

# A Study on the Standardized Classification Scheme of the Various Railway Information Systems

Yong-Ho Choi\*, Tae-Ki An\*\*, Hyoung-Geun Kim\*\*\*

## Abstract

The new information service has been demanded due to the recent mobile internet activation, and the government is promoting the activation of the private use of the public data by putting up the Government 3.0. According to government policy, many public sectors provide public data, but the railway sector is inferior to other public sector. In the case of national railway corporation, urban railway is now operated by 14 corporations such as Seoul Metro through the nation and high-speed railway is now operated by Korea Railroad Corporation and Supreme Railways. It is very difficult to standardize and integrate data due to mutual interests of national railway corporation. This paper describes a way to standardize and integrate rail passengers information collected through research project.

▶Keyword: Railway Information, Standard Classification Scheme, Data Integration, Data Dictionary

## I. Introduction

최근 모바일 인터넷의 활성화로 새로운 정보서비스가 요구되고 있으며 실제 버스나 지하철 등 대중교통과 관련된 공공데이터를 활용한 교통정보 앱(App) 등이 개발되어 일상생활에서 유용하게 활용하고 있다[1]. 과거에는 연구나 정책 개발의 목적으로 연구자나 업무 담당자가 공공데이터를 제한적이고 폐쇄적으로 활용하였으나 최근에는 대중교통 서비스 등과 같이 많은 사람들이 수시로 스마트 디바이스를 통해 공공데이터를 활용하고 있다[2].

하지만 철도정보의 경우 철도운영기관별로 제공하고 있는 정보의 형태, 형식, 수준 및 제공수단 등이 상이하여 이를 활용하기에는 많은 어려움이 있다. 또한 도시철도의 경우 현재까지 실시간 열차 운행 정보 제공의 폐쇄성으로 인해 대부분의 실시간 운행정보 서비스는 철도운영기관에서 제공하는 시간표를 기준으로 제공하고 있다[3].

이러한 문제점들로 인해 철도 관련 서비스 사업자들은 개별적으로 철도정보를 수집하고 이를 통합하는 작업을 통해 서비

스를 제공함에 따라 많은 비용과 시간이 소요되고 있다. 또한 동일한 철도 데이터 수집을 위해 서비스 사업자간 중복 투자가 이루어져 매우 비경제적이다.

철도와 비교되는 버스의 경우 버스운행정보 제공을 위해 국가 및 지자체 차원에서 버스정보시스템을 구축하여 운영하고 있는데 2015년을 기준으로 전국 163개(광역8, 기초 152) 지자체 중 75개(46.0%) 지역에서 구축 및 운영되고 있다. 또한 경기도에서는 시·군을 경유하는 버스가 많기 때문에 일관된 정보를 제공하기 위하여 2006년부터 버스정보를 통합하여 수집·제공하고 있다[4].

본 연구에서는 버스정보제공과 같이 철도운영기관에서 제공하고 있는 정보를 수집하여 이를 표준화 및 통합함으로써 데이터 정확성과 일관성을 유지하고 양질의 데이터 확보를 목표로 하였다.

이를 위해 본 논문에서는 철도정보의 다양성을 고려하여 데이터 분류 체계 및 데이터 사전 설계를 통해 데이터 정확성과 신뢰성을 향상 시킬 수 있는 표준화 설계 방안에 대해 기술하였다.

• First Author: Yong-Ho Choi, Corresponding Author: Tae-Ki An

\*Yong-Ho Choi (yhchoi@u-core.co.kr), Smart R&D Center, U-CORE SYSTEM, Information and Communication Technology, Hanbat National University

\*\*Tae-Ki An (tkahn@krii.re.kr), New Transportation Systems Research Center, Korea Railroad Research Institute

\*\*\*Hyoung-Geun Kim (khg9792@u-core.co.kr), Smart R&D Center, U-CORE SYSTEM

• Received: 2017. 11. 28, Revised: 2017. 12. 12, Accepted: 2018. 01. 09.

• This work was supported by Korea Agency for Infrastructure Technology Advancement Grant.

## II. Related works

### 1. Status of Providing Railway Information

현재 철도정보 데이터 수집을 위해서는 그림에서 보는 바와 같이 공공데이터포털(Fig 1)과 서울 열린데이터 광장(Fig 2)을 통해 수집이 가능하다. 하지만 이곳에서 제공하고 있는 철도정보 데이터들의 경우 각 운영기관에서 주관적으로 제공하는 정보들로 대부분의 정보들이 시간표, 시설정보 등과 같은 정적정보들이 대부분이어서 이를 활용하여 새로운 서비스를 개발하기에는 다소 어려운 실정이다. 또한 운영기관별로 다양한 형태와 형식으로 정보를 제공하다 보니 사용자 입장에서 각 운영기관별 정보를 취합하여 사용하기에는 많은 시간과 비용이 소요되기 때문에 정보 활용성이 매우 떨어질 수밖에 없다.



Fig. 1. DATA.GO.KR[5]



Fig. 2. SEOUL OPEN DATA PLAZA[6]

### 2. Status of Railway Corporations Information

Fig 3은 각 철도운영기관에서 제공하는 역사 편의시설 정보이다. 그림에서 보는 바와 같이 각 운영기관별로 제공하고 있는 수유실 정보 형태를 보게 되면 서울교통공사의 경우 모유수유실, 대전광역시철도공사는 수유방, 대구도시철도공사는 아기사랑방 등과 같은 정보를 각기 다른 형태로 제공하는 것을 확인할 수 있다.

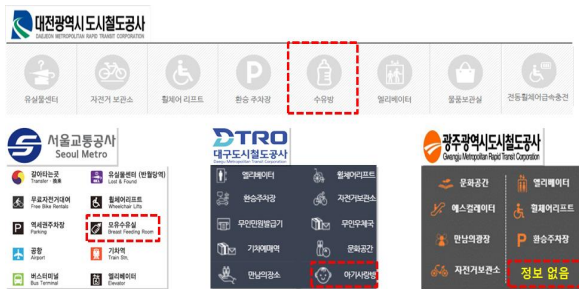


Fig. 3. Status of Nursing Room Information Provided by Operating Agency

대 소요시간을 비교한 것이다. 총 운행구간은 모두 23개 구간으로 서울교통공사는 49분, 네이버는 55분, 다음은 52분이었으며, 정보제공처별 최대 소요시간 차이는 6분으로 나타났다.

이처럼 같은 운행 구간을 정보제공처마다 다른 소요시간을 제공함으로써 철도 이용자의 혼란과 불편을 가중하고 있다.

Table 1. Status of operation time information

	Service Section (Seoul 1Line, Seoul Station→Uijeongbu)	Time (Min)
Seoul Metro [7]		49
Naver [8]		55
Daum [9]		52

## III. Data Standardization Design

### 1. Data Classification Scheme Design

데이터 표준화 핵심은 의미 있는 데이터의 상호 교환이 발생하기 전에 관련된 데이터들을 정확히 식별하고 그 정의를 이해해야 한다는 데 있다[10-11].

따라서 철도정보 데이터 표준화를 위해 표준화가 가능한 데이터를 분석하여 각 데이터 속성(Attribute)에 따라 분류하였으며, 그 속성을 식별하기 위한 체계를 설계하였다.

#### 1.1 Classification Scheme

Fig 4는 데이터 분류체계 모식도로 데이터 항목수가 적은 대/중/소분류는 1자리로 구성하였으며, 세분류/세세분류는 추후 분류 구성이 증가할 것으로 예상되어 2자리로 설정하였다.

Table 1은 서울 1호선 서울역에서 의정부까지의 동일시간

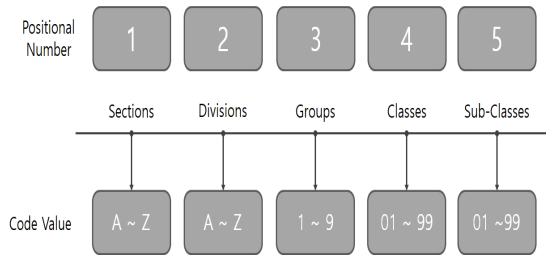


Fig. 4. Definition of Classification System

1.2 Attribute Scheme

속성은 역사, 차량 및 운영정보에서 공통적(단위, 거리, 위치 등)으로 가지는 공통속성과 개별적(온도, 습도, 실시간성 정보 등)으로 가지는 개별특성으로 분류하였으며, Table 2는 속성체계를 정의한 것이다.

Table 2. Attribute Scheme Definition

	Definition	Remarks
Common Attribute	1. Common attribute information of railway information 2. Attributes for managing common attributes	Unit, distance, location, etc.
Individual attribute	1. Individual proprietary attribute information of information 2. Management attributes related to the characteristics of information	Facilities, environmental information, etc.

1.3 Identification Scheme

식별 체계는 최소한의 의미 부여를 통해 역사, 차량 및 운영 정보를 식별할 수 있도록 하는 것으로 Table 3은 식별 체계를 정의한 것이다.

Table 3. Identification Scheme Definition

Master	Form
Station	□+□□+□□□ Region + Railway Operator Code + Station Code
Train	□+□□□+□□□+□□□□ Region + Railway Operator Code + Organization Number + Train Number
Operation	□+□□□+□ Region + Railway Operator Code + Line Code

예를 들면 역사 정보의 경우 공통속성인 지역, 철도운영기관 코드와 개별속성인 역코드를 통해 역사정보를 식별 할 수 있는 구조이다.

2. Data Dictionary

데이터 표준화를 위해서는 상호 교환되는 데이터의 변환을 배제하거나 가능한 최소화되도록 하는 것이 필요하다[12]. 이를 위해 데이터 표준화를 효율적으로 구성할 수 있는 데이터사전을 정의(Table 4)하였다.

Table 4. Data Dictionary Definition

Master	Detail of Function
Word Dictionary	The words extracted for all the words used in the data dictionary are classified by classifying words and type and are defined in consideration of the purpose.
Term Dictionary	Based on the word dictionary, the terms are used to manage standards, general characteristics, and work orientation.
Domain Dictionary	The domains are categorized and shaped by taking into account the physical uses, frequency of use, and physical characteristics of the data.
Code Dictionary	It is managed by selecting codes from terms, defining code values, and defining ownership for each defined code.

Fig 5는 표준화를 위한 데이터사전 흐름도이며, 네 개의 사전기능으로 철도정보 데이터 표준화와 향후 유지보수를 위하여 기본이 되는 데이터 사전관리를 지원하며, 이에 따라 발생하는 테이블 관리, 화면 입출력 항목관리 등의 일관성을 보장한다.

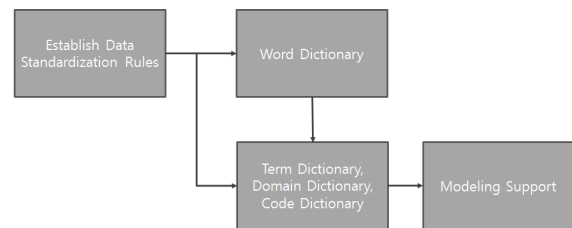


Fig. 5. Data Dictionary Flow Chart

2.1 Word Dictionary

단어사전은 Table 5와 같이 단어명, 정의, 영문명, 영문 약어로 구성되며, 관리 기준으로는 표준성, 일반성 및 대표성 등을 갖는다. 단어 한글명은 명사로 작성하며, 단어의 정의는 누구나 알 수 있도록 간결하고 명료하게 작성한다. 또한 영문 약어는 영문명의 본래 뜻을 알 수 있도록 최대 4자 이내 대문자로 작성한다.

Table 5. Word Dictionary Component

Word Name	Definition	English	
		Name	Abbreviations
floor	The floor of a room is the part of it that you walk on.	floor	FLOR
screen	One of a series of doors at a train or subway station, at the edge of a platform, that prevent access beyond the platform when there is no train to be boarded	platform screen door	SCR

2.2 Term Dictionary

용어사전은 Table 6과 같이 용어명, 정의, 영문명으로 구성되며, 관리 기준으로는 표준성, 일반성, 업무 지향성 등을 갖는다. 용어명은 단어 사전에서 정의된 단어의 조합으로 구성하며, 영문명은 단어 사전에서의 영문약어를 조합하여 구성하였다.

Table 6. Term Dictionary Component

Term Name	Definition	English Name
Transfer Path Classification Code	Categorized by the path of the user to move the station	MV_PATH_DV_CD
Transfer Public Transit Code	Categorized by transit type public transportation	CHTN_PUB_TRFC_CD

2.3 Domain Dictionary

도메인사전은 Table 7과 같이 도메인명, 데이터, 측정단위로 구성되며, 관리 기준으로는 표준성, 유일성 및 업무지향성 등을 갖는다. 이때 표준성은 공통적으로 사용되는 속성을 대상으로 정의하며, 유일성은 동일한 내용의 도메인이 서로 다르게 선언되지 않도록 한다. 또한 업무지향성은 일반화하여 정의하기 보다는 업무 특성을 반영 할 수 있도록 하였다.

Table 7. Domain Dictionary Component

Domain Name	Data			Measurement unit
	Type	Length	Value	
Code	CHAR	2	-	-
Length	NUMBER	10,3	-	M

2.4 Code Dictionary

코드사전은 Table 8과 같이 코드명, 코드값 설명으로 구성되며, 관리 기준으로는 재사용성, 일관성 및 정보 분석성 등을 갖는다. 재사용성은 정의된 코드를 재사용하는 것이 데이터에 대한 이해력을 높이고 코드 관리를 용이하게 한다. 일관성은 코드 업무 범위 내에서 가능한 유일하게 정의되어야 하며, 정보 분석성은 가능한 범위의 데이터를 모두 코드화 하여 관리하는 것으로 사용자가 스스로 텍스트로 입력하는 값을 최소화 시키도록 하였다.

Table 8. Code Dictionary Component

Code Name	Code value	Explanation
Braille Facilities Code	R100	Facility code showing the braille letters of the blind person
Convenience Facility Code	R101	Code for convenience facilities built in the station

IV. Applied Case and Evaluation

1. System Design and Implementation Environment

철도정보 데이터 표준화 및 효율성 검증을 위해 웹서버, 미들웨어, DB서버 이중화로 네트워크(Fig 6) 개발 환경을 구성하였으며, S/W 개발 환경은 Linux Red Hat 서버에 Jboss, Oracle로 구성하였다.

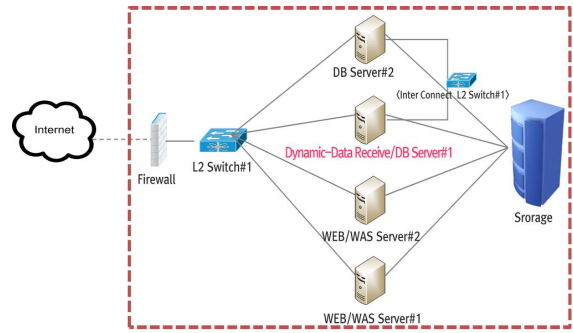


Fig. 6. System Configuration Environment

데이터 표준화 연구 대상은 「전국 철도운영기관에서 제공 중인 철도 이용객 편의 및 안전 시설물 정보(수집기간 2016년 이후, 7,317,127건 이상)」이며, 이를 역사, 차량 및 운영 정보별 분석을 통해 표준화 하였다.

2. Data Standard Classification Scheme

데이터 표준분류체계는 Fig 7과 같으며, 총 71개의 데이터 분류 항목으로 구성되어 있다. 대부분은 분류체계를 고려하여 도시/광역 철도와 일반/고속철도로 분류하였으며, 중분류는 데이터 속성을 고려하여 동적/정적 데이터로 분류하였다. 소분류는 데이터 식별체계를 고려하여 역사, 운영 및 차량정보로 각각 분류하였다.

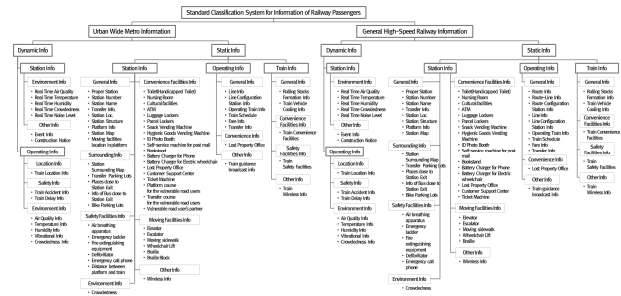


Fig. 7. Standard Classification Scheme of Railway Passengers Information

3. Result of Data Dictionary Application

Table 9는 철도정보 데이터 표준화를 위해 데이터사전을 적용하여 표준데이터를 구분한 결과로 단어사전은 2,801건, 도메인사전은 68건, 용어사전은 1,481건으로 나타났다.

Table 9. Data Dictionary Classification Result

Data Classification	Data Dictionary	Number of Case
Standard Data	Word Dictionary	2,801
	Domain Dictionary	68
	Term Dictionary	1,481

Fig 8은 표준화 대상인 철도정보를 데이터사전에 적용하여 분류한 관계도이다.

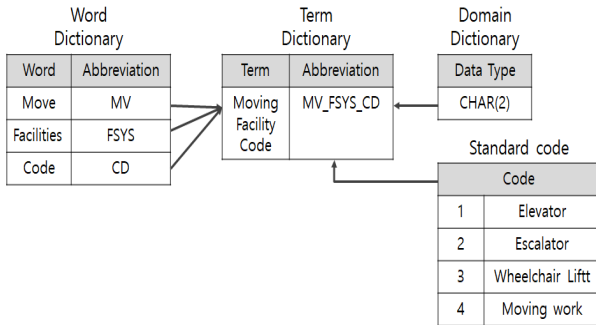


Fig. 8. Standardization Targets and Data Dictionary Relationships

Fig 8에서 보는 바와 같이 용어사전 “이동시설물코드”는 단어사전에 조합된 “이동, 시설물, 코드” 단어의 조합 용어이며, 도메인 사전은 용어사전의 이동시설물코드의 데이터 형태인 “CHAR(2)”를 나타낸다. 표준코드는 상품구분 코드인 “01:엘리베이터, 02:에스컬레이터, 03:휠체어리프트, 04:무빙워크”인 항목을 나타내게 된다.

#### 4. Efficiency Verification through Application Examples

Fig 9는 전국철도정보를 표준분류체계와 데이터사전을 적용하여 세부항목별로 수집·관리를 할 수 있는 시스템이다. 철도이용객 수집관리시스템은 철도운영기관마다 상이한 형태 및 수준으로 제공하고 있는 정보들을 표준화하여 관리함으로써 데이터의 신뢰성을 향상시킬 수 있다.



Fig. 9. Railway Passenger Information Collection and Management System

철도통합정보시스템(Fig10)은 Fig 9에서 수집된 데이터를 표준화 및 가공을 통해 철도이용객에게 정보를 제공하기 위한 시스템으로써 전국 철도 역사 및 차량 정보를 철도 이용객들이 쉽게 접근 할 수 있도록 하기 위해 시인성이 높은 디자인과 PC 및 스마트기기에서도 접근이 가능하도록 하여 철도이용객정보의 이용성을 향상시킬 수 있도록 하였다.



Fig. 10. Railway Integrated Information System[13]

이와 같이 철도데이터의 표준화는 철도데이터에서 가공을 통해 철도정보로써의 가치를 향상시킴으로써 철도 정보를 활용한 다양한 철도 정책과 철도 서비스를 마련하기 위한 기반 요소가 될 것이다.

또한 철도데이터 표준화는 다음과 같은 정성적 효과를 가질 수 있다. 첫째, 철도 분야 정보 시스템 구축 시 표준 데이터를 활용함으로써 신규 데이터베이스 구축을 위한 비용 절감이 가능하다. 둘째, 다양한 공공데이터와 융합 연구를 통해 신규 정보를 가공하기 위한 활용이 용이하다. 셋째, 표준화된 데이터를 민간 및 공공기관에서 활용을 함으로써 정보를 이용하는 국민의 혼동을 줄이고, 데이터의 일관성을 유지할 수 있다.

## V. Conclusions

본 연구는 전국철도운영기관에서 생성되는 철도정보의 효율적인 데이터 관리를 위해 데이터 표준화 체계를 설계하였으며, 또한 데이터 사전을 제시하여 이를 통한 데이터 관리 효율성을 검증하였다. 데이터 표준화 체계는 분류·속성·식별 체계로 구성되며, 표준화된 데이터는 데이터 사전인 단어·용어·도메인·코드 사전을 통해 관리된다. 데이터 표준화 체계와 데이터사전을 통해 관리된 철도정보 데이터는 정보 이용자 간 의사소통의 혼란을 예방할 뿐만 아니라, 데이터의 정확성과 일관성을 유지하여 고품질의 데이터를 확보 할 수 있는 것으로 나타났다. 하지만 철도정보의 경우 운영기관마다 정보의 갱신 주기가 달라 정보의 신뢰성을 확보하기가 어려울 뿐만 아니라 지속적인 수집을 위해서는 많은 비용과 시간이 소요된다.

향후 이러한 문제를 해결하기 위해서는 철도운영기관에서 생성되는 정보를 운영기관이 주기적으로 표준화 DB에 반영 할 수 있는 자동 수집 기술 관련 연구가 필요하다.



## ACKNOWLEDGMENTS

본 연구는 국토교통부 철도기술연구사업의 “ICT기반 철도 이용객 정보제공기술 개발” 연구비지원(17RTRP-B086931-04)에 의해 수행되었습니다.

## REFERENCES

- [1] Korea Local Information Research & Development Institute, “2014 Local Information White Paper”, KLID , 2014, pp84
- [2] Yeon-ung Hong, “A study on the invigorating strategies for open government data”, Journal of the Korean Data and Information Science Society, Vol 25, No. 4, pp769-777, 2014.
- [3] Yeon-suk Kwak, “Development of Urban Railway Information System Applying Real-Time Location Information and Operation Pattern”, The Journal of Korean Institute of Communications and Information Sciences, Vol.42, No.06, pp126-1275, 2017.
- [4] Mi-Yeong Bin, “Research on Enhancing the Operation of the Bus Information System in Gyeonggi-Do”, Gyeonggi Research Institute, pp5, 2016.
- [5] Railway Data Gathering, <https://www.data.go.kr/>
- [6] Railway Data Gathering, <http://data.seoul.go.kr/>
- [7] Route Search, <http://www.seoulmetro.co.kr/kr/cyberStation.do?menuIdx=538>
- [8] Route Search, <http://map.naver.com/>
- [9] Route Search, <http://map.daum.net/>
- [10] Y. Lirov, "Computer Aided Software Engineering of Expert Systems", Expert Systems with Applications, Vol.2, pp.333-343, 1991.
- [11] W. Kozaczynski, "A Knowledge-Based Approach to Software System Understanding", Proc. 6th Annual Knowledge-Based Software Conference, pp.162-170, 1991.
- [12] Byoung-Yup Lee, “Project Management Methodology using Managing Data Dictionary”, Journal of The Korea Contents Association, Vol 9, No. 3, pp72-80, 2009.
- [13] Railway Integrated Information System, <http://visual.railportal.kr/RailContents/altmInfoSys/index.do>.

## Authors



Yong Ho Choi received the M.S. degrees in Information and Communication Technology from Hanbat National University, Korea, in 2015. Yong-Ho Choi is currently a Chief Researcher in the Smart R&D Center U-Core System. He

interested in data interaction, data standardization.



Tae Ki An received the M.S. degree in electronic engineering from Kyungpook National University, Daegu, Korea, in 1996, and the Ph.D. degree in Department of Electrical and Computer Engineering from Sungkyunkwan University, Seoul, Korea in

2011. He is a Chief Researcher in Korea Railroad Research Institute. His research interests are artificial intelligence, pattern recognition, and video analysis.



Hyoung Geun Kim received the B.S. degrees in Computer Engineering in 2005. Hyoung Geun Kim is currently a Senior Researcher in the Smart R&D Center U-Core System. He interested in artificial intelligence, deep learning.