으로 상호운용이 가능한 원칙 기반으로 데이터를 포함하고 있으므로 FAIR에서의 제 1 원칙(Interoperability)인 상호운용성 구현에 사용할 수 있다. 상호운용성 원칙은 서로 다른 데이터가 통합되거나, 분석, 저장, 처리를 위해 애플리케이션과 상호운용되는 것을 뜻한다. 주소정보는 행정정보, 디지털트윈, 인공지능의 다양한 데이터와 연결될 수 있으므로 FAIR 데이터 원칙 적용을 권장한다. 따라서, FAIR 데이터 원칙은 주소 웹 URI의 설계와 활용에 전반적인 기준으로 활용할 수 있다. 그러므로 주소 웹 URI는 주소정보와 관련된 개체의 구조화와 공개된 형태의 접근을 지원하고, 사람과 기계가 읽을 수 있는 언어로 표현되어야 한다.

6 주소의 구성요소 표기법

본 표준에 사용된 URI 구성요소에 대한 패턴 표기법은 RFC 6570(Fielding et al., 2012)에 정의된 ‘URI 템플릿’ 명세를 따른다. 대괄호 ‘[’와 ‘]’는 선택적 구성요소를 기술하는데 사용되며, 대괄호 구성요소 뒤에 별표(‘*’)를 사용하면 구성요소가 임의로 반복(0회 이상)하는 것으로 해석한다. URI 패턴은 크게 두 개의 그룹으로 구성된다. 먼저, 왼쪽의 구성요소는 스킴(scheme)과 호스트(host)를 포함하여 ‘http://{domain}/{collection*}’로 표기한다. 오른쪽 구성요소는 자원의 유형에 따라 표기 방식이 다양하다.

6.1 왼쪽 구성요소의 표기

왼쪽 구성요소의 일반 패턴은 다음과 같다. ‘{prefix} = http://{domain}/{collection*}’

여기서 {domain}은 DNS의 이름이다. {domain} 기관은 위임된 URI 공간에서 {subdomain}.{domain} 형식의 하위 도메인을 만들 수 있다. 하위 도메인의 거버넌스는 상위 도메인에 대한 권한의 범위에 포함되거나 하위 거버넌스 권한에 위임될 수 있다.

{/collection*}은 위임된 URI 공간에 대한 관리 권한의 위임 지점 역할을 한다. 일반적으로 하나 또는 두 개의 짧은 URI 경로 세그먼트를 갖는다. {/collection*}의 경로 세그먼트는 {type} 구성요소에서 사용되는 리터럴 값(예시: “def”, “id”, “doc”, “data”, “so”)을 사용하지 않는 것을 권장한다.

주소는 왼쪽 구성요소는 주소그래프의 서비스에 대한 축약된 키워드로 지정한다. 예를 들어, 행정안전부에서 관리하는 주소와 관련된 URI가 특정한 서브 도메인(예: data)으로 표현한다면, {domain}은 http://data.juso.go.kr로 표현할 수 있다. 도로명주소, 사물주소, 공간주소는 {/collection*}을 이용해서 개별 주소의 유형을 표기한다.

<표 6-1> 주소의 왼쪽 구성요소 표기

항목	왼쪽 구성요소 표기
주소	http://data.juso.go.kr/address
도로명주소	http://data.juso.go.kr/address/address-of-building
사물주소	http://data.juso.go.kr/address/address-of-thing
공간주소	http://data.juso.go.kr/address/address-of-space

6.2 오른쪽 구성요소의 표기

오른쪽 구성요소는 자원의 유형에 따라 서로 다른 표기 패턴을 갖는다.

- ‘id’ 또는 ‘doc’ 등 참조항목: [/{type}][/{concept}/{key}]*[/{concept}][#id]
- ‘def’와 같은 어휘: [/{type}]{/vocabulary*}[/{term}][#{term}]
- ‘data’인 데이터세트: [/{type}]{/dataset*}[/{concept}/{key}]*[/{prop}]

{type}은 URI 패턴에서 공통 {concept}을 공유하는 참조 항목, 참조 데이터, 어휘를 구별하는 데 사용되는 선택적 키워드다. 한편, 사람이 URI가 참조할 수 있는 자원을 추측할 수 있는 역할을 한다. {type} 필드에 사용되는 일반적인 토큰 값은 다음과 같다:

- def: 어휘와 용어
- id: URI 세트와 참조 항목
- doc: 참조 또는 일반문서
- data: 데이터세트와 데이터 항목
- so: 공간 객체

{concept}은 어휘나 URI 집합과 관련된 기본 개념 또는 데이터세트의 이름(예시: 행정구역, 주소, 도로명)을 표기하며, 사람이 읽을 수 있는 형식을 권장한다. 사람이 URI 경로 세그먼트에 사용된 값을 직관적으로 이해할 수 있는 방식이 선호되지만, 엄밀한 의미에서 URI는 사람(및 기계)이 철자만 보고 특정 URI의 의미를 정확하게 식별하는 것을 의미하지 않는다. 그럼에도 불구하고 개발자와 데이터 최종 사용자에게 직관적으로 URI 패턴을 제공하는 것이 유용하다.

{key}는 URI 세트, 어휘 또는 데이터세트에서 한 항목을 다른 항목과 구별하는 키워드다. 일반적인 {key} 값은 게시되는 데이터 내의 기본 키 또는 코딩된 값을 직접 이용한다.

{/vocabulary*}는 어휘(스키마, 코드 목록, 개념 체계 또는 온톨로지)를 기술하기 위한 목적으로 사용한다.

주소와 관련된 오른쪽 구성요소의 표기는 다음과 같다. {prefix}는 주소의 왼쪽 구성요소를 요약하여 표현한 것으로, {prefix} 다음에 오른쪽 구성요소를 표기한다.

<표 6-2> 주소의 오른쪽 구성요소 표기

항목	오른쪽 구성요소 표기
도로명주소와 관련된 문서의 URI	{prefix}/doc/address/road-name-address
주소 온톨로지 어휘의 도로명주소 Class URI	{prefix}/def/address/RoadNameAddress
사물주소 데이터세트의 URI	{prefix}/data/address/address-of-thing

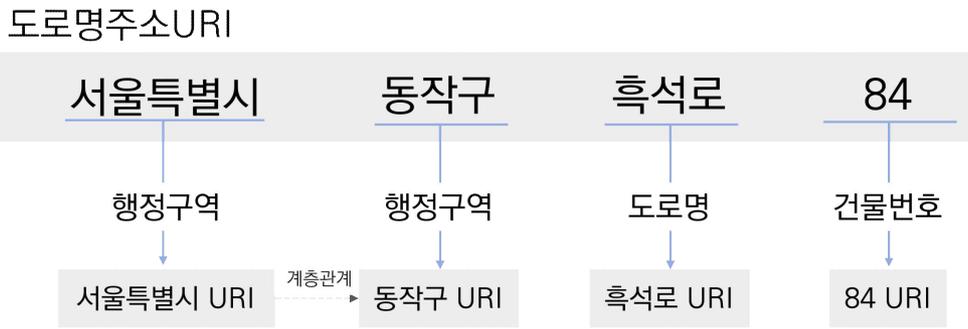
7 URI 설계 패턴

7.1 공통사항

주소그래프는 주소정보를 링크드 데이터 원칙을 적용하여 발행한 데이터세트다. 따라서, 주소그래프는 특정한 도메인({domain})을 통해 발행된다. 해당 도메인에 게시된 모든 데이터세트는 자원을 기술하는 메타데이터를 포함하고, RDF로 기술되며, 역참조 가능한 HTTP 범용 자원 식별자(URI)를 통해 접근할 수 있어야 한다. 월드와이드웹의 구성요소인 HTTP URI는 사물 또는 자원을 고유하게 식별하는 수단을 제공한다. 주소그래프는 주소정보 전반에서 공통의 의미와 공통 식별자를 공유하고, 도로명주소, 사물주소, 공간주소에 포함된 다양한 개체에 대해 포괄적이고 신뢰할 수 있는 식별자를 제공한다. 주소그래프의 개별 데이터세트는 다음의 구성요소를 포함한다:

- 데이터세트를 식별하는 URI
- 품질 특성을 기술하는 메타데이터
- 데이터세트에 정의된 자원의 목록(식별자 URI와 문서 URI)을 참조하는 URI
- 데이터세트 내에서 사용되는 개념과 관계를 정의하는 온톨로지 URI에 대한 참조

일반적으로 주소는 계층적인 구조를 갖는다. 예를 들어, 도로명주소인 ‘서울특별시 동작구 흑석로 84’는 행정구역, 도로명을 조합한 구조이고, 행정구역은 특별시와 자치구의 계층 구조를 갖고 있다.



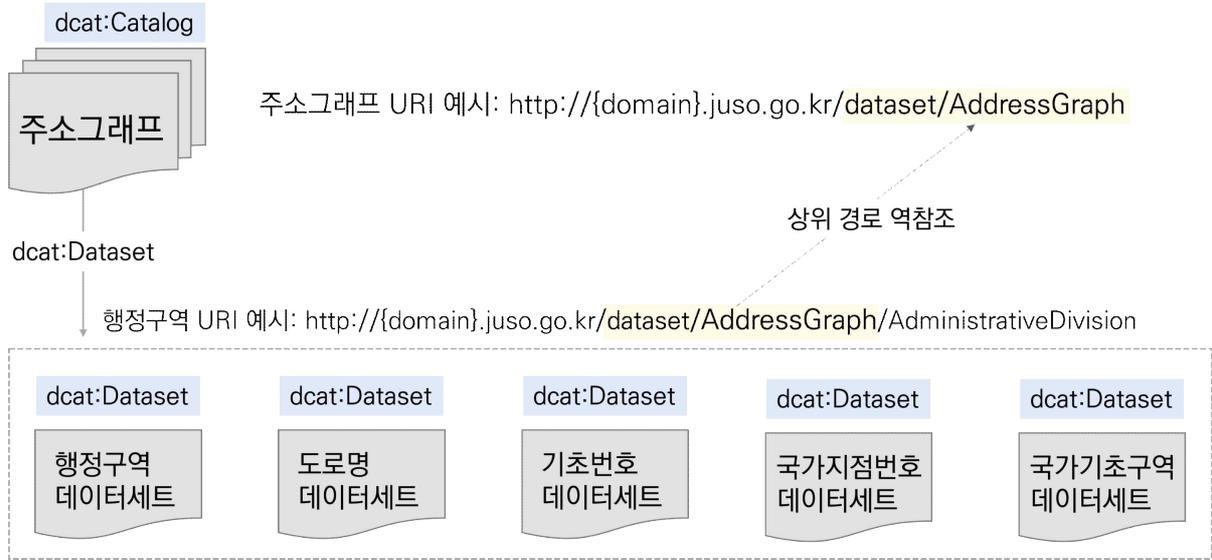
(그림 7-1) 도로명주소의 구조

이 계층 구조를 모델링하기 위해 URI에서 경로 세그먼트를 사용할 것을 권장한다. 구체적인 사항은 <표 7-1>을 참고한다.

<표 7-1> URI의 경로 세그먼트 구성을 위한 원칙

웹 URI 설계 패턴 목록	조건
데이터세트 URI를 확인할 수 있도록 HTTP URI를 사용해야 한다.	필수
데이터세트 URI는 기계 판독 가능한 RDF 표현을 하나 이상 제공해야 한다.	필수
데이터세트 URI는 'dataset' 문자열과 데이터세트의 특성을 설명하는 적절한 식별자 '{datasetid}'가 포함해야 한다. 예시: /dataset/{datasetid}	필수
선택적으로, 모듈화된 데이터세트 URI는 '{module}'로 명명한 식별자로 임의의 개수의 경로 세그먼트로 구성해야 한다. 예시: '/dataset[/]{module}*/{datasetid}'	필수

주소그래프는 주소를 구성하는 다양한 구성요소를 조합한 그래프 구조의 데이터이며, 구성요소에 대한 데이터와 주소그래프는 모두 데이터세트의 유형이다. 그러므로 본 표준에서는 주소그래프와 관련된 데이터세트의 메타데이터는 DCAT 버전 2를 사용하여 표현한다. (그림 7-2)와 같이 데이터세트 URI는 dcat:Dataset 클래스의 원소(member)이다. 한편, 주소그래프를 구성하는 데이터세트는 dcat:Catalog 클래스의 원소로 정의한다. 모든 데이터세트는 URI로 기술하고, dcat:Catalog 클래스의 dcat:dataset 속성으로 연결한다. 이 카탈로그 URI는 데이터세트의 최상위 경로 세그먼트에서 역참조해야 한다. 예를들어, 주소그래프의 dcat:Catalog URI가 '/dataset/AddressGraph'이고, 행정구역 데이터세트 '/dataset/AddressGraph/AdministrativeDivision'는 dcat:dataset 속성으로 참조한다.



(그림 7-2) 주소그래프의 URI 구조

<표 7-2>에서 상세히 언급하듯이, ‘dataset’ 문자열이 포함된 ‘dataset root’ URI는 도메인의 모든 데이터세트 목록을 참조해야 한다.

<표 7-2> 주소 데이터세트 URI를 구성하기 위한 원칙

웹 URI 설계 패턴 목록	조건
데이터세트 URI는 데이터 카탈로그 어휘(DCAT)의 ‘dcat:Dataset’ 클래스의 멤버로 정의하는 것을 권장한다.	추천
모듈화된 데이터세트의 최상위 모듈은 모든 데이터세트가 ‘dcat:dataset’ 속성을 통해 참조되는 ‘dcat:Catalog’ 클래스의 멤버로 선언하는 것을 권장한다.	추천
모든 데이터세트는 더블린 코어의 ‘dct:publisher’ 속성을 통해 정의된 하나 또는 여러 개의 게시자 선언을 권장한다.	추천
데이터세트는 더블린 코어의 dct:license 속성으로 정의하는 것을 권장한다.	추천
데이터세트의 ROOT URI는 해당 {domain}에 있는 모든 데이터세트 URI의 목록 표시를 권장한다. 예시: <code>http://{domain}.juso.go.kr/dataset</code>	추천

또한, 링크드 데이터의 게시와 URI의 표기에 원칙은 다음의 공통 가이드라인을 따른다.

<표 7-3> 링크드 데이터의 게시를 위한 URI 표기 원칙

웹 URI 설계 패턴 목록	조건
데이터세트 URI는 사람이 읽을 수 있는 HTML 표현을 제공해야 한다.	필수
여러 표현이 존재하는 경우, 사용가능한 각 표현에 대해 특정 URI를 검색할 수 있는 수단을 제공하는 것을 권장한다.	추천
데이터세트의 검사 또는 사용에 대한 라이선스는 공통 어휘를 사용하여 기술하는 것을 권장한다.	추천
데이터세트의 메타데이터는 공통 어휘를 사용하여 제공하고, 데이터세트 URI에 대한 관리계획을 포함하는 것을 권장한다.	추천
데이터 게시 시스템은 데이터세트의 URI를 표시하거나 동작에 영향을 미쳐야 한다.	추천

7.2 도메인 구조

데이터세트의 게시자는 데이터세트에 포함된 개체에 적합하다고 생각되는 하위 도메인을 선택할 수 있다. 데이터세트의 하위 도메인 이름은 기존의 모범 사례를 기반으로 주소그래프의 데이터세트에 대한 요구사항을 충족하는 다음 <표 7-4>의 원칙을 따른다.

<표 7-4> 주소 데이터세트의 도메인 구조 표기 원칙

웹 URI 설계 패턴 목록	조건
기본 도메인을 선언하고, 모든 데이터세트는 해당 도메인을 사용해야 한다.	필수
정부의 업무와 기능을 구분하는 키워드를 도메인에 포함한다 (예: administration, address 등)	권장
현재 데이터세트를 담당하는 부서 또는 기관의 이름은 URI에 사용하지 않는다.	필수
하위 도메인은 해당 부서/기관 서버로 리디렉션하는 방식으로 구현되어야 한다.	필수
하위 도메인은 영구적으로 유지하는 것을 권장한다.	추천

7.3 유형별 URI 설계

URI는 정보 자원(Information Resource) 또는 비정보 자원(Non-Information Resource)을 나타내는 데 사용할 수 있다(참고 <표 7-5> ~ <표 7-7>). 예를 들어 이름, 주소 등을 포함하여 사람을 설명하는 문서는 정보 자원이다. 반면, 현실 세계의 실제 사람은 비정보 자원이다. 일반적으로, 웹 환경에서 정보자원의 식별은 서버와 클라이언트의 일련의 통신으로 이루어진다. 서버는 정보 자원(문서, 이미지, 데이터세트 등)의 현재 상태를 나타내는 표현을 반환하고 HTTP 응답 코드 '200 OK'와 함께 클라이언트에 다시 보낸다. 비정보 자원은 직접 역참조할 수 없으므로 서버는 원본(비정보) 자원을 설명하는 정보 자

원에 대한 다른 URI를 클라이언트를 가리키고 HTTP 응답코드는 '303 기타 참조'로 설정하여 응답한다.

<표 7-5> 자원의 유형에 따른 URI

자원의 유형	URI 유형	내용
실세계의 사물 (비정보자원)	Identifier URI	서술문(statement)에서 기술된 현실 세계의 물리적 또는 사물의 식별 - 물리적 사물의 예시: 사람, 건물, 도로 등 - 추상적인 사물의 예시: 정부부처 등
웹 문서 (정보자원)	Document URI	서술문에서 참조할 수 있는 문서의 식별. 문서 URI는 웹의 문서로 직접 확인할 수 있음
실세계 사물의 특성의 정의(정보자원)	Ontology URI	실제 사물의 특성을 정의하는 온톨로지에 포함된 개념과 관계에 대한 정의

7.3.1 해시 URI

해시 URI는 조각 식별자로 해시 기호("#")를 사용하여 나머지 URI를 구분한다. 해시 URI는 일반적으로 다른 처리 응답이 필요한 기본 자원에 종속된 보조 자원을 식별한다. 예를 들어 웹 페이지 또는 텍스트 문서에서 조각 식별자는 페이지 또는 문서 내의 위치를 나타내고, 브라우저는 조각 식별자가 선택되면 문서의 해당 위치로 이동한다. RDF에서 조각 식별자는 보조 자원을 가리키는 데 사용할 수 있다. 해시 URI가 역참조될 때 HTTP 프로토콜은 서버에 URI를 전달하기 전에 조각 부분을 제거해야 한다. 이는 서버가 해시 URI의 나머지 부분을 직접 해석할 수 없음을 의미한다. 따라서 RDF 문서에 해시가 있으면 이 하위 URI가 브라우저에서 직접 제공되지 않으므로 동일한 문서 내에서 다른 비정보 자원을 나타내는 데 사용할 수 있다.

7.3.2 슬래시 URI

URI 경로 구조가 여러 가지 자원 유형을 제공하기 위한 두 번째 솔루션은 문서 URI와 식별자 URI 모두에 '슬래시 URI'를 사용하지만 요청된 식별자 URI가 일반 웹 문서가 아닌 비정보 자원임을 표시하기 위해 HTTP 상태 코드의 '303 See others'를 사용하여 리다이렉션 URI를 반환한다. 이때 해당 자원(문서 정보)의 표현을 위해서 URI에는 'doc'로 변경되어 클라이언트의 요청에 응답한다.

<표 7-6> 유형별 URI 설계 원칙

웹 URI 설계 패턴 목록	조건
슬래시 URI 패턴: 식별자 URI는 토큰 'id', {type}에 대한 참조와 '사물'의 이름 {name}을 포함하는 것을 권장한다. 예시: /id/{type}/{name} → /id/school/2060	추천
해시 URI 패턴: 식별자 URI는 토큰 'resource' 뒤에 데이터세트에 사용된 것과 동일해야 하는 식별자가 포함하는 것을 권장한다. 예시: /resource/{datasetid}#{name} → /resource/schools#2060	추천
슬래시 URI 패턴: URI 패턴은 토큰 'doc', '{type}'에 대한 참조와 '사물'의 이름 {name}이 포함하는 것을 권장한다. 예시: doc/{type}/{name} → doc/school/2060	추천
해시 URI 패턴: 문서 URI는 토큰 'resource' 뒤에 데이터 집합에 사용된 식별자('{datasetid}')를 포함하는 것을 권장한다. 예시: /resource/{datasetid}#{name} → /resource/schools#2060	추천
온톨로지 URI는 간결성을 위해 해시 URI 패턴을 사용하는 것을 권장한다.	추천
개념과 관계의 정의는 'def' 키워드 다음에 온톨로지 이름 '{scheme}', 이후 개념 또는 관계 이름을 '{concept}'으로 표시하는 것을 권장한다. 예시: /def/{scheme}#{concept} → /def/school#phaseOfEducation	추천
클래스의 인스턴스, 즉 실제 개체(비정보 자원)가 온톨로지의 일부로 모델링되는 경우 (예시: 코드 목록, 유한한 개체 집합), 개체에 사용되는 URI는 온톨로지 URI 패턴이 아닌 식별자 URI 패턴을 따를 것을 권장한다.	추천

<표 7-7> 특정 주소(비정보 자원)를 식별하는 슬래시 URI 예시

클라이언트 요청 URI 예시	서버 응답 예시(상태코드와 URI)	
http://{domain}/id/address/2060	상태코드	'303 다른 참조',
	리다이렉션 URI	http://{domain}/doc/address/2060

7.3.3 URI 모듈화

자원의 집합은 데이터세트 계층 구조를 나타내기 위해 임의의 수의 경로 세그먼트를 포함하는 URI로 표시되는 모듈로 그룹화할 수 있다. 각 URI 유형에 대해, 즉 식별자 URI, 문서 URI, 온톨로지 URI의 경우(선택 사항) [{모듈}/]* 단계는 특정 자원이 속한 모듈을 나타낼 수 있다. 데이터세트 모듈의 일부인 경우 이 데이터세트 내의 모든 자원은 반드시 동일한 경로 세그먼트를 사용해야 한다. 예를 들어, 상세주소에 대한 범정부 식별자가 없는 경우 지방자치단체는 모듈 내에 상세주소에 대한 데이터세트를 추가할 수 있다. 다음 예시는 'address'로 모듈을 표현하고, 도로명주소와 지번주소의 데이터를 하위 데이터 집합으로 그룹화하고 있다.

<표 7-8> 특정 주소(비정보 자원)를 식별하는 슬래시 URI 예시

구분	클라이언트 요청 URI 예시(호주 사례)	URI 모듈화 설명(호주 사례)
1	/dataset/address/roadNameAddress	- 'schools'는 우리나라의 주소의 데이터 집합 - 'roadNameAddress'는 모든 도로명주소를 포함하는 'address'의 하위 데이터 집합 - 'landLotNumberAddress'는 모든 지번주소를 포함하는 'address'의 하위 데이터 집합
2	/dataset/address/landLotNumberAddress	

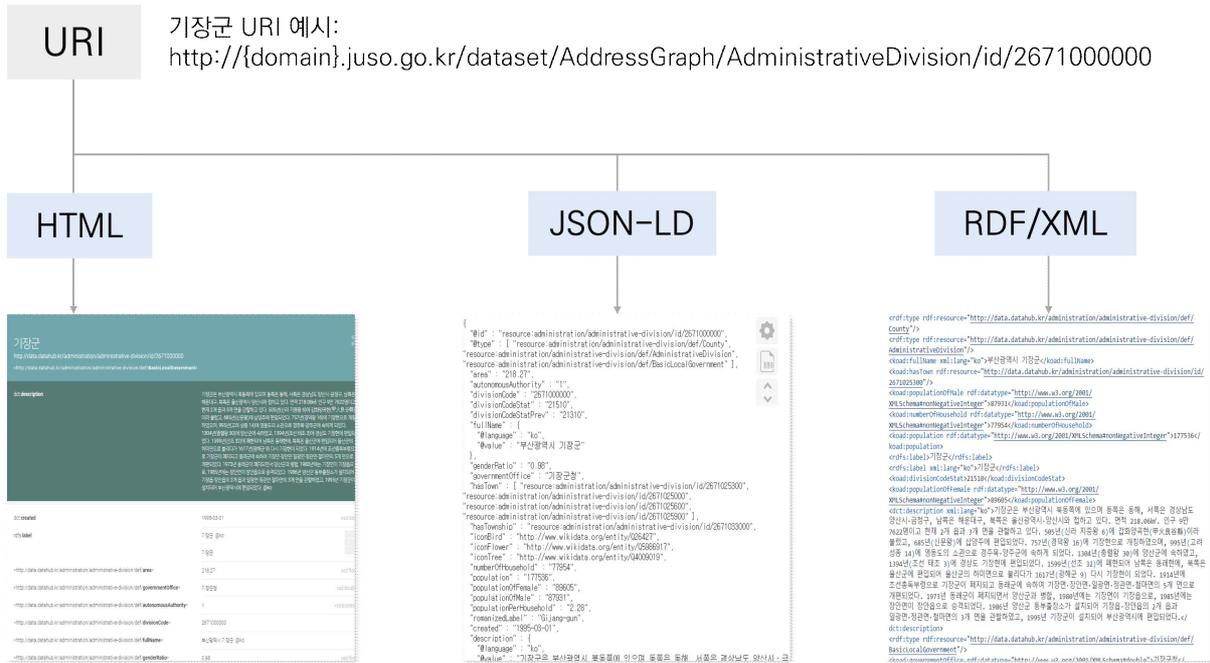
개별 자원은 둘 이상의 모듈에 속할 수 있으며, 각각 다른 컨텍스트와 관련된 둘 이상의 URI로 식별될 수 있다.

<표 7-9> 유형별 URI 설계 패턴

웹 URI 설계 패턴 목록	조건
데이터셋이 모듈의 일부인 경우, 데이터셋 내의 모든 자원은 반드시 동일한 경로 세그먼트를 사용해야 한다.	필수
슬래시 URI 패턴 예시: /[{module}]/*id/{type}/{name} → /act/id/school/2060	추천
해시 URI 패턴 예시: /[{module}]/*/resource/{datasetid}#{name} → /act/resource/schools#2060	추천
슬래시 URI 패턴 예시: /[{module}]/*/doc/{type}/{name} → /act/doc/school/2060	추천
해시 URI 패턴 예시: /[{module}]/*/resource/{datasetid} → /act/resource/schools	추천
예시: /[{module}]/*/def/{scheme}#{concept} → 예시: /act/def/school#phaseOfEducation	추천

7.4 URI 확인

내용 협상(content negotiation)은 동일한 URI에서 서로 다른 버전의 자원을 표현을 제공하는 HTTP 메커니즘이다. 클라이언트 애플리케이션마다 데이터 형식과 언어에 대한 기본 설정이 다르며, 이는 요청의 HTTP 헤더에 표시될 수 있다. 예를 들어, 브라우저는 일반적으로 자연어로 현지화된 HTML을 요청하는 반면, 링크드 데이터 애플리케이션은 일반적으로 RDF를 요청한다. 이때, 서버는 (그림 7-3)처럼 파일 시스템에서 또는 필요에 따라 원하는 콘텐츠를 생성하여 일치하는 형식을 선택하여 클라이언트로 보낸다. HTTP URI를 역참조할 때 다양한 리소스 자원 유형을 구분하는 표준 패턴이 있다.



(그림 7-3) 내용 협상의 예시

<표 7-10> 내용 협상을 위한 URI 설계 원칙

웹 URI 설계 패턴 목록	조건
슬래시 URI의 경우, /id/{type}/{id}에 대한 요청에 대해 '303 기타 참조' 상태 코드가 발행되어야 한다. 이에 대한 응답은 doc/{type}/{id} 으로 리디렉션되는 것을 권장한다.	추천
해시 URI 패턴('/resource/{datasetid}#{id}')의 저장 위치는 /resource/{datasetid}로 하는 것을 권장한다.	추천
자원의 RDF와 HTML 표현이 정보 내용 측면에서 다르지 않은 경우, 서로 다른 표현을 구분하기 위해 파일 확장자(예시: .html, .rdf, .owl)를 사용하는 것을 권장한다.	추천
파일 확장자가 다른 표현을 구분하기 위해 사용되는 경우, 해당 유형은 URI의 다른 부분에 명시적으로 명시한다. 예시: /doc/rdf/school/2060.rdf가 아닌 /doc/school/2060.rdf	추천
자원의 RDF와 HTML 표현이 상당히 다른 경우(즉, 동일한 문서의 두 버전이 아니라 완전히 다른 문서인 경우) 콘텐츠 협상과 함께 "303 다른 문서 참조" 리디렉션을 설정하여 두 개의 개별 문서 URI를 지칭하는 것을 권장한다.	추천
문서의 다른 버전을 나타내려면 문서 URI에 '날짜' 토큰을 사용하여 정보가 특정 '날짜' 또는 '날짜부터 유효함'을 나타내는 것을 권장한다.	추천

온톨로지 URI는 문서 유형이 항상 RDF 또는 OWL인 특수한 종류의 문서 URI다. 따라서 스키마 수준의 정의(즉, 클래스/프로퍼티)만 포함하는 온톨로지의 경우 내용 협상이나 리디렉션이 필요하지 않다. 예를 들어, /def/{scheme}#{concept} 패턴을 가진 해시 URI 온톨로지의 클래스와 속성에서 해시는 자동으로 제거되어 문서 URI인 /def/{scheme}가 된다. 인스턴스가 클래스와 속성이 동일한 경우, 인스턴스의 URI 체계는 식별자 URI 패턴을 따른다.

<표 7-11> 인스턴스 URI 표현을 위한 설계 원칙

웹 URI 설계 패턴 목록	조건
스키마 인스턴스 수준 정보를 모두 포함하는 해시 URI 스키마를 사용하는 경우, 온톨로지에 정의된 문서 URI와 식별자 URI를 올바르게 확인할 수 있도록 파일은 /def/{scheme}와 /resource/{type}에 저장하는 것을 권장한다.	추천

8 사람과 기계가 읽을 수 있는 형식

주소정보의 URI가 확인되면, 해당 URI에 포함된 정보를 다양한 형식으로 제공할 수 있다. 반환되는 문서의 개별 정보의 값은 가능하다면 고유한 URI가 있어야 한다. 예를 들면 다음과 같다.

- RDF/XML: <http://data.datahub.kr/doc/road/M5/junction/24/doc.rdf>
- Turtle: <http://data.datahub.kr/doc/road/M5/junction/24/doc.ttl>
- HTML: <http://data.datahub.kr/doc/road/M5/junction/24/doc.html>
- JSON: <http://data.datahub.kr/doc/road/M5/junction/24/doc.json>
- JSON-LD: <http://data.datahub.kr/doc/road/M5/junction/24/doc.json>

개별 항목에 부여된 고유한 표현 URI를 사용하면 요청과 함께 전송된 헤더를 변경하지 않고도 특정 URI에 액세스할 수 있다. 이를 위해서 사용하는 데이터 형식에 따라 관련된 자원을 함께 제공해야 한다. 자원 목록은 다음과 같다.

- HTML: 적절한 유형 속성이 있는 <링크> 요소를 사용한다.
- RDF: dct:hasFormat 속성을 사용한다.
- Atom: 적절한 유형 속성과 함께 요소를 사용한다.

또한, 적어도 하나의 표현은 기계 판독을 위해서 RDF 그래프를 구성할 수 있는 방식으로 표현해야 한다. RDF 표현은 다음 중 하나를 사용하는 것을 권장한다.

- RDF/XML(가장 많은 도구에서 지원되므로 선호됨)
- GRDDL4 변환
- RDFa5가 내장된 XHTML

RDF 그래프 구성을 위해 다음 형식을 제공하는 것도 허용된다.

- Turtle
- N3

기계가 읽을 수 있는 다른 형식을 제공하는 것을 권장한다.

- JSON
- CSV

사람이 읽을 수 있는 버전의 문서를 HTML로 제공하는 것을 권장한다.

부 록 I

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

웹 URI

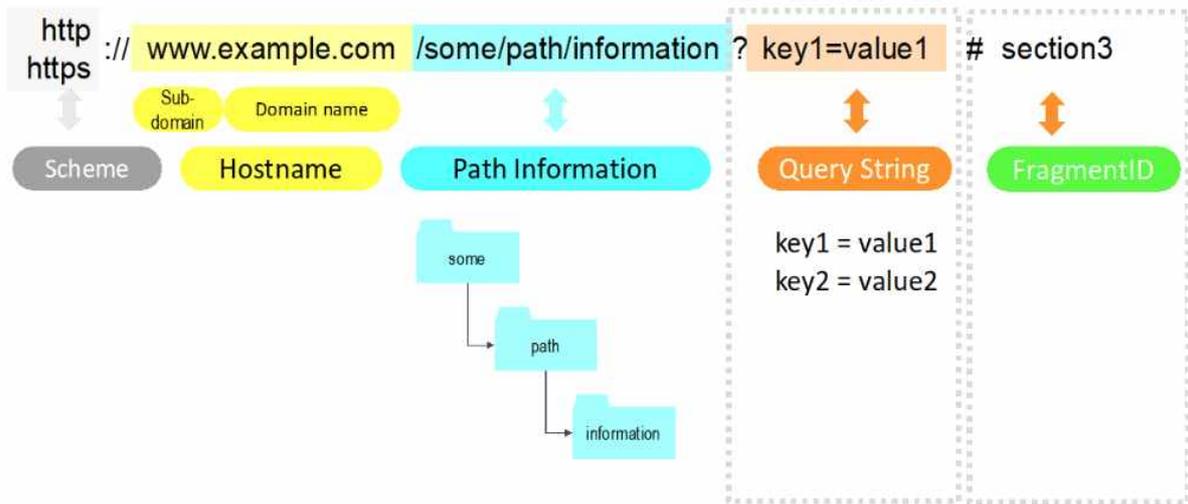
1.1 구성요소

URI(Uniform Resource Identifier)는 월드와이드웹에서 사용되는 단일 글로벌 식별 시스템이다. URI는 전 세계의 엔티티('사물') 또는 개념을 식별하는 일반적인 메커니즘을 제공하여 연결된 데이터를 지원하는 핵심 기술이다. 정부 부서와 기관은 병원, 학교, 도로, 장비 등 담당하고 있는 모든 개체('사물')에 식별자를 할당한다. 이러한 식별자는 특정 사물에 대해 언급하거나 진술할 때 사용된다. 웹 공간에서 식별자(Identifier)는 다양한 형식을 사용한다. 먼저, URI는 Uniform Resource Identifier(통합 자원 식별자)의 약자이고, URL은 Uniform Resource Locator(공통 자원 위치탐색기)의 약자로서 웹페이지에서 정보 탐색 시 사용된다. URN은 Uniform Resource Name(공통 자원 이름)의 약자로 국제적으로 고유하게 사물을 명명하는 데 사용된다. URN과 URL은 HTTP 또는 HTTPS, FTP, TELNET 등 다양한 프로토콜을 사용하는 URI를 포함하는 광범위한 용어다. 즉, URI가 더 포괄적인 용어이므로 모든 URL과 모든 URN도 URI이다. 또한 IRI(국제화된 자원 식별자)는 훨씬 더 광범위한 범주로 범용 문자 세트 및 유니코드의 문자를 지원하나 URI는 ASCII 문자 세트만 지원한다.

식별자의 기능은 크게 두 가지로 구분하며, 첫 번째 기능은 일반적으로 URL, 웹 주소와 연관되고, 두 번째 기능은 URN과 연관된다.

- 웹 공간의 자원을 쉽게 식별하고 접근하는 기능
- 사물이 웹 또는 현실 세계에 존재하는 상황에 상관없이 전 세계적으로 모호하지 않은 이름을 제공할 수 있는 기능

또한, 웹 URI는 이 두 가지 기능을 목적에 맞게 사용할 수 있으며, 구문상으로 HTTP 또는 HTTPS를 프로토콜로 지정하기 때문에 URL과 동일한 형식일 수 있다. 한편, HTTP/HTTPS 웹 프로토콜을 통해 웹 요청을 지원한다는 측면에서 URL처럼 동작하도록 구성할 수 있다. 그러나 현실 세계나 온라인에서 무엇이든 전 세계적으로 모호하지 않은 이름을 지정하는 데 완벽하게 유효한 방법이다. '전 세계적으로 모호하지 않다'는 것은 전 세계적으로 고유하다는 뜻이 아니며, 서로 다른 두 사물을 구분할 수 있으려면 어떤 상황에서든 두 사물을 구분할 수 있는 고유한 URI를 갖고 있어야 한다. 그렇지만 동일한 URI 네임스페이스 또는 도메인 이름 내에서도 모두 동일한 것을 가리키는 URI가 다수 존재할 수 있다. 따라서, 링크된 데이터 원칙을 적용하여 두 URI가 서로 다른 문자열이더라도 둘이 같은 것을 참조한다는 것을 공식적으로 표현하기 위해 두 URI 간에 관계를 만들 수 있다.



(그림 1.1-1) 웹 URI 체계 개요도

(그림 1.1-1은) 웹 URI 체계에 대한 개요도를 나타낸다. 웹 URI 체계는 하나의 프로토콜을 의미하며, ‘http://’ 또는 ‘https://’ 사용을 권장한다. ‘호스트 이름(Hostname)’은 일반적으로 등록된 인터넷 도메인 이름 또는 등록된 도메인 이름의 하위 도메인이다. 도메인 이름 다음에 오는 웹 URI의 나머지 부분은 대소문자를 구분하여 표현한다. URI 경로 정보는 슬래시(‘/’) 문자로 구분된 다수의 문자열로 구성된다. 웹 URI는 문자열로 표현되지만 링크드 데이터 커뮤니티와 REST 인터페이스에서 개념적이고 계층적인 방식으로 구성된 자원의 집합을 나타내기 위해 사용된다.

일반적으로 웹 URI의 구성요소는 가장 광범위한(가장 일반적이고 구체적이지 않은) 카테고리 URI 경로 정보의 왼쪽에 나타나고, 가장 좁은(가장 구체적이지 않은) 카테고리가 URI 경로 정보의 오른쪽에 나타난다.

이러한 웹 URI는 디자인 패턴은 기계뿐만 아니라 사람들도 이해 할 수 있는 패턴이므로, 웹 URI 경로 정보를 오른쪽에서 왼쪽으로 연속적으로 잘라내고 슬래시(‘/’)문자 앞쪽에 위치한 연속된 세그먼트를 제거함으로써 해당 웹 URI에 포함된 정보를 사람들도 이해 가능하다. 그러므로 이러한 웹 URI는 개체에 대한 정보를 보다 광범위하게, 그리고 일반적이며 구체적인 세부 단위로 구성할 수 있다.

1.2 링크드 데이터를 위한 URI

정부 또는 정부 기관에서 링크드 데이터 방식으로 데이터를 웹에 게시하는 경우 웹 URI 기반의 자원 식별자의 적용을 권고한다. 예를 들어, 영국 정부는 공공부문정보(PSI)의 다양한 애플리케이션에서 재사용하기 위해 자원 식별자 URI에 대한 정부 개방표준을 채택하고, 공공부문의 자원을 식별하기 위해 링크드 데이터 원칙을 적용하고 있다(Chief

Technology Officer Council, 2009). 이 문서는 정부 이해관계자가 ‘링크드 데이터 데이터세트’에 설명된 자원에 대한 URI를 정의하고 관리하는 데 도움을 주기 위한 일련의 일반 지침을 제공한다.

1.2.1 HTTP URI 사용

네 가지 원칙(Heath & Bizer, 2011) 중 두 가지 원칙인 ‘URI 사용’과 ‘HTTP URI 사용’은 자원을 웹에 게시하고 재사용함으로써 궁극적으로 데이터 통합·연결을 지원하기 위한 방법이다. HTTP URI를 사용하면 URI를 ‘조회’ 또는 ‘역참조’할 수 있으며, 웹 브라우저를 통해 웹 URI로 식별되는 자원의 표현에 접근할 수 있다.

1.2.2 URI로 식별되는 자원의 기계 판독 가능한 표현 제공

HTTP URI를 ‘역참조 가능’하게 하려면 데이터 게시자 또는 발행자는 자원의 표현 또는 설명(예시: 사람이 읽을 수 있는 HTML 표현 또는 기계가 읽을 수 있는 RDF/XML 표현)을 제공을 위한 인프라(예시: HTTP 서버)를 설정해야 한다. 링크드 데이터 원칙이 적용된 웹 자원은 반드시 RDF를 사용하여 데이터를 게시해야 한다. 즉, RDF를 사용하여 데이터의 의미를 명시적으로 정의해야 하고, 자원을 식별하는 HTTP URI를 통해 기계가 읽을 수 있는 표현을 제공해야 한다(예시: RDF/XML, JSON-LD, Turtle).

한편, 컴퓨터 소프트웨어는 전체 URI를 하나의 문자열로 취급하기 때문에 판단하기 때문에 문자열을 각 구성요소별로 분할하지 않는다. 반면, 링크드 데이터 원칙에서는 온톨로지 제공되는 항목에 표현된 URI를 이용하여 자원에 대한 명시적 접근이 가능케 한다.

부 록 II

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

FAIR 데이터 원칙

II.1 개요

FAIR 데이터 원칙은 데이터의 공유와 재사용성을 강화하기 위해 고안된 가이드 원칙(Wilkinson et al, 2016)이다. FAIR 데이터 원칙은 데이터를 찾을 수 있고(Findable), 데이터에 접근 가능하며(Accessible), 데이터의 상호운용이 가능하고(Interoperable), 데이터를 재사용할 수 있는 원칙을 강조한다. FAIR 데이터 원칙은 오픈 사이언스(Open Science)의 맥락에서 연구 데이터의 재사용성을 증진시키기 위한 도구로 인식되었지만, 현재는 데이터를 보존하고 관리하기 위한 보편적인 프레임워크 역할을 하고 있다(김학래, 2021).

II.2 FAIR 데이터 원칙의 세부 지표

FAIR 데이터 원칙은 탐색성(Findability), 접근성(Accessibility), 상호운용성(Interoperability), 재사용성(Reusability) 요소로 구분된다. 개별 요소는 요소를 정량적으로 평가할 수 있는 세부 지표로 구성된다. 세부 지표는 <표 II.2-1>과 같다.

<표 II.2-1> FAIR 데이터 원칙의 세부 지표(GO FAIR, 2021)

FAIR 항목	조건
탐색성	F1 (메타)데이터는 전역적으로 고유하고 영구적인 식별자가 할당되어야 한다.
	F2 데이터는 풍부한 메타데이터로 기술되어야 한다.
	F3 메타데이터는 기술하는 데이터에 대한 명확하고 명시적인 식별자를 포함해야 한다.
	F4 (메타)데이터는 검색 가능한 자원에 등록 또는 색인화되어야 한다.
접근성	A1 (메타)데이터는 표준화된 통신 프로토콜을 사용하여 식별자로 검색된다.
	A1.1 프로토콜은 개방형이며, 무료이고 보편적인 방법으로 구현 가능해야 한다.
	A1.2 프로토콜은(필요한 경우) 인증과 권한 부여 절차를 허용해야 한다.
	A2 (메타)데이터는 더 이상 사용할 수 없는 경우에도, 계속 접근할 수 있어야 한다.

FAIR 항목	조건	
상호운용성	I1	(메타)데이터는 지식 표현을 위해 공식적이고 접근 가능하며 공유되고 광범위하게 적용 가능한 언어를 사용해야 한다.
	I2	(메타)데이터는 FAIR 원칙을 따르는 어휘를 사용해야 한다.
	I3	(메타)데이터는 다른 메타데이터에 대한 정규화된 참조가 포함되어야 한다.
재사용성	R1	(메타)데이터는 정확하고 관련성 높은 속성이 여러 개 있어야 한다.
	R1.1	(메타)데이터는 명확하고 접근 가능한 라이선스로 공개되어야 한다.
	R1.2	(메타)데이터는 상세화된 출처와 관련이 있다.
	R1.3	(메타)데이터는 도메인과 관련된 커뮤니티 표준을 충족해야 한다.

부 록 III-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 III-2

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

시험인증 관련 사항

해당 사항 없음.

부 록 III-3

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

본 표준의 연계(family) 표준

III-3.1 TTAK.KO-10.1496-Part1, 주소그래프 - 제1부: 주소 체계

III-3.2 TTAK.KO-10.1496-Part2, 주소그래프 - 제2부: 주소 어휘

부 록 III-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.

Chief Technology Officer Council, Designing URI Sets for the UK Public Sector, UK Government, 2009.10.

Tim Berners-Lee, Linked Data – design issues, 2006.07.27.

Heath, T., Bizer, C.(2011). Principles of Linked Data. In: Linked Data. Synthesis Lectures on Data, Semantics, and Knowledge. Springer, Cham. https://doi.org/10.1007/978-3-031-79432-2_2

Roy T. Fielding, Mark Nottingham, David Orchard, Joe Gregorio, Marc Hadley. RFC 6570, Internet Engineering Task Force, 2012.03.

Wilkinson et al., The FAIR guiding principles for scientific data management and stewardship, Scientific Data, 2016.

GO FAIR, FAIR Principles, 2021.

TTAE.OT-10.0427/R1, 데이터 카탈로그 어휘 버전 2, 2020.

김학래, FAIR 원칙: 데이터 관점의 디지털 아카이브 구현을 위한 고려사항, 한국기록관리학회지, 2021.

부 록 III-5

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

영문표준 해설서

해당 사항 없음.

부 록 III-6

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

표준의 이력

판수	채택일	표준번호	내용	담당 위원회
제1판	2023.12.06	제정 TTAK.KO-10.1496-Part3	-	빅데이터 프로젝트그룹 (PG1004)