

# 스마트시티 개방형 데이터허브 아키텍처

Technical Report [1부-1권]

스마트시티  
혁신성장동력 프로젝트

[1-1세부과제]  
주관연구기관-한국전자기술연구원

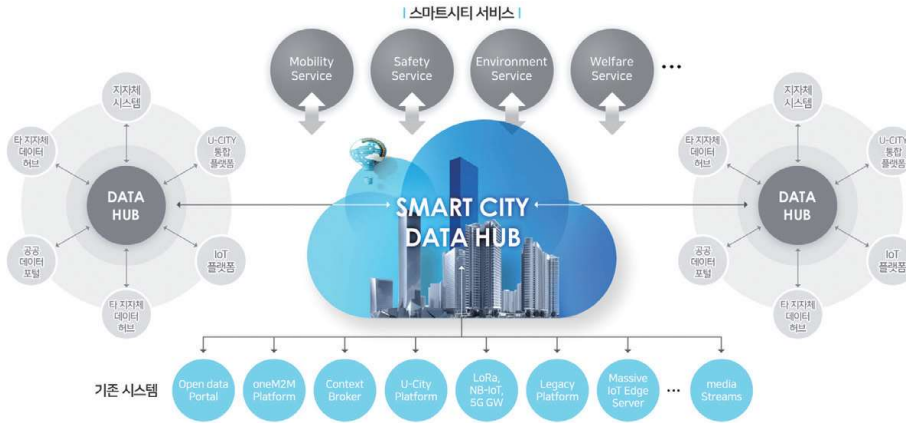
# 1부

## 데이터허브 / 아키텍처

## FACTSHEET

과제명	스마트시티 개방형 데이터허브 아키텍처 및 핵심 기술 개발	연구기간	'18.08 ~ '22.12 (4년 5개월)
		예산	총 109.4억원 (정출금 : 82.5억원)

### 개념도 (서비스 시나리오)



KPI (성과지표)	표준 API 적합성 테스트 통과율	100%	오픈소스의 데이터 허브 기능 제공율	100%	데이터허브 오픈소스 사용자 만족도	80%
------------	--------------------	------	---------------------	------	--------------------	-----

### 과제 개요

- (배경) 도시 내 산재된 데이터 시스템에 도시 데이터가 산재해있어 효율적인 데이터 관리와 활용이 어려워 데이터 중심 스마트시티 구현 어려움
- (목적) 도시 문제를 해결하고 시민의 삶의 질 향상을 위해 도시 데이터 통합 관리 플랫폼을 개발하여 융복합 서비스 개발 활성화 지원

### 주요 연구내용

- 스마트시티 데이터허브 아키텍처 개발
- 스마트시티 데이터허브 플랫폼 오픈소스 개발
- 스마트시티 데이터허브 적합성 시험검증
- 스마트시티 데이터허브 개념검증
- 스마트시티 데이터 거버넌스 연구 (법·제도 등)

### 기술적 차별성

- ETSI NGSI-LD 표준 인터페이스 지원
- Linked Data 기반 도시 공통 데이터 모델 지원
- 데이터 분석 및 데이터 마켓플레이스 지원
- 도시 데이터 시스템 연동을 위한 다양한 어댑터 지원
- 하이브리드 클라우드 인프라 지원
- 통합 인증/인가 및 블록체인 활용 서비스 지원
- 오픈소스 제공 및 웹기반 기술 지원 문서 제공

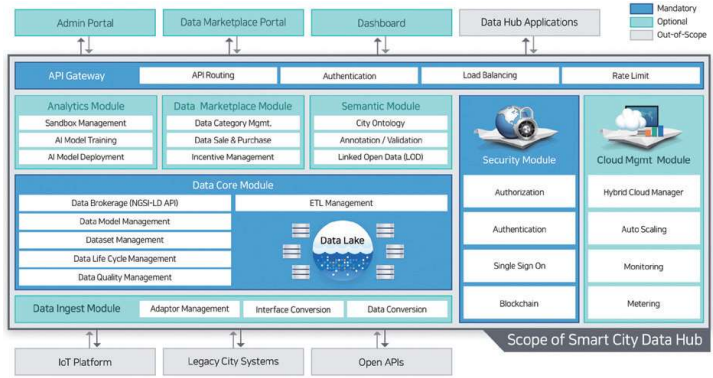
### 기대효과

- 상이한 정보시스템에 파편화되어 관리되는 도시 데이터를 통합 관리하여 융복합 분석 서비스 활성화
- 데이터허브 오픈소스 제공으로 다양한 활용 솔루션 개발 및 구축 확산으로 데이터허브 생태계 조성
- 표준 인터페이스 적용으로 해외 오픈소스 프로젝트와 협업 및 시장 진출 모색 가능

### 참여기관

[주관] [공동]

## 스마트시티 데이터허브 아키텍처



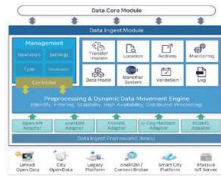
### 기능 모듈로 구성된 스마트시티 데이터허브

- 스마트시티 데이터허브의 주요 요구기능 도출을 통해 단위 기능 모듈을 설계·개발하여 통합된 플랫폼 제공
- 도시 데이터를 활용하는 모듈은 데이터 코어 모듈의 표준 API로 연동
- 지자체의 요구사항에 따라 필수 모듈과 선택 모듈 중에 선택하여 데이터허브 구축·운영 가능
- 향후 추가 기능 모듈을 개발 시, API 게이트웨이 및 보안 모듈과 연계하여 확장 구축 가능

## 데이터허브 기능 모듈

### 데이터 수집 모듈

IoT 플랫폼, Open API, 타 스마트시티 플랫폼 등 도시 인프라의 데이터를 데이터허브로 수집하기 위한 프로토콜 변환 어댑터 제공



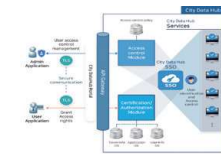
### 데이터 분석 모듈

데이터 전처리, 기계학습 모델 생성/검증 및 실행관리 기능 제공



### 보안 모듈 (인증/인가)

API G/W 연계한 토큰 기반 통합 인증/인가 제공



### 데이터 코어 모듈

데이터 모델, 데이터 셋, 데이터 흐름, 생애주기, 관리 및 데이터 저장/조회 기능 제공  
다양한 데이터 저장소 제공



### 데이터 마켓플레이스 모듈

데이터 상품 판매 및 구매 위한 마켓플레이스 포털 제공



### API 게이트웨이

API 라우팅, 외부 요청 수 제한, 보안통신, 토큰검증 등 기능제공



### 실증을 통한 시사점

- 주차 혼잡도 예측 서비스, 코로나19 역학조사지원시스템 등 데이터허브를 활용한 융복합 분석 서비스를 개발하여 대용량 데이터 수집과 분석 서비스 활용의 효율성 검증함
- 활용 확산을 위한 지속적인 기술 지원과 보다 많은 실증 적용을 통해 데이터허브의 사용성을 개선하고 도시 데이터 통합 운영과 융복합 서비스 개발 지원 위한 시스템 고도화 개발 지속 필요

### 타 지자체 확산방안

- 스마트시티 데이터허브 오픈소스(CityDataHub) 공개 및 기술지원 활동을 통해 오픈소스 솔루션 구축 및 오픈소스 기반 상용 솔루션 개발 추진
- 지자체에 기 구축된 U-City 통합플랫폼 및 빅데이터 플랫폼과의 연계 활용 방안 제시 및 개념검증
- 기존 실증한 오픈소스 기반 데이터허브(예: 시흥시, 부천시) 사례를 통해 구축 및 운영 가이드 지원



#### 연구책임자

- 한국전자기술연구원
- 성낙명 팀장
- nmsung@keti.re.kr



#### 집필자

- 한국전자기술연구원
- 정승명 선임
- sm.jeong@keti.re.kr

• 목차 •

제1장

개요

- 1. 배경 및 목적 ..... 21
- 2. 특징 및 기대효과 ..... 24

제2장

연구 개발 성과

- 1. 아키텍처 ..... 29
- 2. 모듈 정의 및 구성 ..... 32

제3장

데이터허브 검증

- 1. 개념검증(Proof of Concept) ..... 48
- 2. 시험검증 ..... 50

제4장

확산 방안

- 1. 도시 데이터허브 확산방안 ..... 52

제5장

Lesson Learned

- 1. 맺음말 ..... 56
- 2. 향후 과제 ..... 57

• 🔍 용어 정리 •

용어	정의
CityDataHub	스마트시티 혁신성장동력프로젝트 사업을 통해 개발된 오픈소스 기반의 스마트시티 데이터허브 <a href="http://citydatahub.kr">http://citydatahub.kr</a>
Context Broker	NGSI 및 NGSI-LD 인터페이스를 구현한 데이터 플랫폼
DMZ	내부와 외부 네트워크 중간에 위치한 네트워크 관리 영역으로 침입차단 시스템 등으로 접근 제한 등을 수행하지만 외부 네트워크에서는 접근이 가능한 영역(Demilitarized Zone)
ETL	데이터 분석을 위한 데이터 추출, 변형 및 적재 작업
FIWARE	유럽연합(EU) 지원 프로젝트로 시작한 IoT 및 스마트시티 오픈소스 제공 단체
JSON-LD	Linked Data를 지원하는 JSON 메시지 포맷 (JavaScript Object Notation for Linked Data)
Linked Data	서로 다른 데이터 간의 연결로 구성된 데이터 간의 웹
NGSI-LD	컨텍스트 관리 프레임워크와 모델링에 따른 데이터 인터페이스 표준 (Next Generation Service Interfaces-Linked data)
oneM2M	사물통신, IoT 기술을 위한 요구사항, 아키텍처, API 사양, 보안 솔루션, 상호 운용성 등 IoT 플랫폼을 지원하는 국제 단체표준 기술
RDBMS	MySQL, PostgreSQL, Oracle 등으로 알려진 관계형 데이터베이스
RDF	기계가 이해할 수 있는 시맨틱 기반 데이터 형식
REST API	통상적으로 HTTP로 구현하는 리소스 기반 API(Representational State Transfer API)
SPARQL	시맨틱 데이터베이스에서 지원하는 데이터 질의 언어
SQL	관계형 데이터베이스에서 지원하는 데이터 질의 언어
URI	웹주소(HTTP)와 같이 웹 자원을 식별하는 식별자

## · 그림 목차 ·

〈그림 1-1〉 스마트시티 데이터허브 기반 융복합 서비스 .....	22
〈그림 1-2〉 스마트시티 혁신성장동력 사업단의 데이터허브 개발 및 실증 .....	23
〈그림 1-3〉 도시 데이터 시스템 통합 및 효율적 데이터 활용을 위한 데이터허브 .....	25
〈그림 1-4〉 스마트시티의 링크드 데이터 개념도(출처: NEC Laboratories Europe)...	27
〈그림 2-1〉 스마트시티 데이터허브 참조구조 .....	31
〈그림 2-2〉 스마트시티 데이터허브 네트워크 구성도 예시 .....	32
〈그림 2-3〉 데이터 수집 모듈 개념도 .....	33
〈그림 2-4〉 데이터 수집 모듈의 어댑터 실행 관리 화면 .....	34
〈그림 2-5〉 데이터 코어 모듈 관리자의 데이터 모델 설정 화면 .....	36
〈그림 2-6〉 데이터 코어 모듈 개념도 .....	37
〈그림 2-7〉 분석 모듈 개념도 .....	38
〈그림 2-8〉 기계학습 모델 생성을 위한 파라미터 설정 .....	39
〈그림 2-9〉 데이터 마켓플레이스 모듈 개념도 .....	41
〈그림 2-10〉 데이터 코어 모듈 관리자의 데이터세트 설정 화면 .....	42
〈그림 2-11〉 시맨틱 모듈 개념도 .....	43
〈그림 2-12〉 인증·인가 모듈 개념도 .....	44
〈그림 2-13〉 블록체인 모듈 개념도 .....	45
〈그림 2-14〉 API 게이트웨이 개념도 .....	46
〈그림 2-15〉 클라우드 관리 모듈 개념도 .....	47
〈그림 3-1〉 CityDataHub를 활용한 주차 혼잡도 예측 PoC 개념도 .....	48
〈그림 3-2〉 CityDataHub 활용 코로나19 역학조사지원시스템 개념도 .....	50
〈그림 3-3〉 스마트시티 데이터허브 시험검증 절차 .....	50
〈그림 4-1〉 스마트시티 데이터허브와 통합 플랫폼 연계 활용 방안 .....	53
〈그림 4-2〉 CityDataHub 활용 확산방안 .....	54



## 1 | 배경 및 목적

## D 데이터 중심의 스마트시티 구현

- 도시 효율성을 높이고 시민의 삶의 질을 향상시키는 스마트시티가 필요하다.
- 스마트시티 혁신성장동력 프로젝트 사업단에서는 교통, 에너지 등 사회적 인프라에 첨단 ICT 기술 접목하여 데이터 양방향 흐름을 추구함으로써 도시 관리의 효율성을 높이고 시민 삶의 질을 향상시키는 네트워크화된 도시를 스마트시티로 정의한다.
- 과거 인프라 구축 중심의 스마트시티와 달리, 현재의 스마트시티는 AI, IoT, 빅데이터, 클라우드, 블록체인, 5G 등 최신 ICT 기술을 활용하여 인프라/데이터 플랫폼을 연결하여 이를 활용한 다양한 도시 서비스를 제공하는 데 목적이 있다.

## ● 도시 데이터를 활용한 데이터 중심 스마트시티 구현

- 위의 스마트시티 개념은 도시 데이터의 원활한 수집과 활용을 강조하고 있다. 우리가 살고 있는 도시에는 도처에 데이터가 흘러넘치고, 이러한 데이터는 서로 다른 정보시스템에 산재되어 운영되고 있다.
- “데이터는 이제 새로운 석유다(Data is the new oil)”라는 구호가 현실화되기 위해서는 도시에 산재한 서로 다른 종류와 품질의 데이터를 모으고 정제하여 필요한 곳에 공급해 효율적으로 활용하기 위한 노력이 필요하다.
- 이러한 역할을 수행하기 위해 스마트시티 데이터허브 개념이 제시되었고, 「스마트시티 혁신성장동력 프로젝트 사업(국토교통부/과학기술정보통신부, '18~'22)」을 통해 데이터 허브를 설계, 개발, 구축, 운영 및 검증하였다.

## 1 융복합 데이터와 도시 서비스

### ① 융복합 데이터로 부가가치 높은 서비스 제공

- 과거의 스마트시티 실증이나 시범 사업에서는 분야별 도시 서비스가 개별적으로 개발되고 운영되었다. 예를 들어, 교통 서비스는 도시 내 교통 인프라에서 데이터를 수집하여 교통 서비스만을 제공했다. 이전에도 융복합 기술과 서비스 개념은 존재했으나 도시에서는 서비스 분야나 운영·관리 주체에 따라 시스템이 나뉘어 있었다. 따라서 데이터를 한 곳에 모아 융복합하고 활용하는 사례는 많지 않았다.
- 스마트시티 데이터허브는 시스템 관점에서 도시 내 수많은 정보시스템을 연동하고 데이터 수준에서도 이를 융복합할 수 있도록 지원한다. 예를 들어 전기 요금이 저렴할 때 충전하고, 미세먼지 발생과 연계하여 교통을 관제하는 등 교통·에너지 또는 교통·환경 분야를 융합한 서비스가 가능해진다.



〈그림 1-1〉 스마트시티 데이터허브 기반 융복합 서비스

## 1 스마트시티 데이터허브 개발 및 실증

### ① 스마트시티 데이터허브 공통 모델 개발

- 본 사업의 데이터 기반 스마트시티 구현을 위한 대표적인 성과인 스마트시티 데이터허브는 공통 핵심기술 개발을 담당하는 1핵심 사업에서 데이터허브의 참조 모델을 설계하고 개발하였다.



- 스마트시티 혁신성장동력프로젝트 사업 내 1핵심의 주도와 2, 3핵심 연구기관의 참여로 기술분과위원회를 운영하여 데이터허브의 유즈케이스, 요구사항, 참조구조 그리고 인터페이스와 프로토콜을 협력하여 개발하며 스마트시티 데이터허브 공통 모델을 정의하였다.
- 또한 도시 실증을 수행하는 2, 3핵심 사업에서는 데이터허브 공통 모델을 기반으로 확장하여 각각 대구와 시흥의 상이한 요구사항을 반영한 데이터허브 구현 및 구축을 수행하였다.

### ● 스마트시티 데이터허브 오픈소스

- 1핵심은 레퍼런스 데이터허브 구현을 오픈소스(CityDataHub)로 제공한다. 공통 모델 기반으로 상세 아키텍처를 설계하고 인터페이스와 프로토콜을 구현하였으며, 제공하는 다양한 모듈을 데이터 플랫폼으로 통합하여 오픈소스 소프트웨어로 공개하였다.
- 오픈소스 제공을 통해 이를 활용한 데이터허브 구축뿐만 아니라 다양한 데이터허브 상용 솔루션 개발 및 개방형 생태계 조성을 추진한다.



〈그림 1-2〉 스마트시티 혁신성장동력 사업단의 데이터허브 개발 및 실증

## 2 | 특징 및 기대효과

### **D** 도시 데이터 통합관리

#### ● 도시 데이터 인프라를 통합한 도시 데이터 통합관리 및 활용

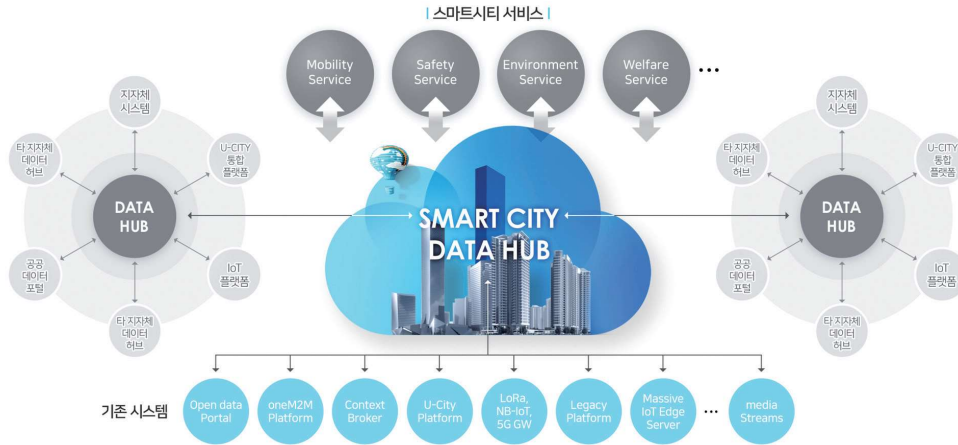
- 도시에는 이미 많은 데이터 활용 시스템들이 운영되고 있다. IoT 플랫폼, 빅데이터 플랫폼, 스마트시티(U-City) 통합플랫폼, 그리고 다양한 행정시스템에 도시에서 발생하는 다양한 데이터가 수집되고 개별적으로 활용되고 있다.
- 스마트시티 데이터허브는 기존 시스템에 산재한 데이터를 수집하여 여러 서비스에서 활용할 수 있도록 통합관리한다. 서로 다른 분야의 서비스에서 활용 가능하도록 원천 데이터를 활용도가 높은 공통 데이터 모델에 맞게 변환하여 저장하고 표준 인터페이스로 데이터 조회 및 활용을 지원한다.

#### ● 기존 도시 데이터 시스템과 데이터허브 간 공생(共生)

- 데이터허브는 기존 도시 데이터 시스템을 대체하는 개념이 아닌 기존 시스템의 데이터를 융복합하여 활용하기 위한 상호협력적인 시스템으로 볼 수 있다. 도시 인프라에서 데이터를 수집하는 기존의 시스템들은 하드웨어, 네트워크 등 해당 분야의 요구사항을 만족하는 시스템으로 지속 운영될 것이다. 아울러 데이터허브는 종래 시스템에서 수집한 데이터를 통합관리하고 활용하는 도시 데이터 생태계 조성의 마중물 임무를 수행할 것이다.

#### ● 도시 데이터 플랫폼에 허브 앤 스포크(Hub & Spoke) 모델 적용

- 허브 앤 스포크는 기존 도시 데이터 시스템을 데이터허브와 연동하여 데이터를 통합 관리하는 모델로 이해할 수 있다. 예를 들어 스포크에 해당하는 공공 데이터 포털이나 oneM2M 플랫폼의 도시 데이터를 데이터허브로 수집하여 통합관리하고 융복합 활용할 수 있다.
- 뿐만 아니라 데이터허브 간에도 Hub & Spoke 모델로 운영하는 개념으로 확장할 수 있다. 예를 들어 기초지자체와 광역지자체의 데이터허브 연계 모델 또는 인접 지자체 간의 데이터 연계(Federation) 활용 방안을 생각해볼 수 있다. 예를 들어, 하천 데이터나 광역 교통 데이터는 여러 지자체에서 연계 활용할 데이터로 볼 수 있다.



〈그림 1-3〉 도시 데이터 시스템 통합 및 효율적 데이터 활용을 위한 데이터허브

## 표준 인터페이스 및 데이터 모델 지원

### NGSI-LD 인터페이스 표준

- 스마트시티 데이터허브는 데이터 저장, 조회 등의 기능 구현에 ETSI(유럽통신표준화기구)에서 개발한 NGSI-LD 표준 인터페이스를 적용하였다. 이는 데이터를 생성·수집하는 IoT나 공공 데이터 플랫폼보다 상위 계층에서 데이터를 통합 관리하기 위한 표준이다.
- NGSI-LD 표준을 적용한 데이터허브는 NGSI-LD 표준을 활용한 국내외 다른 솔루션과도 상호운용성을 가지는 장점이 있다. 스마트시티 데이터허브 플랫폼에 적용된 NGSI-LD 인터페이스를 적용한 도시 서비스는 다른 데이터허브와도 연동 가능하다. NGSI-LD는 국내뿐만 아니라 FIWARE 등을 통해 해외에서도 이를 활용한 다양한 소프트웨어가 개발되고 있어 향후 연계 활용이 가능하다.
- ETSI에서 NGSI-LD 표준에 대한 시험 표준을 개발하고 있으며 향후 다양한 표준 구현 솔루션 간의 상호운용성을 검증하기 위한 시험행사를 개최할 것으로 예상된다.
- 표준 기술은 표준화 활동을 통해 누구나 기술 개발에 참여할 수 있는 개방성을 띠며, 제품화 적용 시 해당 표준을 사용하는 다른 솔루션과 생태계를 구축할 수 있는 추가적인 장점도 있다.
- NGSI-LD 표준 인터페이스는 강력한 데이터 조회(Query) 기능을 제공한다. 위치 기반 조회, 시간기반 조회 그리고 관계형 데이터베이스에서 사용하는 SQL 수준

의 풍부한 조회 조건을 지원한다. 또한 도시 데이터 통합관리 측면에서 하나 이상의 NGSI-LD 기반 플랫폼을 연합(Federation)해서 사용할 수 있는 인터페이스를 추가 개발하고 있다.

- 또한 NGSI-LD는 FIWARE 오픈소스로 구현된 과거 OMA (Open Mobile Alliance) NGSI 인터페이스의 후속 표준격으로 링크드 데이터(Linked Data) 개념을 포함하고 있으며, 도시 데이터 통합관리 플랫폼에 적용 시 향후 데이터 관계성에 기반해 융복합 데이터 서비스로 활용이 가능하다.

## Linked Data와 융복합 서비스

### 스마트시티 공동 데이터 모델

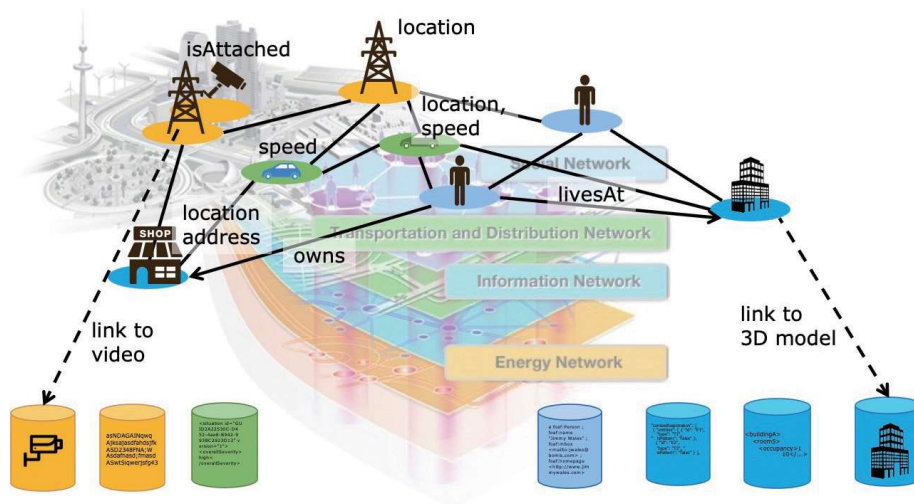
- 도시 데이터는 서비스 분야 그리고 서비스 제공자와 같은 다양한 이해관계자에 따라 상이한 데이터를 정의해서 사용한다. 여기서 데이터가 상이하다는 것은 데이터의 표현(Syntax)과 의미(Semantic)가 다르다는 것이다. 예를 들어 날씨 관측 데이터를 여러 시스템으로부터 수집하여 빅데이터 분석이나 관련 애플리케이션 개발에 활용하려는 데, 데이터를 제공받은 시스템마다 데이터 모델(스키마)이 상이하면 이를 정제·변형해서 사용하는 데 시간과 비용이 많이 들 수 있다.
- 따라서 시스템 연동 관점에서 동일한 표준 인터페이스를 사용하는 것뿐만 아니라 도시 데이터의 공통 플랫폼으로서 수집된 데이터 활용성을 높이기 위해 공통 데이터 모델을 사용할 필요가 있다.
- 스마트시티 혁신성장동력프로젝트 사업에서는 서비스 실증을 담당하는 2, 3핵심에서 공통 데이터 모델을 정의하여 사용하였으며 향후 데이터허브가 확산되어 활용할 때도 공통 데이터 모델을 정의하고 관리하는 체계가 운영될 예정이다.

### 스마트시티와 링크드 데이터

- 도시에는 분산된 시스템에 다양한 데이터가 개별(Silo)적으로 관리되고 활용되고 있으며, 스마트시티는 시스템들의 시스템(System of systems)이라고 불리는 것처럼 시스템 간의 연계를 통해 산재한 데이터를 통합해서 활용할 필요가 있다. 그리고 단순히 데이터를 한곳에 모아서 사용하는 수준이 아니라 서로 다른 데이터 간에 연계 활용

이 필요하다. 데이터 연계는 데이터 매시업 또는 서비스 융복합과 같은 맥락에서 이해할 수 있다.

- 도시의 모든 것은 데이터로 표현할 수 있다. 예를 들어, 자동차처럼 사물로 존재하는 것이든 아니면 주차 혼잡도와 같이 개념 혹은 가상의 사물 또한 도시 데이터로 표현된다. 복잡하게 얽혀 있는 도시 데이터는 서로 관계를 맺고 있다. 상이한 도시 데이터 간의 관계성을 표현하는 것은 스마트시티에서 링크드 데이터를 구현하는 것이며, 이를 통해 데이터 융복합 그리고 서비스 융복합을 실현할 수 있다.



〈그림 1-4〉 스마트시티의 링크드 데이터 개념도 (출처: NEC Laboratories Europe)

### ● 링크드 데이터를 지원하는 데이터허브

- NGSI-LD는 그 이름에서 의미하는 것처럼 **Linked Data**를 지원한다. NGSI-LD 표준에서 ‘엔티티’로 표현되는 도시 데이터는 고유한 식별자(Uniform Resource Identifier, URI)를 가지고, 데이터(예: 주차장)의 다양한 속성값(예: 현재 가용 주차면 수)뿐만 아니라 다른 데이터와의 관계(예: 주차장 내 주차면 엔티티)를 나타낼 수 있다.
- 데이터 간의 관계를 추가하여 연결된 데이터를 표현할 수 있으며 서로 다른 데이터가 융합되는 개념으로 이해할 수 있다. 예로 성남시 야탑동의 공영주차장의 실시간 주차 현황 데이터는 성남시 날씨 관측이나 예보 데이터와 연결할 수 있다. 실시간 가용 주차면 수 데이터는 날씨 관측 데이터와 결합하여 기계학습에 응용해 날씨 예보에 따른 주차 가능 면적 수를 예측하는 서비스 등에 활용할 수 있다.

- 데이터 융복합은 데이터 분석의 ETL(Extraction, Transform and Loading)과 같은 절차로 새로운 데이터를 생성하는 것뿐만 아니라 NGSI-LD와 같이 데이터 간의 연결성을 표현하여 이종 데이터를 함께 활용할 수 있는 큰 그림으로도 이해할 수 있다.

## **1** 링크드 데이터와 시맨틱

### **1** 도시 데이터와 시맨틱웹

- 시맨틱웹은 일반 웹과는 다르게 사람이 이해할 수 있는 문서로 구성된 웹이 아닌 기계가 이해할 수 있는 데이터로 구성된 웹을 의미한다. 이를 위해서는 기계가 데이터를 이해할 수 있는 용어(Vocabulary)를 정의해서 데이터 표현에 사용해야 하며 기계가 이해할 수 있는 RDF(Resource Definition Framework) 형식을 따라야 한다.
- 스마트시티에서 채용한 NGSI-LD는 이러한 시맨틱웹의 핵심 개념을 구현하고 있다. RDF 형식의 시맨틱 데이터와 같이 모든 데이터(Entity) 타입 및 식별자, 속성(Property), 관계(Relationship)를 '온톨로지'로 정의하였고 데이터로 표현될 때 모두 고유한 URI를 가진다.
- 관계성을 가지는 데이터는 그래프로 표현된다. 현재 NGSI-LD에서 지원하는 인터페이스는 REST(REpresentational State Transfer) API로서 그래프 데이터 순회 기능을 지원하지는 않는다. 대신 NGSI-LD로 구축한 도시 데이터를 시맨틱 RDF 데이터로 변환하여 시맨틱웹의 대표적인 데이터 질의 기술인 SPARQL(SPARQL Protocol and RDF Query Language)을 이용하여 데이터 링크를 순회하며 데이터 그래프 안에서 원하는 데이터를 탐색하거나 추론하는 등의 활용이 가능하다.



## 1 | 아키텍처

## D 참조구조

## ● 도시 데이터 통합관리 및 활용을 위한 다양한 모듈 지원

- 스마트시티 데이터허브는 도시 데이터의 효율적인 통합관리를 위한 다양한 기능을 모듈 단위로 제공한다. 지자체는 각자의 요구사항에 따라 필요한 기능을 지원하는 데이터허브를 구축하여 활용할 수 있다.
- 스마트시티 데이터허브 참조구조는 도시 데이터 시스템 및 다양한 데이터허브 응용 서비스와 연동하는 데이터허브로 구성된다. 개념적으로 데이터허브는 모듈형 구조를 지원하며, 시스템 구축 입장에서 기능 모듈을 필수 모듈과 선택 모듈로 구분할 수 있다.
- 도시 데이터 통합관리 측면에서 데이터 코어 모듈은 시스템 구축에 필요한 최소 필수 모듈이다. 여기에 통합 인증·인가 보안 기능이 포함된 보안 모듈 및 API 게이트웨이 또한 필수 모듈로 분류할 수 있다.
- 이 외의 모듈은 도시별로 필요에 따라 구축하여 활용할 수 있는 선택 모듈로서 데이터 마켓플레이스, 데이터 분석, 클라우드 관리 모듈 등이 있다.

## ● 마이크로 서비스 아키텍처(MSA) 지원

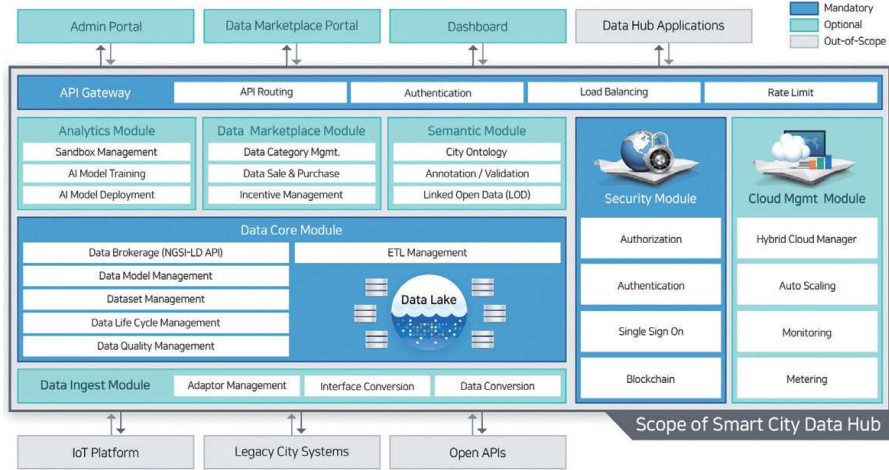
- 데이터허브는 마이크로 서비스 아키텍처를 지원하며 데이터허브의 각 기능 모듈은 하나 이상의 마이크로 서비스로 구성되어 있다.
- 서비스 레지스트리를 통해 전체 서비스를 관리하고 서비스 검색을 통해 서비스 간의 통신을 동적으로 관리할 수 있다.

## ● 이벤트 기반 아키텍처 지원

- 마이크로 서비스 간, 모듈 간 발생하는 이벤트를 전달하기 위해 **Apache Kafka** 오픈소스의 메시지 큐를 활용한 이벤트(Publish/subscribe) 기반 아키텍처를 지원한다. 예를 들어, 데이터 모델 변경 이벤트, 데이터셋 변경 이벤트, 데이터 구독을 통한 통지 메시지 발생 이벤트 등을 다른 마이크로 서비스 또는 다른 모듈로 전달이 가능하다.

## ● 아키텍처 기반 데이터 흐름

- 도시에서 발생하는 다양한 데이터는 데이터 코어 모듈에 저장된다. 데이터허브에서는 데이터 모델(스키마)을 관리하며, 데이터는 데이터셋 단위로 저장된다. 그리고 **NGSI-LD API**로 다른 모듈 또는 외부 서비스에 데이터를 제공한다.
- 데이터 모델이 존재하지 않는 데이터, 예컨대 데이터 가공이 추가로 필요한 데이터도 데이터 코어에 저장할 수 있다. 이 경우 해당 데이터는 **Apache Kafka**에 임시로 저장되며 별도의 추가 가공 후 스키마 기반의 데이터로 변환하여 데이터 저장소에 저장할 수 있다.
- 도시 데이터는 **NGSI-LD API**를 구현한 소프트웨어를 통해 데이터 코어 모듈에 데이터를 저장할 수 있고, 도시 내 레거시 시스템의 데이터는 데이터 수집 모듈을 통해 데이터를 취합하고 **NGSI-LD**로 변환하여 데이터 코어 모듈에 저장할 수 있다.
- 데이터 코어 모듈은 데이터셋 별로 데이터의 저장소를 선택할 수 있다. **PostgreSQL**, **Hive**, **HBase** 중 하나를 선택할 수 있으며 이는 데이터의 특성, 트래픽량 등을 고려하여 선택할 수 있다.
- 데이터 분석 모듈은 저장된 데이터를 기반으로 배치 분석을 수행하고, 그 결과를 데이터 코어 모듈에 적재할 수 있다. 데이터 코어 모듈은 원천 데이터는 물론 분석 결과까지 저장할 수 있게 되며 이렇게 저장된 데이터는 데이터 마켓플레이스를 통해 외부에 공개될 수 있다.
- 데이터허브 애플리케이션은 데이터 코어 모듈의 **NGSI-LD API**를 연동하여 도시의 원천 데이터, 가공 데이터, 분석 데이터를 획득할 수 있으며, 이를 통해 도시 서비스를 제공할 수 있다. 데이터허브 애플리케이션도 서비스에서 생성된 데이터를 데이터 코어 모듈에 저장하여 다른 이해관계자에게 제공할 수 있다.



〈그림 2-1〉 스마트시티 데이터허브 참조구조

## 1 구축 방안

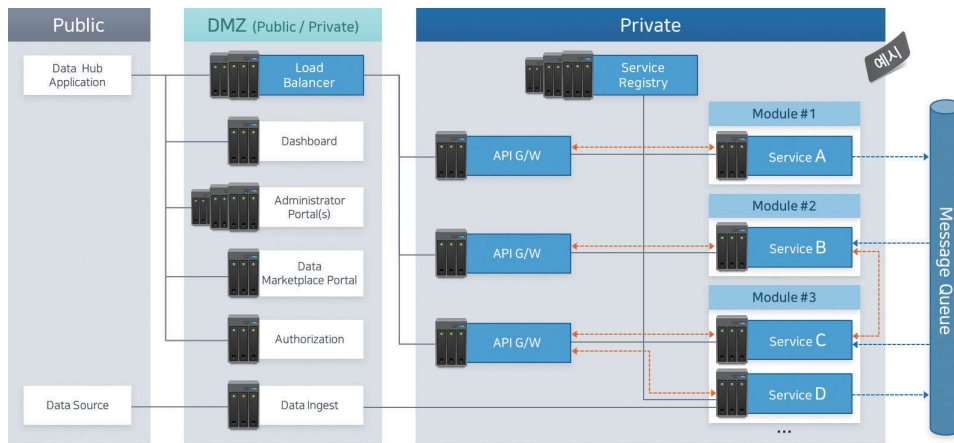
### ● 하드웨어 구성

- 스마트시티 데이터허브는 소프트웨어이기 때문에 온프레미스, 클라우드의 제약 없이 구성할 수 있다. 온프레미스에 구축할 시에는 구축하는 보안 요구사항에 따라 L3 또는 L4 스위치, 방화벽, 망연계 장비 등의 네트워크 장비의 도입이 필요할 수 있다.

### ● 네트워크 구조

- 데이터허브에 대한 네트워크 구조는 구성하는 주체에 따라 다양하게 구성이 가능하나, 기본적인 원칙은 API 게이트웨이를 통해 내부 API들을 보호하고 웹 애플리케이션은 API 게이트웨이를 통해 내부 API들을 호출하는 구성이다.
- 다음 그림은 DMZ 구간에 로드밸런서(예: L4 스위치, HA Proxy)를 두고 API 게이트웨이를 통해 서비스들의 API를 호출하고 이외의 애플리케이션의 경우 로드밸런서로 API들을 호출하도록 구성한 예시이다.
- 내부 서비스는 서비스 레지스트리(Eureka)를 통해 연동되도록 구성하였다.

• **Note** • 서비스 메쉬(Service Mesh) 도입도 논의가 되었으나 데이터허브가 지자체에 구축될 시 온프레미스(On-premise)로 구성되며, 온프레미스 환경에서의 클라우드 구성이나 쿠버네티스(Kubernetes) 환경 구성이 현실적으로 어렵다는 점으로 인해 도입하지 않았다.



〈그림 2-2〉 스마트시티 데이터허브 네트워크 구성도 예시

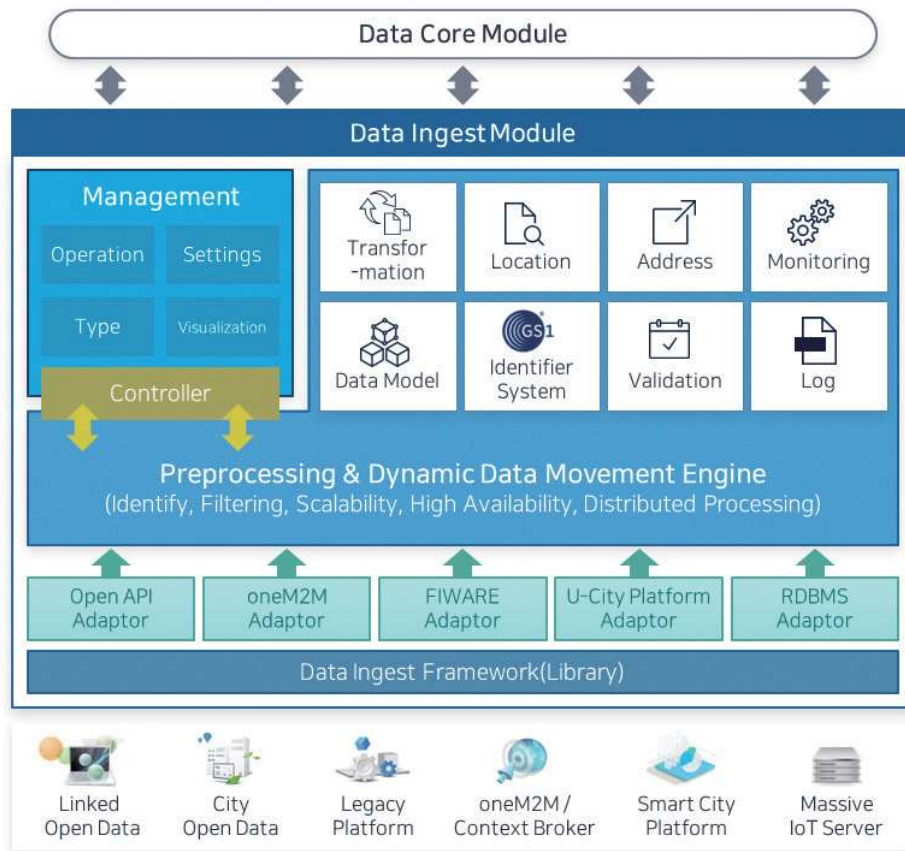
## 2 | 모듈 정의 및 구성

### 1 데이터 수집 모듈

#### ● 도시 데이터 인프라로부터 데이터를 수집하여 데이터허브로 저장

- 도시 데이터는 IoT 플랫폼, 공공 API 서비스, 다양한 정보시스템에 산재되어 존재한다. 이러한 시스템은 상이한 데이터 모델과 데이터를 제공하기 위한 별도의 인터페이스 및 프로토콜을 사용한다.
- 따라서 기존 도시 데이터를 데이터허브로 수집하여 활용하기 위해서는 데이터 모델과 프로토콜을 변환하기 위한 어댑터가 필요하다. 데이터 수집 모듈은 다양한 어댑터를 제공하며 도시 데이터 수집 개발을 지원하고 수집 현황을 관리하는 기능을 제공한다.
- IoT 플랫폼의 경우 대표적인 IoT 표준인 oneM2M 표준 플랫폼과 연동할 수 있는 어댑터를 제공하며, 공공 데이터 포털의 API 템플릿을 제공하여 실제 연동하는 외부 데이터 시스템과의 데이터 연계를 손쉽게 구축할 수 있게 한다.
- 또한 스마트시티(U-City) 통합플랫폼, NGSI 인터페이스를 사용하는 FIWARE Context Broker 연동 어댑터, 그리고 표준 SQL을 사용하는 RDBMS 연동을 위한 어댑터를 제공한다.

- 또한 데이터 수집, 변환, 검증 현황에 대한 모니터링 기능을 제공하여 도시 데이터 수집을 체계적으로 관리할 수 있는 기능을 제공한다.



〈그림 2-3〉 데이터 수집 모듈 개념도

### ● Flume 기반의 데이터 수집 모듈 오픈소스 제공

- CityDataHub의 데이터 수집 모듈은 Apache Flume 오픈소스 프레임워크를 기반으로 개발되어 Flume에서 제공하는 에이전트 및 어댑터 개념을 활용한다. 사용자는 데이터 수집 단위(대상 시스템 및 인스턴스)별로 어댑터를 설정하며, 다수의 어댑터를 에이전트로 묶어 실행 관리를 수행한다.

순번	Adapter 명	시각/종료	인스턴스...	상태	Messages in / hour	Messages out / hour	Bytes in / hour	Bytes out / hour	Failure Messa... / hour	모니터링
1	성남시 기상관측 어댑터	시각	1	Stop	83	5	7327	3689	15	모니터링

〈그림 2-4〉 데이터 수집 모듈의 어댑터 실행 관리 화면

## 데이터 코어 모듈

### 표준 API를 적용한 도시 데이터 관리 기능

- 데이터 코어 모듈은 ETSI NGSI-LD 표준 API를 지원한다. 이는 데이터허브에서 외부 도시 서비스에 데이터를 제공하는 인터페이스가 표준을 따르기 때문에 서비스가 지자체에 구축된 데이터허브 시스템들과 호환될 수 있음을 의미한다.
- 데이터허브에 저장된 데이터를 활용하는 제3의 모듈을 NGSI-LD API로 연동하여 활용할 수 있으며, 국내뿐만 아니라 외국의 서비스와 플랫폼과도 표준 API로 상호 운용이 가능하다. API 사용은 표준 기반 기술 생태계가 조성됨을 의미한다.
- NGSI-LD API는 데이터 저장/갱신/삭제, 최신 데이터 조회, 이력 데이터 조회, 위치 기반 조회(Geo-query), 구독/통지 등 다양한 데이터 관리 기능을 제공하고 있다.

### 데이터 모델과 데이터셋 관리 기능

- 이질적인 데이터는 스마트시티 서비스 호환성 및 이식성을 저해할 뿐만 아니라 서비스 개발 및 데이터 분석에 있어 데이터 정제에 많은 시간과 비용을 소모하게 한다. 스마트시티 데이터허브는 도시에서 수집한 데이터를 관리하고 도시 서비스에 제공함에 있어 공통 데이터 모델(스키마)을 사용한다.
- CityDataHub 오픈소스의 데이터 코어 모듈 관리자는 데이터허브에서 사용할 데이터 모델을 등록하고 데이터 인스턴스를 그룹으로 관리할 수 있는 데이터셋을 생성하여 데이터를 저장하고 사용한다.
- 데이터 모델은 NGSI-LD 표준에서 사용하는 JSON-LD (Javascript Simple Object Notation-Linked Data) 포맷 데이터를 정의하는 스키마를 관리할 수 있다. 관리자 화면

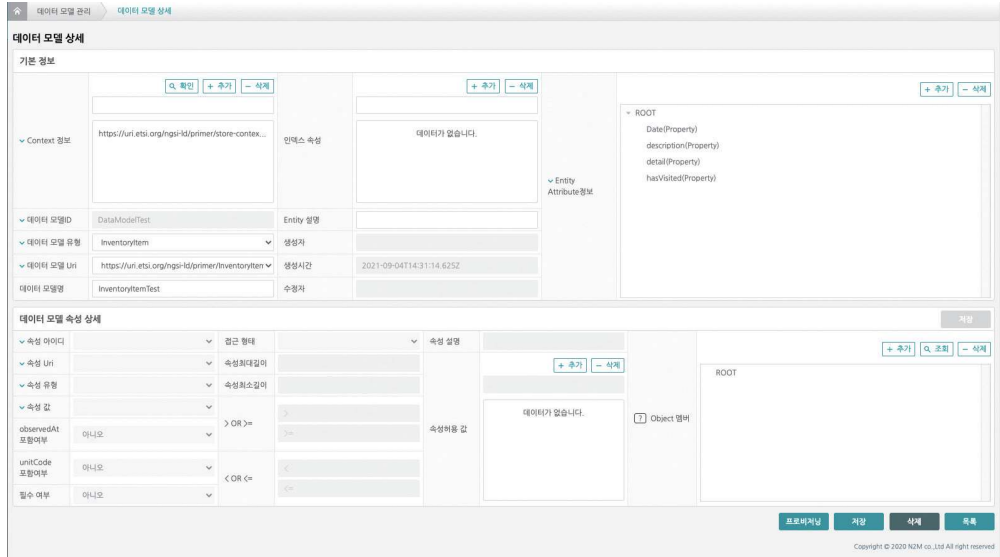


을 통해 컨텍스트, 속성별 데이터 포맷, 유효범위, 필수여부, 하위속성 등을 정의할 수 있다.

```
{
  "@context": [
    "http://uri.citydatahub.kr/ngsi-ld/citydatahub-contexts.jsonld"
  ],
  "id": "urn:citydatahub:OffStreetParking:yatap_01",
  "type": "OffStreetParking",
  "name": {
    "type": "Property",
    "value": "yatap_01"
  },
  "location": {
    "type": "GeoProperty",
    "value": {
      "type": "Point",
      "coordinates": [
        127.1293735,
        37.4114424
      ]
    }
  },
  "address": {
    "type": "Property",
    "value": {
      "addressCountry": "KR",
      "addressRegion": "Gyeonggi-do",
      "addressLocality": "Seongnam-si",
      "streetAddress": "8th Seungin-ro",
      "addressTown": "Yatap-dong"
    }
  },
  "totalSpotNumber": {
    "type": "Property",
    "value": 110
  },
  "availableSpotNumber": {
    "type": "Property",
    "value": 24,
    "observedAt": "2022-07-15T08:09:55.000+09:00"
  },
  "maximumAllowedHeight": {
    "type": "Property",
    "value": 2.2
  },
  "refParkingSpots": {
    "type": "Relationship",
    "value": [
      "urn:datahub:ParkingSpot:yatap_540",
      "urn:datahub:ParkingSpot:yatap_541"
    ]
  }
}
```

### NGSI-LD 주차장 데이터 예시

- 도시 데이터 시스템으로부터 데이터 수집하기 전에 먼저 데이터 모델을 등록하고 데이터를 저장할 데이터셋을 생성한다. 데이터셋은 예를 들어 같은 동네에 설치된 다수의 미세먼지 센서의 측정 데이터를 표현할 수 있다.

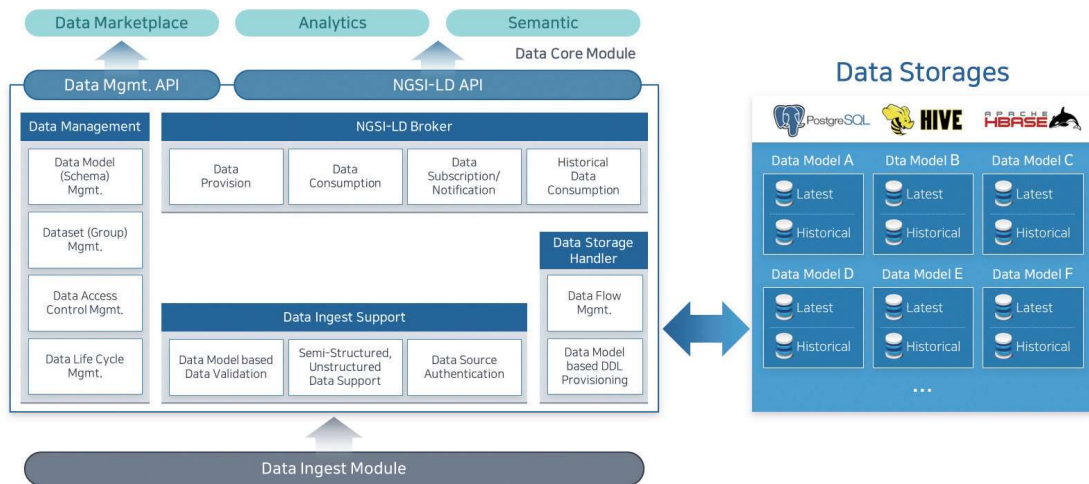


〈그림 2-5〉 데이터 코어 모듈 관리자의 데이터 모델 설정 화면

- 데이터셋 소유권, 라이선스 등 다양한 메타데이터를 입력하여 관리할 수 있다. 또한 데이터셋 단위로 접근제어를 수행할 수 있으며, 데이터 모델 기반 검증 및 생애주기를 관리하는 단위로 사용된다.

## ● 다양한 데이터 저장소 유형 지원

- 오픈소스의 데이터 코어 모듈은 기본으로 RDBMS인 PostgreSQL과 하둡(Hadoop) 기반 빅데이터 스토리지인 Hive와 HBase를 지원한다. 사용자는 데이터셋 설정 GUI에서 데이터셋마다 저장되는 스토리지를 다르게 설정할 수 있으므로 수집 데이터나 활용 서비스 특성에 맞게 스토리지를 선택하여 운용할 수 있다.



〈그림 2-6〉 데이터 코어 모듈 개념도

## 1 분석 모듈

- 데이터허브의 도시 데이터를 활용한 데이터 분석 도구 제공
- 기존 AI를 활용한 데이터 분석 도구 및 서비스와 차별화되게 스마트시티 데이터허브의 분석 모듈은 데이터허브의 데이터 코어 모듈에 저장된 실제 데이터를 활용한 내재화된 분석 도구라는 특징을 가진다.
- 데이터 분석가는 분석 모듈에서 제공하는 다양한 데이터 분석 도구를 이용해 데이터허브의 데이터를 확인하고 전처리한 후 원하는 AI 알고리즘으로 학습 모델을 생성하고 검증할 수 있다.
- 생성된 AI 학습 모델을 일정 시간 간격으로 실행하여 AI 기반 데이터 예측 서비스를 개발할 수 있다. 매시간 수집되는 날씨 및 미세먼지 예보 데이터를 활용, 향후 24시간 동안 가용 주차면을 예측하는 서비스를 개발할 수 있다.



〈그림 2-7〉 분석 모듈 개념도

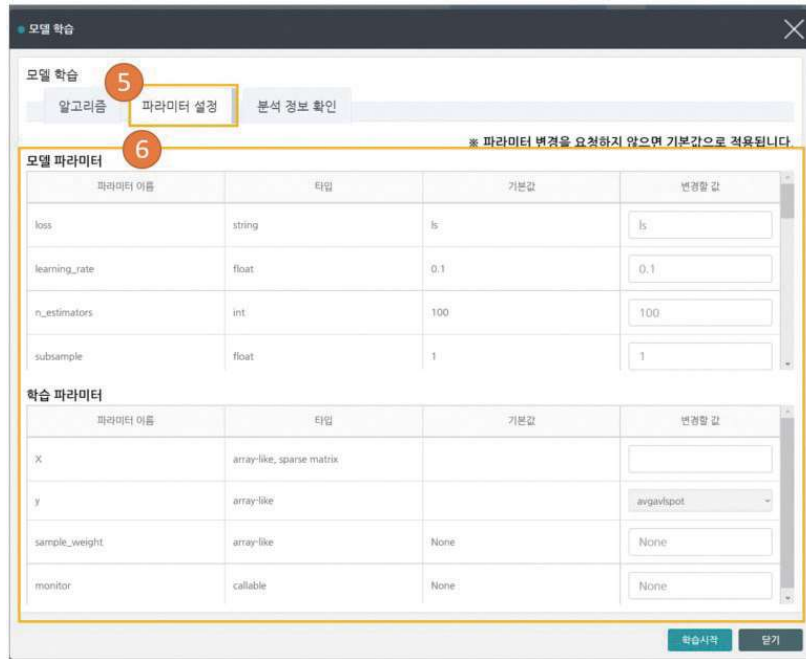
## ● ETL 도구

- 데이터 분석 서비스 개발을 위해 우선 사용 가능한 데이터를 Apache Hue를 통해 조회하여 확인한다. 이후 학습 모델 생성을 위해 필요한 데이터를 가공하기 위해 ETL 과정을 거치는데 CityDataHub는 오픈소스 ETL 도구인 Apache NiFi를 제공하여 데이터 분석가가 ETL을 수행할 수 있다. 사용자의 편의를 위해 자주 사용하는 NiFi 데이터플로우를 템플릿으로 제공하여 몇 가지 설정 변경으로 ETL을 손쉽게 수행할 수 있다.
- ETL을 통해 수집된 도시 원천 데이터를 변형시킬 수도 있고 이종 데이터를 하나의 데이터로 융합하여 AI 모델 학습에 필요한 데이터를 만들어낼 수 있다.

## ● 분석 도구

- 데이터 분석가가 사용하는 CityDataHub의 데이터 분석 도구는 웹 기반 서비스로 제공되며, ETL 결과 데이터를 AI 학습으로 진행하기 위한 데이터 전처리, 학습 알고리즘 선택 및 파라미터 설정, 학습 그리고 학습된 모델의 검증을 수행한다.
- 분석 도구는 현재 Scikit-learn 파이썬 라이브러리의 전처리 및 학습 알고리즘 20여 종을 지원한다.

- 생성된 모델은 분석가의 요청에 의해 분석 모듈 관리자가 배치 실행을 관리한다. 학습에 사용한 ETL 프로세스와 AI 모델 그리고 모델 실행 결과를 저장할 데이터 경로 정보를 통해 AI 모델 실행에 따른 예측(추론) 결과를 데이터허브에 다시 저장하여 외부 서비스에서 이를 활용할 수 있는 체계를 가지고 있다.



〈그림 2-8〉 기계학습 모델 생성을 위한 파라미터 설정

## 데이터 마켓플레이스 모듈

### 데이터 마켓플레이스 서비스 제공

- 데이터 마켓플레이스 모듈은 데이터허브 사용자에게 데이터 코어 모듈에 저장된 도시 데이터를 유통할 수 있는 데이터 마켓플레이스 포털 서비스를 제공한다.
- 유통 대상 데이터는 데이터 수집 모듈이 수집하여 데이터 코어 모듈에 저장한 원천 데이터이거나, ETL 및 분석 과정에서 생성된 가공 데이터일 수 있다. 또한 데이터 유통 방법은 데이터를 무료로 제공하거나 원하는 정책을 설정하여 유료로 거래하는 방법을 지원한다.
- 데이터 상품 출시 시 데이터 코어 모듈에 저장된 데이터의 원하는 특정 속성을 선택하거나 속성명을 변경할 수 있으며, 데이터 상품 구매 후 서비스에서 데이터가 필요

할 때 조회하거나 통지 메시지를 통해 신규 가용 데이터 발생 시 이를 이벤트 메시지로 수신할 수 있다.

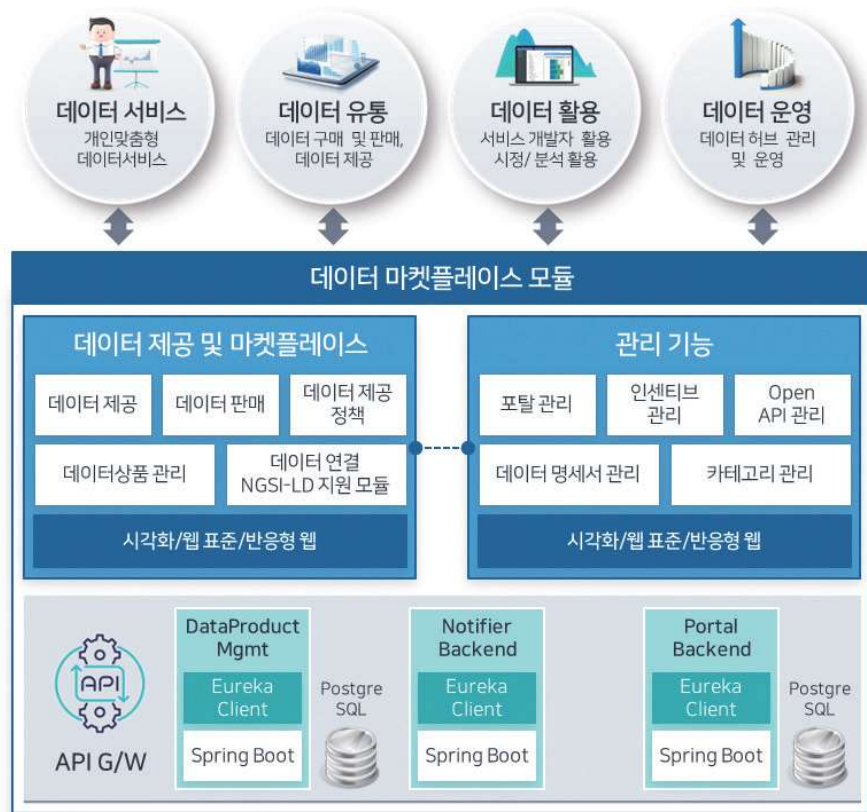
### ● 데이터 스토어의 기본적인 비즈니스 모델을 데이터허브에 구현

- 데이터 마켓플레이스 모듈은, 데이터 스토어 개념을 스마트시티 데이터허브에 구현한 것으로, 차이점으로는 데이터 유통을 위해 사용자가 데이터 마켓플레이스 포털에 데이터를 별도로 업로드하지 않고 데이터 코어 모듈에 저장된 자신의 데이터를 상품으로 출시하여 다른 사용자가 포털에서 검색하여 활용할 수 있다.
- 출시하는 데이터 상품은 판매자의 선택에 따라 무료/유료로 제공할 수 있다.
- 판매자가 데이터 상품을 출시하면 이를 관리자가 검수하여 승인하고 다른 사용자가 이를 구매·이용할 수 있는 데이터 유통 모델을 제공한다.
- 또한 마켓플레이스 관리자와 이용자에게 데이터 상품 관리를 위한 카테고리 관리, 판매 및 구매 데이터에 대한 이력 관리 기능 등을 제공한다.

### ● 데이터 유통 활성화와 블록체인 연계

- 블록체인의 다양한 활용 방안 중, 데이터 유통거래 활성화를 위한 인센티브 플랫폼으로 데이터 마켓플레이스와 연동을 제공한다.
- 데이터 사용자는 구매한 상품에 대해 평점을 남기고 인센티브를 지급받을 수 있으며, 이는 사용자의 블록체인 지갑에 토큰을 발행하는 형식으로 동작한다.
- 발행된 인센티브 토큰은 사용자 이용 정보에서 확인할 수 있으며 향후 발급받은 인센티브를 활용한 추가적인 서비스 응용이 가능하다.



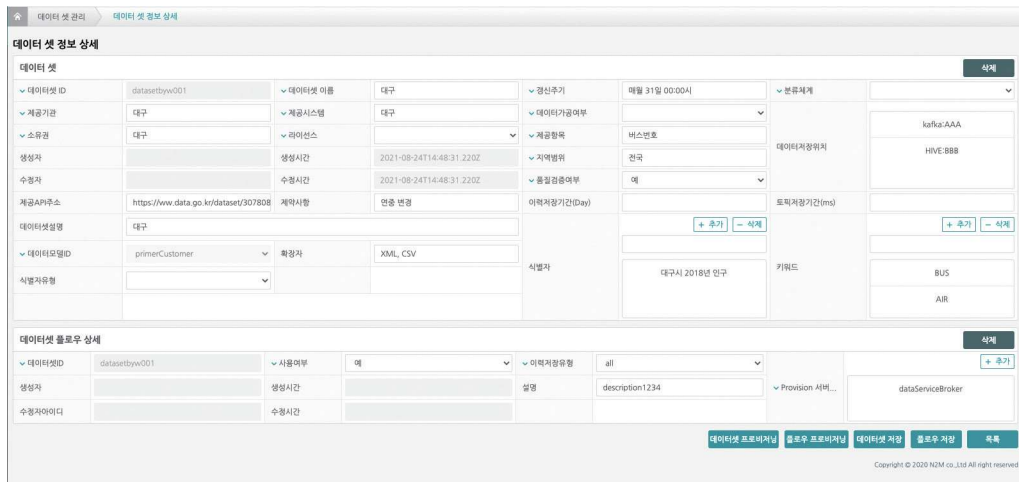


〈그림 2-9〉 데이터 마켓플레이스 모듈 개념도

## 시맨틱 모듈

### 시맨틱 웹기술을 적용한 스마트시티 데이터 활용 기술 제공

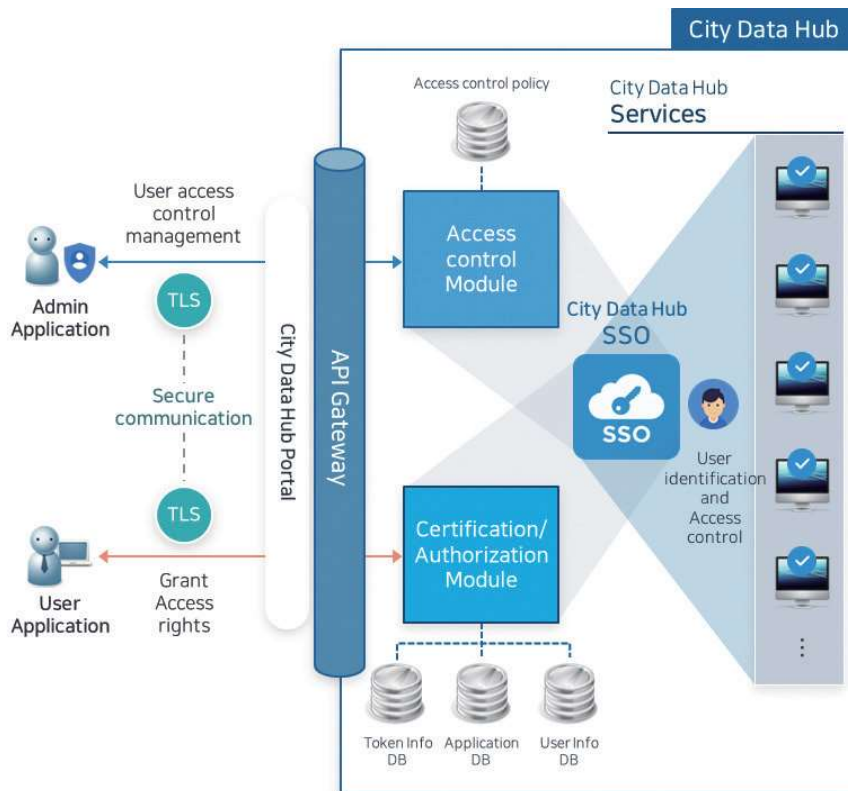
- 시맨틱 모듈은 데이터 코어 모듈에 저장한 데이터를 스마트시티 온톨로지 기반 시맨틱 주석(Annotation) 데이터로 변환하여 시맨틱 데이터베이스인 트리플 스토어에 구축하고 시맨틱웹 기술을 활용하는 기능을 제공한다.
- 시맨틱웹의 **RDF** 형식으로 구축된 데이터는 **SPARQL** 질의를 통해 복잡한 데이터 질의 수행, 연관성 추론 등 데이터 코어 모듈의 **REST API**에서 제공하지 않는 다양한 질의를 수행할 수 있다.



〈그림 2-10〉 데이터 코어 모듈 관리자의 데이터셋 설정 화면

- 시맨틱 데이터 구축을 위해 스마트시티 공통 온톨로지를 정의하였으며 이는 교통, 환경 등과 같은 도메인 온톨로지로 확장 정의하여 RDF 데이터 구축에 활용된다. 스마트시티 공통 온톨로지 활용 확산을 위해 oneM2M 표준에 이를 반영(OneM2M TR-0061 (Smart City Ontology)하였다.
- 기존 데이터를 시맨틱 데이터로 변환하여 저장하기 위해서는 필요한 온톨로지를 정의하고 이를 소프트웨어로 구현하는 과정이 필요하며 이를 돕기 위해 CityDataHub에서는 시맨틱 주석 템플릿 생성 도구(Annotation Template Creation Interface, ATCI)와 시맨틱 주석기(Annotator)를 제공한다. 이후 생성된 시맨틱 데이터의 온톨로지 정합성을 검증하고 오픈소스로 제공되는 Virtuoso 트리플 스토어에 이를 저장한다.
- 이와 같은 종래 시맨틱웹 응용 기술을 데이터 코어 모듈과 연동하여 제공할 뿐만 아니라 SPARQL에 익숙하지 않은 기존 REST API 개발자를 고려해 시맨틱 데이터 이용을 위한 REST API도 개발되어 함께 제공된다.
- RDF로 구축한 데이터도 그래프 형태로 SPARQL을 활용한 다양한 질의나 정보 추출이 가능하며 또 다른 그래프 데이터 형식인 Labeled Property Graph(LPG)로 데이터 코어의 데이터를 변환하여 저장하고 이를 활용한 연구도 수행되었다. NGSI-LD의 엔티티와 엔티티 간의 관계를 기존 LPG로 표현 가능하며 그래프를 순회하며 원하는 데이터를 빠르게 탐색할 수 있다. CityDataHub의 시맨틱 모듈은 대표적인 그래프 데이터베이스인 Arango를 활용한다. 이를 통해 NGSI-LD 데이터를 LPG

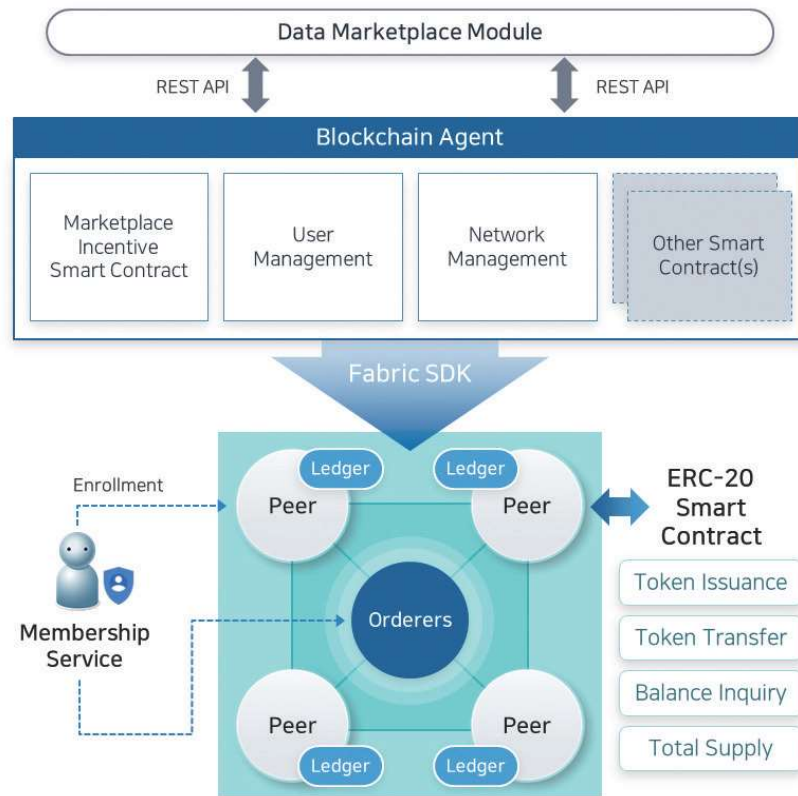




〈그림 2-12〉 인증·인가 모듈 개념도

### ● 블록체인 활용 서비스

- 프라이빗 블록체인인 Hyperledger Fabric을 활용한 플랫폼에 ERC-20 기반 스마트 컨트랙트 기능을 제공한다.
- 다양한 블록체인의 유즈케이스 중 CityDataHub에서는 데이터 마켓플레이스 서비스의 인센티브 관리를 위해 블록체인 사용자 관리 및 스마트 컨트랙트 기능을 제공한다.
- 사용자마다 블록체인 지갑을 생성하고 발급된 인센티브 토큰을 블록체인에 기록하여 이를 다른 서비스 시나리오에 활용할 수 있다.

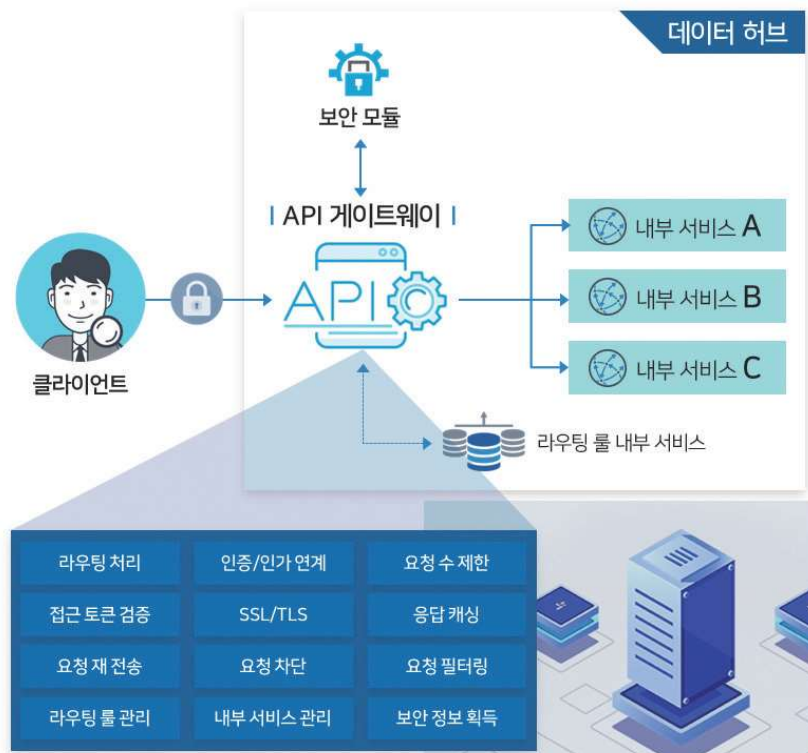


〈그림 2-13〉 블록체인 모듈 개념도

## 1 API 게이트웨이

### ● 외부 사용자 및 서비스로부터 데이터허브를 보호

- 데이터허브 시스템 보호를 위해 외부 클라이언트(서비스)는 데이터허브 내부 모듈과 직접 통신할 수 없으며, 외부의 요청 메시지를 API 게이트웨이가 이를 먼저 수신하고, API 게이트웨이는 요청을 해당하는 모듈로 라우팅한다.
- API 게이트웨이는 외부 클라이언트와 데이터허브 간 통신의 기밀성을 보장하며, 내부 모듈로 라우팅하기 전 요청 메시지에 포함된 토큰을 검증하고 토큰에 적용된 역할 기반 접근 제어를 수행한다. 해당 모듈로 메시지를 전송하기 전, 해당 클라이언트가 해당 모듈에 접근할 권한이 있는지 사전에 판단한다.
- 또한 시스템 과부하를 방지하기 위해 외부 클라이언트의 시간당 최대 요청 수를 제한하거나 필요시 부적절한 클라이언트의 요청 메시지의 라우팅을 차단하여 내부 모듈을 외부로부터 보호한다.



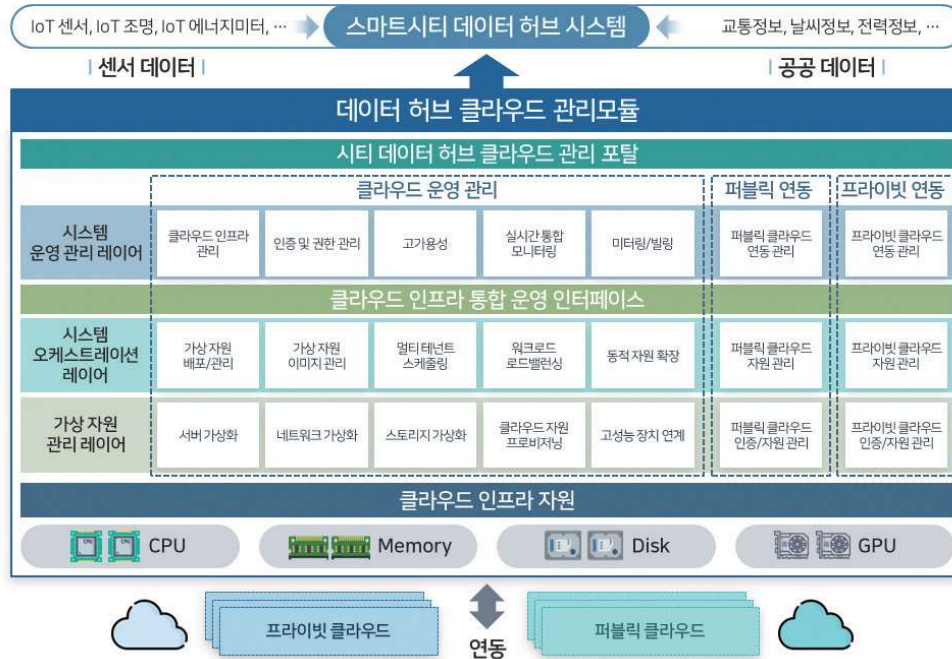
〈그림 2-14〉 API 게이트웨이 개념도

## 클라우드 관리 모듈

### 하이브리드 클라우드 관리 기능 제공

- 다양한 기능 모듈을 지원하는 것처럼 스마트시티 데이터허브는 다양한 시스템 구축 방안을 지원하며 클라우드 컴퓨터 환경에서 데이터허브를 운용할 때 클라우드 관리 모듈을 활용할 수 있다.
- 데이터허브의 클라우드 관리 모듈은 기존 퍼블릭(예:AWS) 또는 프라이빗(예: OpenStack) 클라우드뿐만 아니라 퍼블릭과 프라이빗 클라우드를 함께 운용하는 하이브리드 클라우드를 구성하고 사용할 수 있다. 예를 들어, 데이터 관련 규제나 데이터 특성상 프라이빗 클라우드에 데이터 코어 모듈을 구축하고 공개 활용 가능한 데이터를 포함한 데이터 분석 환경을 퍼블릭 클라우드에서 제공하여 초기 인프라 구축 비용 및 컴퓨팅 자원의 효율적인 운용을 하이브리드 클라우드로 제공할 수 있다.

- CityDataHub의 클라우드 관리 모듈은 퍼블릭 클라우드의 경우 AWS, Azure 그리고 국내의 TOAST 클라우드를 지원하며, 프라이빗 클라우드의 경우 오픈소스인 OpenStack을 지원한다. 지원하는 클라우드 서비스에 대한 통합 자원 관리, 대시보드, 모니터링 및 미터링 기능을 제공한다.



〈그림 2-15〉 클라우드 관리 모듈 개념도

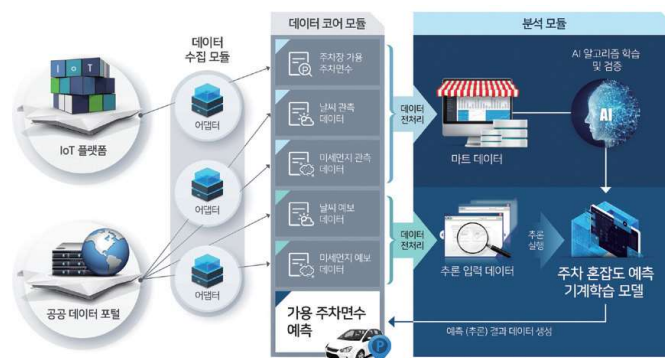


## 1 | 개념검증(Proof of Concept)

## 주차 혼잡도 예측 PoC

## 데이터허브를 활용한 융복합 서비스

- 스마트시티 데이터허브는 도시의 다양한 데이터를 통합관리하는 것뿐만 아니라 여러 분야의 데이터를 융복합한 서비스, 그리고 융복합 분석 서비스를 실증하는 것이 중요한 목표이다. 레퍼런스로 개발한 CityDataHub를 활용해 융복합 분석 서비스 개발 가능성을 검증하고자 주차 혼잡도 예측 서비스를 기획하고 개발하였다.
- 데이터는 국내외에서 스마트시티 구축 및 다양한 IoT 솔루션 개발에 활용되고 있는 oneM2M 표준의 오픈소스 플랫폼인 Mobius에서 수집하는 실시간 주차 데이터를 활용하였다. 추가로 주차 혼잡도에 영향을 미치는 데이터로서 공공 데이터 포털을 통해 날씨와 미세먼지 데이터를 수집하였다. 데이터 수집에는 수집 모듈의 oneM2M 및 공공 데이터 어댑터를 활용하였다.
- 공영주차장의 주차 가용면 수와 날씨 및 미세먼지 이력 데이터를 학습하여 날씨와 미세먼지 예보 데이터를 입력하여 주차 가용면 수를 예측(추출)하는 AI 학습 모델을 개발하였다.



〈그림 3-1〉 CityDataHub를 활용한 주차 혼잡도 예측 PoC 개념도



## ● 데이터허브에 내재화된 분석 모듈 활용

- CityDataHub가 제공하는 분석 모듈이 기능상 상용 서비스 대비 제한적이긴 하지만 데이터 코어 모듈과의 연동을 기본으로 제공하는 내부 서비스라는 측면에서 장점이 있으며 주차 혼잡도 예측 PoC에 이를 활용하였다.
- 데이터 코어에 수집된 데이터로 마트 데이터를 생성하여 학습 모델을 생성하고, 이를 일정 시간마다 실행하여 혼잡도 예측 데이터를 데이터 코어에 저장하였다. 혼잡도 예측치는 주차 서비스 애플리케이션으로 노출하여 사용자가 앞으로의 주차 혼잡도를 사전에 확인하고 주차 가능한 주차장을 선택할 수 있게 하였다.

## D 코로나19 역학조사지원시스템

### ● 역학조사 데이터 통합관리 및 분석 시스템

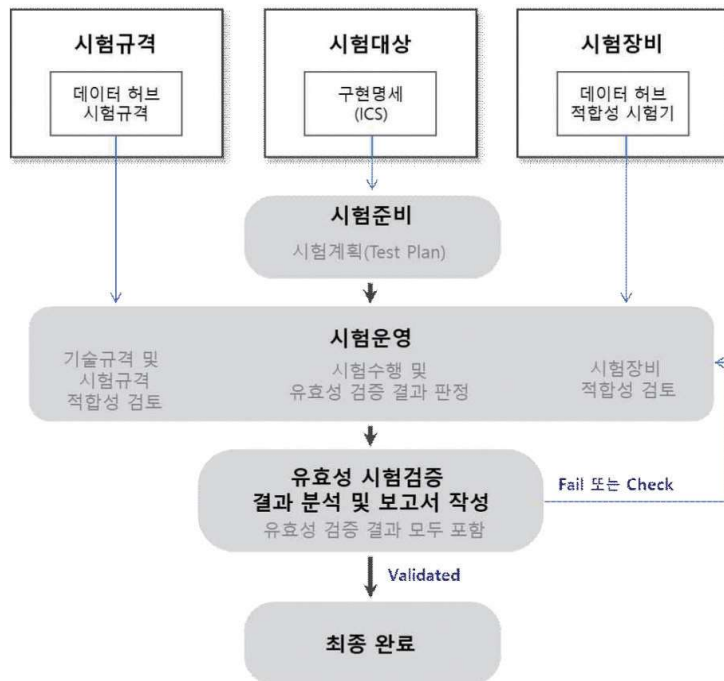
- 2020년 초, 여전히 코로나19 역학조사는 역학조사관의 확진자 인터뷰, 공문 기반 기초 데이터 확보 및 수기 데이터 분석 등으로 인해 가파르게 증가하는 확진자 역학조사에 어려움이 증대되고 있었다.
- 당시 프로토타입 개발을 마친 CityDataHub를 활용해 외부 시스템 연동을 구축하여 기초 데이터를 빠르게 수집하고, 수집된 데이터를 정제하고 분석하여 역학조사관이 활용할 수 있는 확진자 동선 정보를 제공할 수 있었다.
- 이동통신사 기지국의 확진자 위치 이력 데이터는 1명당 일일 수만 건의 데이터가 수집되고, 수집한 데이터는 오차를 포함하고 있어 이동 속도, 동선 분포 등으로 정제하여 기초 동선 파악에 활용하였다. 또한 신용카드사의 카드 결제 이력과 QR 출입 이력 데이터를 수집하여 동선 파악에 활용하였다.
- 생성된 다수의 확진자의 동선 정보는 확진자 동선 교차점 분석을 통해 주요 감염 경로 및 핫스팟을 분석하여 감염위험지역을 파악하여 감염 확산 방지에 활용하고 있다.
- 2020년 4월부터 운영 중인 코로나19 역학조사지원시스템은 신속한 데이터 기반 역학조사 지원 업무에 활용되고 있으며 법무부 출입국 기록 등 다양한 시스템 연동 추가 개발을 통해 심층 역학조사 지원시스템으로 고도화를 추진 중이다.
- 보다 상세한 데이터허브 기반 코로나19 역학조사지원시스템 개발 내용은 별편을 통해 확인할 수 있다.



〈그림 3-2〉 CityDataHub 활용 코로나19 역학조사지원시스템 개념도

## 2 | 시험검증

### 시험검증 절차



〈그림 3-3〉 스마트시티 데이터허브 시험검증 절차

## ● 데이터 코어 모듈 적합성 시험

- 스마트시티 데이터허브의 시험검증은 데이터 코어 모듈의 적합성 시험으로 진행하였다. 데이터허브의 다양한 기능 모듈 중에 필수 모듈에 해당하며 스마트시티 혁신성장동력프로젝트 사업에서 자체적으로 정의한 API 명세가 아닌 외부 표준화 기구에서 개발한 인터페이스를 지원하고 있어 데이터 코어 모듈을 적합성 시험 대상으로 선정하였다.
- 적합성(Conformance) 시험이란 규격에 맞게 정확하게 구현하였는지 검증하는 절차로 검증된 시험기 소프트웨어를 통해 사전에 정의한 시험 항목을 실행하여 항목별 성공 및 실패 결과를 확인한다. 데이터 코어 모듈의 적합성 시험은 NGS-LD API 표준에 대한 적합성과 데이터 모델의 적합성 시험으로 구성된다.
- ETSI NGS-LD API는 적합성 시험규격(Test Purpose Description, ETSI GS CIM 013)을 정의하고 있으며 프로토타입 버전인 Robot 프레임워크 기반 시험기를 공개하고 있다.

## ● 시험규격 및 시험기 검증 기반 실증도시 데이터허브 검증

- 2022년 기준 ETSI의 시험규격뿐만 아니라 시험기는 초기 버전이라 CityDataHub의 데이터 코어 모듈 구현을 시험하고 결과를 분석하면서 표준 구현 이슈, 시험규격 이슈 및 시험기 이슈로 나누어 결과를 분석하고 1핵심의 오픈소스 구현물뿐만 아니라 2, 3핵심의 스마트시티 데이터허브를 검증할 수 있는 시험규격과 시험기를 갖추게 되었다.
- 검증 완료한 NGS-LD 규격 및 시험기 이슈는 ETSI 표준회의에 참석하여 이를 공유하고 수정하는 표준화 업무도 수행하였다.
- 스마트시티 혁신성장동력프로젝트 사업 내에서 수행하는 2, 3핵심 데이터허브 검증뿐만 아니라 다른 지자체의 데이터허브 검증에도 적합성 시험검증 체계를 바탕으로 인증체계가 구축될 예정이다.

## 1 | 도시 데이터허브 확산방안

## 1.1 정부 지원 사업 연계

## ● 데이터허브 보급 확산사업

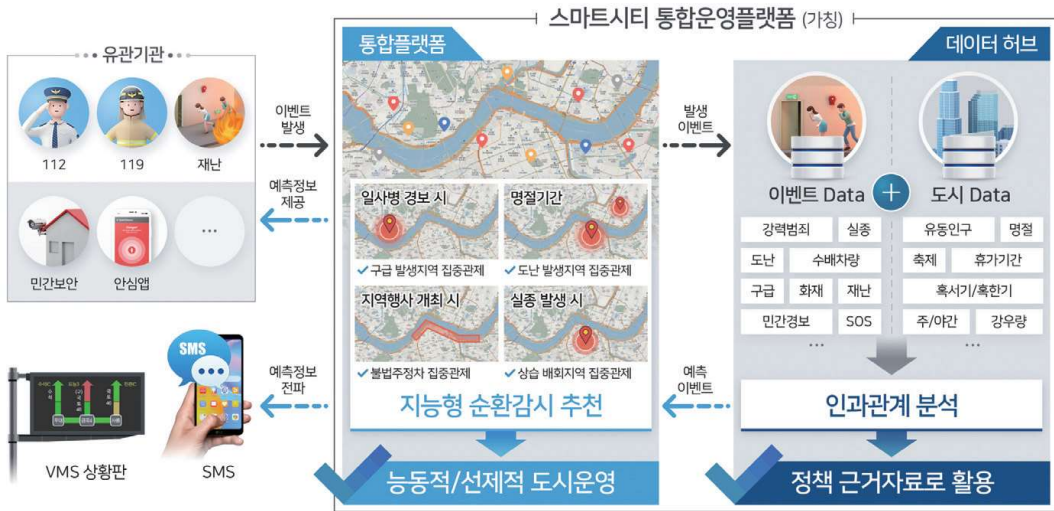
- 스마트시티 혁신성장동력프로젝트 사업을 과학기술정보통신부와 공동 지원한 국토교통부는 2022년 추가 사업을 통해 스마트시티 데이터허브 확산을 지원하고 있다. 스마트시티 지역거점 조성사업과 데이터허브 보급사업을 통해 대구와 시흥 이외의 지자체 실정에 맞는 스마트시티 데이터허브를 구축하여 다양한 도시 서비스 제공을 목표로 한다.

## ● 통합 플랫폼 연계 활용 모델 제시

- 2022년 기준 108개 지자체에 구축된 스마트시티(U-City) 통합플랫폼은 5대 연계 서비스와 지자체별 S-서비스를 제공하고 있다. 부천시는 국토교통부의 스마트 캠퍼스 챌린지 사업을 통해 기존 운영 중인 통합플랫폼과 스마트시티 데이터허브 시스템 연동을 구축하고 두 시스템을 연계 활용하는 방안을 구현 검증하였다.
- 통합플랫폼으로 수집되는 도시 이벤트 데이터를 데이터허브에 축적하고, 다양한 분석에 활용하여 데이터허브를 이용한 도시운영 모델을 제시하였다.
- 통합플랫폼 외에도 지자체에 기존 구축된 빅데이터 플랫폼과의 연계 활용 모델을 제시하고 검증하여 스마트시티 데이터허브의 지자체 확산방안을 도출할 예정이다.

## ● 지자체를 위한 데이터허브 가이드 제공

- 스마트시티 혁신성장동력프로젝트 사업에서 개발된 스마트시티 데이터허브와 데이터허브 활용 서비스 개발 실증사례를 제공한다.
- 데이터허브 확산, 활용을 위한 데이터 거버넌스 구성과 관련하여 데이터 관리 가이드, 품질관리 가이드 및 보안 가이드를 제공한다.
- 향후 지자체 보급과 솔루션 개발에서 활용할 수 있도록 데이터허브 구축 및 운영 가이드 제시가 필요하다.



〈그림 4-1〉 스마트시티 데이터허브와 통합 플랫폼 연계 활용 방안

## 1 CityDataHub 보급

### ● 표준 및 오픈소스 생태계 기반 확산

- 스마트시티 데이터허브에 구성에 필수적인 데이터 코어 모듈에 사용되는 ETSI의 NGSI-LD는 국내 데이터허브 뿐만 아니라 유럽의 스마트시티 프로젝트에도 활용되고 있다.
- 대표적인 IoT 표준인 oneM2M과도 연동이 가능한 어댑터를 제공하며 연계 활용 사례를 확보하고 있어서 표준을 이용한 연동 기술 개발, 활용 확산 및 표준 기반 생태계 조성에 활용할 수 있다.
- CityDataHub 오픈소스는 소스 코드뿐만 아니라 개발자 지원을 위한 설치와 사용을 위한 가이드를 온라인으로 제공한다.

- 오픈소스 연계에 있어서는 유럽과 남미 중심으로 활용도가 높은 FIWARE 오픈소스 재단에서도 NGSI-LD 지원 솔루션을 제공하고 있어 해외 표준화 그리고 오픈소스 기관과의 협력을 통해 활용 확산을 추진할 수 있다.



〈그림 4-2〉 CityDataHub 활용 확산방안

## 인증 프로그램

### 시험검증 역량기반 인증 프로그램 운영

- 대구와 시흥 외의 지자체에 데이터허브가 확산되어 운영될 때 해당 지자체는 데이터허브에 대한 인증 프로그램을 요구할 것으로 예상된다. 상호운용성에 기반한 서비스의 이식성뿐만 아니라 구축한 시스템에 대한 유지보수 측면에서도 시스템 인증은 필수적이기 때문이다.
- 앞서 설명한 데이터허브 시험검증 체계는 스마트시티 혁신성장동력프로젝트 사업단 내에서 실증 완료되었다. 이를 바탕으로 데이터허브 인증 프로그램이 설립되고 운영될 것으로 기대된다. 특히 CityDataHub와 같은 오픈소스 솔루션이 인증을 수여하면 이를 활용한 다양한 솔루션이 추가 인증을 통해 기술 생태계가 확대되는 계기가 될 것으로 보인다.

## 1 도시 데이터 모델 표준화

### ● 데이터허브 활용 서비스 확대를 위한 데이터 모델 거버넌스

- 스마트시티 데이터허브는 외부로부터 수집한 데이터를 공통 데이터 모델로 변환 저장하여 데이터를 분석하고 활용하는 데 필요한 가공 비용을 줄이고, 도메인 간 융복합 활용을 용이하게 하는 장점을 가지고 있다.
- 데이터허브를 활용한 도시 서비스 활성화를 위해서는 비즈니스 모델 도출도 중요하지만, 사용자가 공통 또는 표준 데이터 모델을 활용하는 것이 중요하다. 인터페이스는 같지만 데이터 모델이 상이하다면 서비스 간에 이식성도 없고 이를 다시 변환해서 사용하는 불편함이 있다.
- 다만 도시 데이터는 그 분야가 방대하고 분야별로 표준 모델이 존재하기도, 부재하기도 하다. 표준 데이터 모델의 경우 NGSI-LD 인터페이스에 맞게 준용하여 활용할 필요가 있고, 그렇지 않은 서비스 데이터는 표준화 단체가 아닐지라도 스마트시티 그리고 데이터허브 이해관계자가 모여 공통 데이터 모델을 정의하고 활용할 필요가 있다.
- 데이터 모델은 기술적으로 옳고 그름의 문제가 아닌 비즈니스 측면의 이슈이다. 따라서 데이터허브에서 사용할 데이터 모델을 제안하고 논의하고 이를 제공할 수 있는 체계가 필요하다.
- 유럽의 경우에는 스마트 데이터 모델 협의체를 통해 공통 데이터 모델 확산을 위한 활동을 스마트시티 프로젝트를 중심으로 수행하고 있다. 앞서 설명한 바와 같이 데이터 모델링은 서비스 혹은 비즈니스 측면에 강해 국내와 해외의 모델 정의가 다를 수 있으며, 국내 데이터허브 활용 기반 공통 데이터 모델 확산이 더 우선순위로 추진되어야 하기에 유럽의 사례와 비슷한 데이터허브 공통 데이터 모델 거버넌스 모델이 2022년에 시범적으로 운영될 예정이다.

## 1 | 맺음말

## ● 스마트시티 데이터허브

- 데이터허브는 기획부터 실증까지 방대한 사업을 통해 개발된 시스템으로 설계, 개발 및 검증에 있어 여러 세부과제와 기관과의 협력이 필요한 고난도 과제이다.
- 데이터허브 모델 및 아키텍처 검증을 위해 프로토타입 기반으로 기술검증(PoC)을 수행하였고 개발된 데이터허브를 2개 도시에 성공적으로 적용, 실증하였다.

## ● CityDataHub 오픈소스

- 2022년 4월 공개한 CityDataHub는 데이터허브 개발의 한 단락을 마무리함과 동시에 새로운 장을 여는 계기가 되었다. 특히 오픈소스를 활용해 확장 개발된 형태로 시흥시에 적용되었고, 부천시에는 통합플랫폼 연계 모델 검증을 위해 CityDataHub가 사용되었다. 해외 실증사례로서 스페인의 산탄데르(Santander)에 구축된 것뿐만 아니라 다른 스마트시티 연계 사업에도 CityDataHub를 활용한 제품화가 추진되고 있다.
- 데이터허브 오픈소스를 공개 이후에도 국내외 데이터허브 확산, 활용, 유지보수를 위해 지속적인 리소스를 투입할 예정이다.
- 개발의 끝이 아닌 유지보수와 활용 확대를 위한 지속적인 리소스 투입을 통해 국내외 데이터허브 확산을 추진할 예정이다.



## 2 | 향후 과제

### ● 스마트시티와 링크드 데이터

- 스마트시티는 다양한 분야로 이루어져 있고 그 안에서 수많은 데이터가 융복합되어 활용될 수 있는 잠재성이 있다. 이러한 관점에서 링크드 데이터 또한 큰 활용 가능성을 지니고 있으며 기술적으로 구현한 링크드 데이터를 활용한 데이터 구축 및 인터페이스 구현을 떠나 진정한 활용 서비스 발굴 및 개발검증 또한 추진해볼 만하다.

### ● 데이터허브 활용 표준 개발

- 데이터허브는 국토교통부의 스마트시티 국제 표준화 과제와 연계하여 추가적인 표준 기술을 2023년까지 개발하고 있으며 NGSI-LD 표준을 제정한 ETSI에서 NGSI-LD를 적용해 데이터허브를 개발 검증한 사례 외에도 데이터허브를 활용하기 위한 표준을 개발하고 있다.
- IoT 표준과 데이터허브의 연계는 oneM2M과 NGSI-LD 표준 인터페이스 간 연동 기술을 정의하는 것으로 기존 구축된 oneM2M 시스템에 기존 표준에서 제공하는 인터페이스를 활용하여 NGSI-LD를 지원하는 데이터허브와 데이터 연동을 구축하기 위한 표준을 개발한다.
- 데이터허브 페더레이션 또한 표준을 개발 중이며 이는 기존 NGSI-LD 표준에서 정의한 페더레이션 아키텍처 개념을 실제 구현하기 위한 방안과 절차를 제시한다.

### ● 스마트시티 데이터허브 고도화 추진

- 2022년까지 개발과 실증을 마쳐 구축 확산을 진행 중인 스마트시티 데이터허브 1.0은 향후 지속적인 기술 고도화 및 활용 도시 서비스의 지속적인 실증 추진이 필요하다.
- 데이터 기반으로 인접 도시와 상황 연결된 스마트시티로, 행정구역 경계를 넘어 도시 간 자원 및 서비스 공유를 통해 효율적이고 최적화된 도시 생활환경을 제공하는 메타시티를 통해 구현될 수 있는 서비스 발굴을 주요 목표로 설정할 수 있다.

## 참고문헌

---

- 스마트시티 데이터 허브 (<http://citydatahub.kr>)
- 스마트 데이터 모델 협의체 (<http://smartdatamodels.org>)