

동적주제도 시스템의 통합 관리를 위한 DB 구축 연구

홍종유*, 김성환
(주)제이비티

DB Construction research for Integrated management of Dynamic Thematic Map System

Jongyoo Hong*, Sunghwan Kim
JB Technology Co., Ltd.

요약 동적주제도 시스템은 고정/이동 플랫폼을 통해 여러 유형의 센서로부터 다양하고 방대한 정보를 수집하고 저장한다. 주요 데이터는 수집원·객체추출 데이터, 주제별동적정보·이동맥락정보·동적주제도 데이터셋으로 구성된다. 영상정보, 센싱정보, 분석정보 등 다양한 데이터를 포함하고 있으며, 다양한 참조 관계가 필요하기 때문에 각 데이터의 접근성을 보장하면서 정보의 통합이 가능한 관리 체계가 필요하다. 이를 위해 데이터의 흐름, 소유자, 주요 공정을 식별하고 데이터의 세부 내용을 파악했다. 정형·비정형의 데이터가 포함되어 있고, MongoDB 및 Postgres 등을 사용하고 있었다. 이런 정보를 관리하기 위한 방법으로 데이터 카탈로그를 분석했다. 데이터 카탈로그는 데이터레이크에 주로 적용되어 정형·비정형 데이터를 모두 수용할 수 있고, 비즈니스 메타데이터까지 포함하기 때문에 동적주제도 시스템의 DB를 통합 관리하기에 매우 적합하다. 데이터 카탈로그는 비즈니스·테크니컬·운영 메타데이터로 구성된다. 비즈니스 영역은 비즈니스 관점에서 모든 사용자가 동일한 명칭과 개념을 사용하도록 지원한다. 테크니컬 영역은 기존의 메타데이터와 유사하게 테이블, 컬럼 등의 정보를 제공한다. 운영 영역은 보고서, 대시보드 등 데이터 활용정보를 제공하며 데이터 리니지를 포함하기도 한다. 본 연구에서는 동적주제도 DB 관리 시스템의 기능을 데이터 카탈로그 기반으로 구성했다. 사용자는 카탈로그를 검색하고 실행데이터를 조회할 수 있다. 쿼레이션 기능과 DB 표준화 관리 기능도 통합해 제공한다. 관리자는 사용자의 이용 현황을 모니터링하고 시스템을 관리한다. 이와 같이 데이터 카탈로그를 이용하면 다양한 종류의 데이터와 여러 계층의 사용자를 위한 정보를 효과적으로 관리할 수 있다. 동적주제도 시스템이 데이터 카탈로그 기술을 채택하면 효율적인 정보 관리와 사용자 편의성을 동시에 제공할 수 있을 것으로 기대된다.

Abstract A dynamic thematic map system collects and stores a variety of extensive information from various types of sensors through fixed/mobile platforms. The main data consist of collection sources and object extractions, dynamic information by subject, heterogeneous context information, and a dynamic subject map dataset. Since it contains various data, such as image information, sensing information, and analysis information, and requires various reference relationships, a management system that can integrate information while ensuring accessibility to the data is required. To do this, we identified data flows, owners, and key processes, and identified details of the data. Structured and unstructured data were included, and MongoDB and Postgres were used. We reviewed data catalogs as a way to manage this information. Since data catalogs are mainly applied to data lakes, they can accommodate both structured and unstructured data, and can even include business metadata, so they are very suitable for integrated management of a DB for the Dynamic Subject Map System. The data catalog consists of business, technical, and operational metadata. The business domain supports the use of the same nomenclature and concepts from a business perspective. Similar to existing metadata, the technical area provides tables and columns. The operational area provides data utilization through reports and dashboards, and includes data lineage. In this study, the functions for the Dynamic Subject Map System's DB management are configured based on the data catalog. Users can search catalogs and view real data. A curation function and a DB standardization management function are integrated and provided. The administrator monitors the status of users and manages the system. In this way, by using a data catalog, various types of data and information for users of various classes can be effectively managed. If the Dynamic Subject Map System adopts data catalog technology, it is expected to provide efficient information management and user convenience.

Keywords : Dynamin Thematic Map, DBMS, Data Catalog, Metadata, Unstructured Data

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었음(과제번호 RS-2022-00143782).

*Corresponding Author : Jongyoo Hong(JB Technology Co., Ltd.)

email: turk182@ejbt.co.kr

Received March 21, 2023

Revised April 21, 2023

Accepted May 12, 2023

Published May 31, 2023

1. 서론

4차 산업혁명 시대로 접어들면서 사회간접시설(SOC)의 패러다임이 생산 기반 시설에서 국민생활 편의시설로 확장되고 있다. 또한, 디지털 뉴딜사업의 일환으로 SOC의 디지털 전환이 진행되면서 SOC의 정적 정보 뿐 아니라 이용 현황, 품질 상태, 외부 환경요인 등 다양한 동적 정보의 구축이 요구되고 있다. 도심의 모빌리티는 자율주행차, 퍼스널모빌리티, 드론 등으로 다양화하고 있고, 물류 이동의 폭증으로 인해 다양한 도시 안전문제도 발생이 늘어나고 있다. 이러한 변화로 인해 신속하고 정확한 동적 공간정보 인식 및 정보갱신을 제공할 수 있는 동적 객체 모니터링의 필요성이 커지고 있다.

현재 운전/보행자의 제보, CCTV, 프로브카 등으로 지상 동적정보를 모니터링하고 있으나, 갭신 주기 및 공간적 정보 인식 범위의 한계에 따라 위급상황 발생 시 골든타임을 놓치는 경우가 많은 실정이다. 이에 동적 공간정보의 상황 변화를 준 실시간으로 인식하고, 맞춤형 정밀도와 해상도의 정보를 제공하는 기술 개발이 필요하다. 다양한 도시 주제도 구축을 위해서는 도시 내 운용 가능한 여러 수단을 통해 데이터를 최대한 확보하고 정확도 향상을 위한 기술을 개발해 적용하는 것이 필수적이다[1].

이런 배경으로 도시 내 이동 객체와 상태 변화 등의 생활 SOC 동적 공간정보를 고정/이동 플랫폼으로 지속적으로 준 실시간 감지·추적해 다양한 사회문제를 해결하기 위한 동적정보 주제도 구축 기술이 개발되고 있다. 고정플랫폼은 CCTV, 검지기, 통신AP, IoT 센서 등 위치가 고정되어 있는 센서 기반의 데이터 수집을 위한 플랫폼을 의미하며, 이동플랫폼은 무인항공기(ex. 드론)와 같이 객체를 추적하거나 구역을 순찰할 수 있는 이동형 센서 기반의 데이터 플랫폼이다. 각 플랫폼에서 수집한 정보는 정제·추출·식별·분석의 과정을 거쳐 동적주제도로 제공된다.

데이터의 측면에서는 고정/이동 플랫폼을 통해 여러 가지 유형의 센서로부터 다양하고 방대한 정보를 수집하고 저장하게 된다. 각 모듈은 이 정보에 접근해 객체 식별, 분석 등의 고유 기능을 수행하고 최종적으로 동적주제도를 제공하게 된다. 이를 위해서는 수집된 정보들을 체계적으로 통합 관리할 수 있는 데이터베이스가 필요하다.

따라서 본 연구에서는 동적주제도 구축 목표 시스템에 대한 이해를 바탕으로, 시스템에서 다루어지는 정보

들을 효율적으로 관리할 수 있는 동적 통합정보 데이터베이스의 구축 방안을 수립하고, 시스템 개발에 적용할 수 있는 DB 설계를 수행한다. 구축 방안으로 제시하고 있는 데이터 카탈로그는 빅데이터의 성장과 맞물려 그 필요성과 역할이 더욱 주목받고 있다. 이런 배경으로 빅데이터에 초점을 맞춘 연구와 사례가 주를 이루지만, 그것이 데이터 카탈로그의 활용 범위를 제한하는 것은 아니다. 본 연구에서는 데이터의 통합 관리를 위해 데이터 카탈로그가 제공하는 다양한 정보와 사용자 관점에 주목하고자 한다.

본론에서는 동적주제도 시스템의 주요 기능과 구성을 소개하고, 데이터의 유형과 속성에 대한 분석 결과를 통해 관계형 데이터베이스 중심 설계의 한계를 설명한다. 이어서, 이의 대안으로 데이터 카탈로그를 소개하고, 동적주제도 시스템을 위한 적용 방안과 기대 효과를 제시한다.

2. 본론

2.1 목표 시스템

2.1.1 시스템 개요

동적주제도 시스템은 도시 내의 보도/차도/보차혼용도에서 이동하는 객체와 상태 변화 등의 생활 SOC 동적 공간정보를 고정/이동 플랫폼으로 지속적으로 준 실시간 감지하여 다양한 사회문제를 해결하기 위한 시스템이다. Fig. 1은 동적시스템의 개념을 설명한 그림이다.

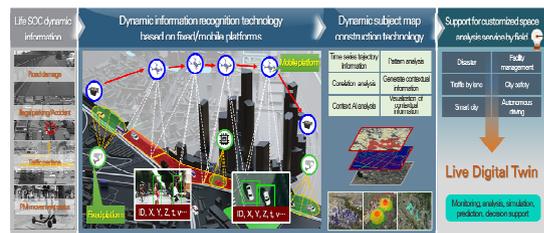


Fig. 1. Concept of Dynamic Thematic Map System

시스템의 정보 수집 대상은 Fig. 1에서 보는 바와 같이 이동형 객체와 고정형 객체로 구분할 수 있다. 이동형 객체는 차량, 퍼스널모빌리티 등이 해당되고 고정형 객체는 도로 상태나 시설물 변화 정보 등이 해당된다. 이러한 정보가 고정 플랫폼과 이동 플랫폼을 통해 수집되고 인공지능 빅데이터 분석, 이동 맥락정보 분석 등을 거친

정보는 다양한 동적주제도로 가공되어 유관기관, 지자체, 기업, 시민 등의 수요자에게 제공된다.

이러한 서비스를 제공하기 위한 동적주제도 시스템의 구성도는 Fig. 2와 같다.

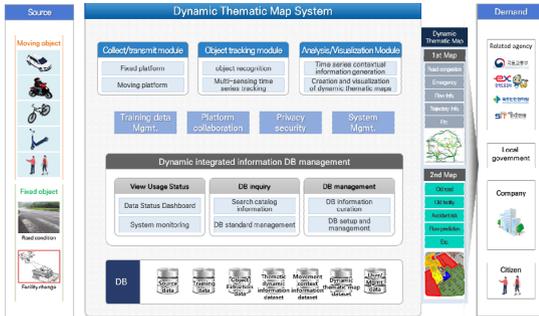


Fig. 2. Diagram of Dynamic Thematic Map System

시스템의 구축 기술은 크게 세 가지로 나누어 볼 수 있는데, 고정 플랫폼 기반 동적정보 수집 기술, 이동 플랫폼 기반 동적정보 수집 기술, 동적정보 분석·예측 및 표현 기술이다.

고정 플랫폼에서는 CCTV, WiFi, 블루투스 등의 고정되어 있는 센서장비와 연계하여 객체 검출(Object Detection)기술, 객체 추적(Multi Camera Tracking) 기술을 활용한 실시간 이동체 검출/추적 기술과 도시 내 이동체의 연속적인 시계열 위치정보를 지속적으로 추적하는 기술을 개발한다.

이동 플랫폼에서는 동적정보 수집용 드론, 이동플랫폼 운용 시스템 구축, 동적정보 인식·통합 알고리즘 개발 및 고정-이동플랫폼 협업형 연속적 객체 인식 기술을 개발한다.

동적정보 분석·예측 및 표현에서는 고정·이동 플랫폼에서 수집되는 객체 단위 이동체 정보와 정적데이터를 연계한 이동체의 시계열 패턴 및 맥락정보 생성 기술, 이동체의 분석·예측을 위한 맥락 기반 AI적용 기술을 개발한다. 그리고, 동적정보를 활용한 사용자 맞춤형 동적주제도 구축 및 갱신 기술로 동적주제도를 시각화한다.

2.1.2 데이터

동적주제도 시스템은 다양한 데이터를 관리하게 된다. 고정·이동 플랫폼에서 수집된 정보를 비롯해 분석 결과도 여기에 포함된다. Table 1은 시스템에 포함된 주요 데이터의 목록이다.

Table 1. Key data list

No.	Data	Detail
1	Source data	Video, sensing information
2	Object extraction data	Location, time, status information
3	Thematic dynamic information dataset	Moving object tracking information
4	Movement context information dataset	Moving object analysis-prediction information
5	Dynamic thematic map dataset	Dynamic thematic map information

수집원데이터는 고정플랫폼 및 이동플랫폼으로부터 수집되는 원천 정보이다. 고정플랫폼에서 수집되는 정보는 영상정보와 비영상정보가 있는데, 영상정보는 CCTV를 이용해 수집하고 비영상정보는 무선AP와 각종 센서가 포함된다. 이동플랫폼에서는 드론의 카메라를 이용한 영상정보를 수집한다. 영상과 센서의 원천 정보를 포함하는 수집원데이터는 대용량 빅데이터의 특성을 가지고 있다.

객체추출데이터는 수집원데이터의 객체 정보이다. 수집원데이터를 분석해서 객체를 식별하게 되고, 식별된 객체의 위치, 시간, 상태 정보를 추출하게 된다. 이 정보가 객체추출데이터에 저장된다.

주제별동적정보데이터셋은 객체의 연속 정보이다. 동적주제도는 변화하는 정보를 대상으로 하기 때문에 식별된 객체는 이동체로 볼 수 있다. 이 단계에서 수집된 객체정보에서 이동체의 추적 정보를 생성한다. 객체정보의 연속 해석, 위치 추적 등의 방법으로 도출된 이동체의 추적정보가 주제별동적정보데이터셋에 저장된다.

이동맥락정보데이터셋은 이동체의 분석·예측 정보이다. 지금까지 살펴본 데이터가 이동체(객체)를 식별하고 해석한 결과였다면, 이동맥락정보는 이동체의 시계열 패턴과 맥락정보를 포함한다. 맥락정보 기반의 이동체 분석 및 예측을 위해서 AI기술을 활용하며 그 결과물이 이동맥락정보데이터셋에 저장된다.

동적주제도에데이터셋은 동적주제도를 생성하기 위한 정보이다. 목표시스템에서 제공하게 될 동적주제도가 구체화되면, 각 주제도 생성에 필요한 정보로 구성된 정보가 만들어지고 동적주제도에데이터셋에 저장된다. 이를 위해서 주제별동적정보데이터나 이동맥락정보데이터가 사용될 수 있고, 별도의 정보들도 활용될 수 있다.

이상과 같이 동적주제도 시스템은 영상정보, 센싱정보, 분석정보 등 다양한 데이터를 포함하고 있으며, 시스

템 내에서는 분석 모델에 따라 다양한 참조 관계가 요구되기 때문에 각 데이터의 접근성을 보장하면서 통합할 수 있는 관리 체계가 필요하다.

2.2 데이터 분석 및 모델링

2.2.1 데이터 흐름 분석

데이터베이스를 구축하기 위해서는 각 데이터의 내용과 특성 정보가 필요했기 때문에, 우선 데이터의 흐름을 파악하고 소유자 및 주요 공정을 식별했다.

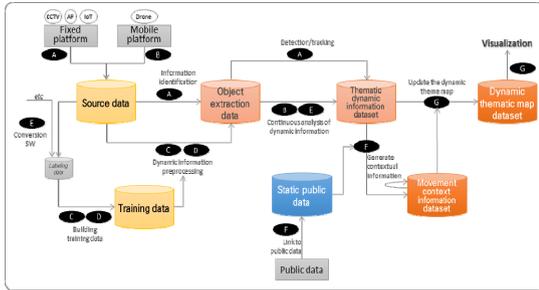


Fig. 3. Data Flow Chart

Fig. 3에서 보는 바와 같이 데이터의 흐름은 고정플랫폼과 이동플랫폼에서 시작된다. 검은색 원의 부호는 각 연구 수행 기관인데, 연구 현황에 대한 추가 조사를 진행하기 위해 파악했다.

고정플랫폼과 이동플랫폼에서 수집된 정보는 「수집된 데이터」에 저장된다. 저장된 데이터로부터 객체를 식별하고, 그 정보는 「객체추출데이터」에 저장된다. 이 과정에서 딥러닝과 영상분석 알고리즘이 적용된다. 그 다음으로 위치감지와 추적, 동적정보 연속해석에 의해 이동체 추적정보를 생성하고 「주제별동적정보데이터셋」에 저장한다. 계속해서 정적공공정보와 이동체 추적정보를 이용한 맥락정보 기반의 분석·예측 정보가 「이동체맥락정보데이터셋」에 저장된다. 마지막의 「동적주제도데이터셋」은 이동체 추적정보와 이동체맥락정보를 이용해 생성된다.

여기서 설명한 데이터의 흐름은 기본이 되는 흐름이며 모든 참조 관계를 설명하는 것은 아니다. 식별, 분석, 예측 연구의 확장성을 위해 모든 자료의 접근 권한과 방법이 제공될 필요가 있다.

각 데이터에 대한 좀 더 상세한 내용을 파악하기 위해, 식별된 내용을 바탕으로 각 연구 기관의 데이터 현황 및 설계 자료를 수집했다.

2.2.2 데이터 모델링 및 한계

수집한 자료를 토대로 데이터의 상세한 내용을 파악했다. 데이터는 정형, 비정형 유형이 모두 포함되었으며, 그에 따라 DBMS도 MongoDB, Postgres 등으로 나뉘었다. 아직 구체화되지 않은 부분도 있지만, 수집된 자료에서 공통된 내용을 Table 2와 같이 정리할 수 있었다.

Table 2. Common information of construction data

No.	Category	Common information
1	Object	Classification of objects, detailed format
2	Location	3D location information of objects
3	Sensor	Information collection sensor type, unique number
4	Time	information acquisition time

첫 번째, 객체(이동체)에 대한 정보가 포함된다. 시스템에서 식별하는 객체는 사전에 정의되며, 식별된 객체의 분류와 세부 형식 정보가 포함된다.

두 번째, 객체에 대한 위치정보이다. 객체에 대한 위치 정보는 고도를 포함하는 3차원 형식으로 저장되고, 기본적으로 경위도 좌표를 사용한다.

세 번째, 정보를 수집한 센서에 대한 정보이다. 센서의 유형과 고유정보를 들 수 있다. 고유정보에는 일련번호와 세부 사양이 포함될 수 있고, 고정된 센서인 경우에는 위치 정보가 추가될 수 있다.

네 번째, 정보를 취득한 시간 정보이다. 동적주제도에서는 이동체의 정보를 제공하기 때문에 시간은 구간 형태가 될 수도 있다.

수집된 정보의 분석으로 도출한 공통 정보 항목을 정리해 Table 3과 같이 공통컬럼을 설계했다.

Table 3. Common column design

No.	Entity	Attribute	Scope
1	Date and time	datetime	Public
2	sensor type code	string	Public
3	sensor identification number	string	Public
4	Object code (large category)	string	Public
5	Object code (medium category)	string	Public
6	Object code (small category)	string	Public
7	latitude	float	Public
8	longitude	float	Public
9	altitude	float	Public

공통컬럼은 데이터베이스 설계 시 Fig. 4와 같이 적용하는 방안이 검토되었다. 테이블의 전반부에 공통 컬럼을 삽입하고, 그 이후 영역에 업무 컬럼을 배치하는 방식이다.

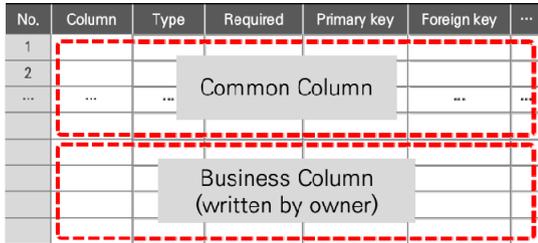


Fig. 4. Common column application method (example)

하지만 이러한 공통컬럼은 MongoDB와 같은 NoSQL DB에 적용하기에는 적합하지 않다. 반복적인 정보의 처리로 어플리케이션 특성에 따라 상당한 성능 저하를 가져올 수도 있다. 관계형 데이터베이스(RDB, Relational DataBase)에서는 활용될 수도 있겠지만, RDB와 NoSQL을 모두 포함해 관리할 수 있는 체계를 도입할 필요가 있다.

2.3 데이터 카탈로그

2.3.1 데이터 카탈로그 관리 기법

데이터 카탈로그는 조직 내에서 구성된 데이터 자산 인벤토리이다. 데이터 카탈로그는 다른 데이터에 대한 기본 정보를 제공하고 그에 대한 설명을 포함하는 메타데이터 역할을 하는데, 이러한 정보를 이용해서 사용자는 필요로 하는 데이터를 빠르게 검색하고 쉽게 활용할 수 있다[2].

데이터 카탈로그는 기존의 메타데이터를 확장한 것으로 이해할 수 있는데, 빅데이터와 같이 비즈니스 영역에서 관리하는 데이터가 기하급수적으로 증가하면서 그 효용가치도 커지고 있다. 데이터웨어하우스, 데이터레이크, 데이터카탈로그의 흐름을 이해하기 위해 Table 4에 정의와 용도를 간단히 정리했다.

데이터웨어하우스는 다양한 운영시스템에서 생산된 데이터를 통합해 관리하기 위한 데이터 접근 전략이었다. 하지만 이런 방식은 빅데이터 시대의 비정형 데이터에는 적합하지 않았다. 변환, 통합의 비용이 컸기 때문에 수집한 데이터는 일단 그대로 보관하고 필요할 때 데이터를 가공하는 데이터레이크 방식이 등장했다. 비정형의

Table 4. Data warehouse, data lake, data catalog

Category	Definition	Usage
data warehouse	Repository of data extracted, transformed and consolidated from various systems	Storage of Structured Data
data lake	A repository that temporarily stores collected data	Storage of unstructured data
data catalog	Information about an organization's data assets	Support data search and utilization

빅데이터 처리가 가능하기는 했지만, 문제도 있었다. 방대하고 다양한 정보가 쌓이면서 데이터의 활용이 쉽지 않아 데이터 늪(data swamp)이라고 불리기도 했다[3]. 반면에 데이터의 중요성은 계속 증가했기 때문에 데이터 자산을 관리할 필요성은 커졌고, 데이터 카탈로그의 등장은 자연스러운 수순이었다.

데이터 카탈로그는 기존 메타데이터의 확장으로 볼 수 있으며[4], 크게 세 가지의 정보를 포함한다. 테크니컬/비즈니스/운영 관점의 정보이며[5], 기존의 메타데이터는 테크니컬 메타데이터에 해당한다. 운영 메타데이터는 종종 활용, 프로세스 관점으로 분류하기도 한다. 데이터 계보(Lineage)는 테크니컬 또는 프로세스로 분류하기도 하는데, 여기서는 운영 관점으로 분류했다. 데이터 카탈로그의 정보는 Table 5와 같다[6].

Table 5. Organization of the data catalog

Category	Content	Example
Business Metadata	<ul style="list-style-type: none"> Defining data from a business perspective Help business users use the same nomenclature and concepts. 	<ul style="list-style-type: none"> Business Term Definition Data Ownership/Stewardship Data Compliance Factors
Technical Metadata	<ul style="list-style-type: none"> information about data Information from the point of view of data design and construction 	<ul style="list-style-type: none"> table, column Reference, connection
Operational Metadata	<ul style="list-style-type: none"> Contents on the flow of data creation/movement/transformation (Data Lineage) Data utilization information such as reports and dashboards 	<ul style="list-style-type: none"> Data movement path Data conversion process Data usage status of the system

Table 6은 메타데이터 관리시스템과 데이터 카탈로그를 항목별로 비교한 표이다. 관리항목에 대한 내용은 위와 다른 부분이 있지만, 표는 원문 내용을 그대로 인용했다[7].

Table 6. Comparison of Metadata Management Systems and Data Catalog(Source : Seonwoong Yoon, Make Data Catalog, 2021, pp.39)

Item	Metadata Management System	Data Catalog
Managed Data Objects	<ul style="list-style-type: none"> Relational DB (table/column) 	<ul style="list-style-type: none"> Relational/NoSQL DB (table/column) File systems such as HDFS (folders/files/fields) Event message (topic) Query, Data API, Knowledge (Article), Report/Dashboard, etc.
Meta data management items	<ul style="list-style-type: none"> Description of table/column Column data type and size Relationship between tables (PK, FK, etc.) 	<ul style="list-style-type: none"> Biz meta, such as data name, description, tag, and Q&A Technical meta information such as data lineage and data profiling Utilization meta such as query, API, report/dashboard, etc.
Responsible for entering metadata	<ul style="list-style-type: none"> Data modeler DB manager 	<ul style="list-style-type: none"> Catalog Agent: Automatic creation/recommendation Data Steward Users: Description, Tags, Ratings/Reviews, Glossary
data search	<ul style="list-style-type: none"> Exact 'table name' or 'column name' 	<ul style="list-style-type: none"> Table name, file name, column name, field name Business (work) keywords 'Data Owner' Name, 'Data Steward Name', etc.
tool extension	<ul style="list-style-type: none"> Data modeling tools data dictionary 	<ul style="list-style-type: none"> Interactive query service, glossary (internal) Data preprocessing tool Data analysis tools

데이터 카탈로그는 데이터 레이크에서 주로 사용되는 만큼 관계형 DB와 NoSQL DB의 정보를 모두 포함할 수 있다. 또한, 동적주제도 시스템은 여러 분석 모듈이 수행되는 환경이기 때문에 비즈니스 메타데이터와 프로세스 메타데이터가 제공하는 정보가 효과적으로 활용될 수 있다. 시스템에서 관리하는 데이터와 분석 업무의 특성을 고려했을 때, 데이터 카탈로그는 동적주제도 시스템에 적용하기에 적절한 것으로 판단했다.

2.3.2 데이터 카탈로그 설계

데이터 카탈로그는 대상이 되는 데이터의 수집/처리/분석/활용과 관련된 업무를 지원·조정·연계·통합하는 역

할을 수행하기 때문에, 정형화하기 어렵다[7]. 이런 점을 고려해 동적주제도 시스템에 적용하기 위한 데이터 카탈로그 구축 방안을 마련하기로 했다.

우선 데이터 카탈로그의 정보 수집 대상이 될 수 있는 시스템의 데이터 객체를 Table 7에 정리했다.

Table 7. Data objects in the system

Category	Data Object
DBMS	Database, Schema, Table, Column
File	Filesystem, Folder, File, Field
Event	Event message, Topic
Operation/Statistics	Queries, Data API, Reports, Dashboards
Etc	User-created objects, etc

다음 Table 8은 세 개의 메타데이터 영역에서 각 데이터 객체에 대해 관리하는 정보를 정리한 것이다[7].

Table 8. Items by metadata type(Source : Seonwoong Yoon, Make Data Catalog, 2021, pp.177)

Metadata type	Metadata type
Business	<ul style="list-style-type: none"> Data name, description, tag, data owner, data steward, popular user, user rating/review, author
Technology	<ul style="list-style-type: none"> Object name, hierarchy, schema list, table list, column list, folder list, file list, topic list, data profiling, SQL information, URL information, report/dashboard image, application object, application query, analysis process, Analysis
Operation	<ul style="list-style-type: none"> Query List, Data API List, Report/Dashboard List, Data Lineage

이런 내용을 기반으로 동적주제도 시스템의 데이터 카탈로그에서 제공할 주요 기능을 Table 9와 같이 선정했다.

Table 9. Data catalog function of dynamic thematic map system

Category	Function
User	Catalog search
	Catalog Inquiry
	Daat Inquiry
	Data query response
	Catalog curation
Manager	DB standardization management
	Data Utilization Status Board
	Category management
	Data Security Management
	system monitoring

사용자는 다양한 방법으로 '카탈로그 검색'을 이용하고, 결과에 대한 '카탈로그 조회'는 데이터의 비즈니스/테크니컬/운영 메타데이터 정보를 확인할 수 있다. 확인 후 '실데이터 조회'를 이용하거나, 추가 정보가 필요하면 '데이터 질의 응답'으로 관리자와 소통할 수 있다. 데이터 Steward는 '카탈로그 큐레이션'을 통해 본인의 비즈니스 카테고리를 관리하고, 'DB 표준화 관리'를 통합된 환경에서 이용할 수 있다.

관리자는 대시보드 형태의 '데이터 활용 현황판'을 조회하고, '카테고리 관리'에서 비즈니스 카테고리나 데이터 유형별 카테고리를 관리한다. 개인정보를 비롯한 '데이터 보안관리'와 시스템의 상태, 문제점을 확인할 수 있는 '시스템 모니터링' 기능도 제공된다.

여기에서 선정한 기능은 '동적주제도 시스템의 DB 통합 관리'를 달성하기 위한 것으로, 데이터 카탈로그의 모든 기능을 제공하지는 않는다. 머신러닝을 이용한 메타데이터 추천 등은 포함하지 않았으며, DB 표준화 관리는 일반적으로 데이터 카탈로그에서 제공되는 기능은 아니지만, DB 관리 기능에 추가했다. 이를 통해 사용자와 관리자는 통합된 환경에서 동적주제도 시스템 DB의 전반적인 내용을 조회하고 관리할 수 있다.

이런 기능이 제공되기 위해서는 메타데이터 수집, 데이터 프로파일링, 데이터 활용 현황을 수집할 수 있는 서버 기능의 구현이 필요하다. 이에 대한 내용은 보다 기술적인 측면의 접근이 필요한 부분으로 상세 설계 과정에서 구체화 될 것이다.

2.3.3 활용 및 기대 효과

데이터 카탈로그는 동적주제도 시스템의 모든 데이터베이스에 대한 정보를 다양한 관점에서 제공할 수 있다. 데이터 카탈로그의 기능과 DB 표준화 관리 업무를 통합해 제공함으로써 DB 관리자 및 업무 담당자는 각자의 업무 관점으로 데이터베이스를 바라보고 원하는 정보를 검색할 수 있다.

또한, 데이터베이스를 체계적으로 관리할 수 있는 환경과 지원 도구를 제공함으로써 데이터베이스의 정보를 보다 편리하게 현행화시켜 관리할 수 있다.

관리자에게 제공되는 도구는 시스템에 입력된 데이터 카탈로그 정보와 실제 구축된 데이터베이스에 접근해 운영 현황을 점검할 수 있다. 예를 들어, DB 표준화 정책에 어긋나거나 최신화 되지 않은 카탈로그 정보를 검색해 지정된 업무 담당자에게 제공할 수 있을 것이다. 또한 전체 데이터베이스의 활용 통계 정보는 시스템의 정확한

현황과 문제점, 개선사항을 파악할 수 있는 수단을 제공하기 때문에 전반적인 시스템의 활용성을 높은 수준으로 유지할 수 있는 기반을 제공하게 될 것이다.

데이터 측면에서는 비정형/정형의 구분 없이 모든 데이터를 단일 환경에서 관리할 수 있다. 이는 시스템의 모든 데이터에 대한 확인과 모니터링이 가능함을 의미하기 때문에 데이터 표준화, 현행화, 정합성 등에서 개선 효과를 기대할 수 있다.

3. 결론

데이터의 통합 관리를 위해 반복되는 정보를 공통 컬럼으로 정의해 관리하는 방식은 빅데이터를 다루는 NoSQL DB에 적용하기에는 적합하지 않았다.

메타데이터 관리시스템은 데이터를 구성하는 정보를 관리함으로써 데이터의 유지 보수 및 변경 관리를 효율적으로 수행하도록 지원하지만, DBA 등 개발 레벨의 정보를 다루며 관계형 DB가 주요 대상이기 때문에 마찬가지로 NoSQL에는 적합하지 않았다.

데이터 카탈로그는 데이터 레이크의 데이터 늪(data swamp) 문제를 해소할 수 있는 방안으로 주로 소개되고 있지만, 데이터 관리 관점에서도 많은 장점을 가지고 있다. 구조적, 비구조적 데이터를 모두 수용하고, 비즈니스 관점의 정보를 포함하면서 활용 범위를 메타데이터 관리 시스템보다 크게 확장했다.

본 연구 내용을 기반으로 동적주제도 시스템의 데이터 관리 기능이 구현될 것이다. 또한, 통합 DB의 관리 기능이 필요하거나 유사한 시스템을 구축하는 경우에도 참고가 될 수 있을 것이다.

동적주제도 시스템은 목표시스템의 운영 계획에 맞춰 데이터 카탈로그의 정제된 기능을 제공하겠지만, 필요에 따라 데이터 카탈로그의 자동화 Agent, Data API, Machine Learning 등의 기능을 적극적으로 활용한다면 더욱 풍부하고 강력한 데이터 관리 시스템을 갖출 수 있을 것이다.

References

- [1] KAIA, Digital live land information technology development project planning final report, planning report, Science & Technology Strategy Institute, Korea, pp544-585.

- [2] Tibco, Data Catalog, Available From:
<https://www.tibco.com/ko/reference-center/what-is-a-data-catalog> (accessed Feb. 9, 2023)
- [3] 2e Consulting, Technology to collect and use data, Available From:
<https://www.youtube.com/watch?v=XJEqvvmWxo> (accessed Jan. 10, 2023)
- [4] KOREA Data Agency, Effective data management strategy using Metadata, Available From:
<https://dataonair.or.kr/db-tech-reference/d-lounge/technical-data/?mod=document&uid=236053> (accessed Feb. 9, 2023)
- [5] Oracle, Data Catalog, Available From:
<https://www.oracle.com/kr/big-data/data-catalog/what-is-a-data-catalog/> (accessed Feb. 9, 2023)
- [6] 2e Consulting, Understanding the Data Catalog, Available From:
<https://www.youtube.com/watch?v=Zq90TPgFhj8> (accessed Jan. 10, 2023)
- [7] Seonwoong Yoon, Make Data Catalog, p.223, GoodLand, 2021, pp31-40, pp162-163, pp177.

홍종유(Jongyoo Hong)

[정회원]



- 2002년 2월 : 인하대학교 컴퓨터 공학과 (공학사)
- 2008년 11월 ~ 현재 : ㈜제이비티 이사

<관심분야>

정보통신, GIS, SW공학

김성환(Sunghwan Kim)

[정회원]



- 1998년 2월 : 건국대학교 전기공학과 (공학사)
- 2017년 2월 ~ 현재 : ㈜제이비티 상무

<관심분야>

정보통신, GIS, 3D, 디지털트윈