

빅데이터 플랫폼 및 센터

2021

데이터 품질관리 가이드

부록 산출물 서식 >>





부록1 >>

1장	정량적 품질진단 규칙 예시	
	1. 개요	07
	2. 진단규칙 예시 설명	08
2장	데이터 품질관리 산출물 서식	
	1. 데이터 품질진단 산출물 서식	46
	2. 데이터 관리 산출물 서식	77
	3. 데이터 연계키 서식	84
	4. 연계표준사전 서식	85

부록2 >>

1장	데이터 연계키	
	1. 개요	97
	2. 데이터 연계키 발굴	99
	3. 데이터 연계키 관리체계	108
2장	연계표준사전	
	1. 개요	113
	2. 연계표준사전 표준화	115
	3. 연계표준사전 관리체계	124
3장	소유권 및 이용약관 개선 정의서	
	1. 개요	132
	2. 해외 데이터 소유권 관련 사례 분석	139
	3. 빅데이터 플랫폼 이용약관 현황	150
	4. 데이터 소유권 및 이용약관 개선	155



부록 1 >>

1장 정량적 품질진단 규칙 예시

2장 데이터 품질관리 산출물 서식



01

개요



데이터 품질 지표별 진단대상 예시를 15개 분야로 나누어 설명한다.

- [1] 진단 대상 데이터 모델 사례
- [2] 진단예시 > 단독 완전성
- [3] 진단예시 > 조건 완전성
- [4] 진단예시 > 구조 완전성
- [5] 진단예시 > 코드 유효성
- [6] 진단예시 > 형식 유효성
- [7] 진단예시 > 여부 유효성
- [8] 진단예시 > 날짜 유효성
- [9] 진단예시 > 범위 유효성
- [10] 진단예시 > 선후관계 유효성
- [11] 진단예시 > 참조 무결성
- [12] 진단예시 > 중복 일관성
- [13] 진단예시 > 표준 일관성
- [14] 진단예시 > 논리관계 정확성
- [15] 진단예시 > 파생항목 정확성

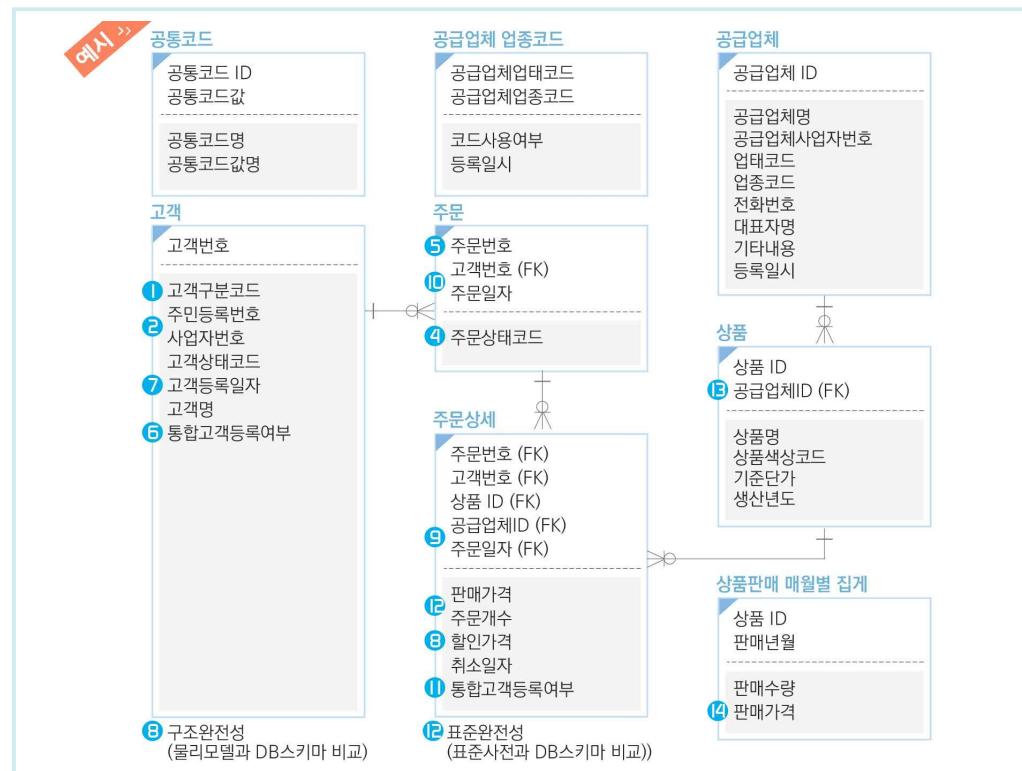
02 진단규칙 예시 설명

2.1 진단 대상 데이터 모델 사례

상품판매 DB 데이터 모델을 사례로 데이터 품질관리 지표별 정량적 진단규칙을 도출하는 예시와 상세한 도출 방법을 설명한다.

아울러 각 품질 상세지표별로 플랫폼의 진단 사례를 추가 예시로 설명한다.

▶ 진단 대상 ERD 논리 모델



▶ 데이터 품질 지표별 진단대상 예시

품질지표	품질상세지표	진단대상	비고
완전성	단독 완전성	고객구분코드	예시1
		트랙명/항목등록일시/테이블순번	예시2
		계량기용량	예시3
		아파트코드	예시4
	조건 완전성	주민등록/사업자번호	예시5
		병력당뇨내용	예시6
		소유권변경일	예시7
유효성	구조 완전성	물리모델<->DB스키마	예시8
		주문상태코드	예시9
		대상자성별코드	예시10
	코드 유효성	시도코드	예시11
		차로유형구분코드	예시12
		주문번호	예시13
		(채번)일련번호/전화번호/우편번호	예시14
		우편번호	예시15
		가맹점전화번호	예시16
	형식 유효성	CAS등록번호	예시17
		통합고객등록여부	예시18
		결혼여부명/자녀여부명	예시19
		주소표준화	예시20
		고객등록일자	예시21
		년도/월/년월/시간(일자시분초)	예시22
	날짜 유효성	발생년도	예시23
		할인가격	예시24
		캠핑장화장실수/10미터기온/기관일련번호	예시25
		시간대별이용자비용	예시26
		휴업폐업1등급비용	예시27
		생체검사조직학적등급수	예시28
	범위 유효성	온도	예시29
		주문일자<->배송일자	예시30
		생체검사판독일자	예시31
	선후관계 유효성	고객번호	예시32
		환자대체번호	예시33
		CAS등록번호	예시34
일관성	참조 무결성	통합고객등록여부	예시35
		사망일자	예시36
		표준사전<->DB스키마	예시37
	중복 일관성	상품(공급업체ID, 색상코드)	예시38
		배송비여부/배송비	예시39
		음주기간년수	예시40
정확성	논리관계 정확성	판매가격	예시41
		방사선치료총선량	예시42
	파생항목 정확성	대중교통이용정보	예시43

2.2 진단 예시 > 단독 완전성

(예시 1) 고객구분코드 : 단독 완전성 품질지표를 기준으로 '고객구분코드' 항목의 업무규칙 진단기법에 의한 진단규칙 도출 예시

1. 진단대상 분석(업무규칙)

1.1 '고객구분코드' 항목의 NOT NULL 제약조건(구조) 확인

- 단독완전성 항목이 NOT NULL 구조인 경우 품질진단 대상에서 제외하며 NULL 값이 허용되는 경우 품질진단 대상이 될 수 있다.

테이블명	컬럼명	NULL여부
고객	고객번호	N
	고객구분코드	Y
	주민등록번호	Y

(조회 SQL - Oracle 11g 기준)

```
SELECT TABLE_NAME AS 테이블명
      , COLUMN_NAME AS 컬럼명
      , NULLABLE AS NULL여부
FROM ALL_TAB_COLUMNS
WHERE TABLE_NAME ='고객';
```

1.2 고객 테이블 데이터 구성 확인

고객번호	고객구분코드	주민등록번호	사업자번호
E0001	C02		10-181-39750
E0002		76****_*****	
E0003	C02		60-381-13200

(조회 SQL - Oracle 11g 기준)

```
SELECT 고객번호
      , 고객구분코드
      , 주민등록번호
      , 사업자번호
FROM 고객 ;
```

1.3 고객구분코드 업무규칙 정보 확인

- '고객구분코드' 항목의 값은 물리적으로는 NULL을 허용하는 구조이나 업무적으로 NOT NULL 이어야 한다.

2. 업무규칙 대상 분석 결과 및 진단규칙 도출

- 업무규칙 정보를 참고하여 '고객구분코드' 항목에 NULL 값이 있는 경우 오류로 도출하는 진단 규칙을 적용한다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT 고객번호
      , 고객구분코드
FROM 고객
WHERE 고객구분코드 IS NULL ;
```

(예시 2) 트랙명/항목등록일시/테이블순번 : 업무 규칙에 의해 값이 항상 존재(Not Null)해야 한다고 정의된 컬럼은 해당 컬럼의 입력 값이 널(Null)이거나 공백문자로만 입력된 경우 오류 데이터로 도출 예시

1. 진단대상 분석

1.1 업무규칙 정의에 'NULL'을 허용하지 않는 경우 단독 완전성 진단 대상으로 진단 SQL수행

- 오류로 추출되는 대상은 'NULL' 또는 공백만 포함된 문자열로 입력된 데이터

테이블명	컬럼명	데이터 타입	NULL 허용여부
숲길이용패턴	트랙명	character varying	N
첨부파일 정보	항목등록일시	timestamp	N
귀산촌 이용자 정보	테이블순번	numeric	N

(조회 SQL - PostgreSQL 기준)

* 컬럼의 데이터 타입이 문자형

```
SELECT '숲길이용패턴' as TableName, '트랙명' as ColName, count(case when
length(trim(COALESCE(트랙명,''))) < 1 then 1 end) as null_cnt from 숲길이용패턴
```

* 컬럼의 데이터 타입이 날짜형

```
SELECT '첨부파일 정보' as TableName, '항목등록일시' as ColName, count(case when
COALESCE(항목등록일시,'9999-12-31') = '9999-12-31' then 1 end) as null_cnt from 첨부파일 정보
```

* 컬럼의 데이터 타입이 숫자형

```
SELECT '귀산촌 이용자 정보' as TableName, '테이블순번' as ColName, count(case when
COALESCE(테이블순번,99999999999999999999) = 99999999999999999999 then 1 end) as
null_cnt from 귀산촌 이용자 정보
```

- 실제 DB Schema로부터 컬럼별 not null constraint 설정 확인

테이블명	컬럼명	데이터 타입	Null 허용여부
숲길이용패턴	월요일방문수	numeric	N
숲길이용트랙포인트	트랙포인트기록일시	timestamp	N
선사정보	선박회사명	character varying	N

(조회 SQL - PostgreSQL 기준)

```
SELECT a.table_name as 테이블명,a.column_name as 컬럼명, a.is_nullable as null허용여부
```

```
FROM information_schema.columns a
WHERE table_catalog = 'XXXXX'
AND table_schema = 'XXXXX'
```

* 컬럼의 데이터 타입이 문자형이 아닌 숫자형 또는 날짜형인 경우 DB에 실제로 not null constraint 가 설정되어 있다면 진단 대상에서 제외할 수 있음

(예시 3) 계량기용량 : 행정구역별 미납정보 주택용 테이블의 '계량기용량' 항목 프로파일 검사 도출 예시

1. 진단대상 분석

1.1 "계량기용량" 항목의 NOT NULL 제약조건(구조) 확인

- 단독완전성 항목이 NOT NULL 구조인 경우 품질진단 대상에서 제외하며 NULL 값이 허용되는 경우 품질진단 대상이 될 수 있다.
- 업무규칙 정보를 참고하여 '계량기용량' 항목에 NULL 값이 있는 경우 오류로 도출하는 진단 규칙을 적용한다.
- 행정구역별 주택용 계량기 용량 컬럼의 값은 항상 존재해야 한다.

테이블명	컬럼명	데이터 타입	NULL 허용여부
행정구역별 미납정보 주택용	계량기용량	numeric	N

(조희 SQL - MySQL 기준)

* 컬럼의 데이터 타입이 문자형
select 행정구역별 미납정보 주택용, 가스타입, 계량기용량, 계량기용량 ERR_VAL from 행정구역별 미납정보 주택용 where 계량기용량 is null

(예시 4) 아파트코드 : 데이터 업무 요건에 맞게 항상 채워져 있는지 진단 예시

1. 진단대상 분석

1.1 '아파트코드' 항목의 NOT NULL 제약조건(구조) 확인

- 단독완전성 항목이 NOT NULL 구조인 경우 품질진단 대상에서 제외하며 NULL 값이 허용되는 경우 품질진단 대상이 될 수 있다.
- 업무규칙 정보를 참고하여 '아파트코드' 항목에 NULL 값이 있는 경우 오류로 도출하는 진단 규칙을 적용한다.
- 지역별 음식물쓰레기 배출 추정정보의 아파트코드 컬럼의 값은 항상 존재해야 한다.

테이블명	컬럼명	NULL 허용여부
지역별 음식물쓰레기 배출 추정정보	아파트코드	N

(조희 SQL - PostgreSQL 기준)

select 배출년,
아파트코드,

배출량

+R3:R7from 지역별 음식물쓰레기 배출 추정정보 where COALESCE(CAST(아파트코드 AS TEXT),")
= ";

2.3 진단 예시 > 조건 완전성

(예시 5) 주민등록/사업자번호 : 조건 완전성 품질지표를 기준으로 '고객구분코드', '주민등록번호' / '사업자번호' 항목의 업무규칙 진단기법에 의한 진단규칙 도출 예시

1. 진단대상 분석(업무규칙)

1.1 '주민등록번호', '사업자번호' 목의 NOT NULL 제약조건(구조) 확인

- 조건완전성 진단 대상(주민등록번호, 사업자번호)항목이 NOT NULL 구조인지 확인 필요하며 조건의 대상이 되는 항목의 구조가 NULL 값이 허용 되는 경우 품질진단 대상이 될 수 있다.

테이블명	컬럼명	NULL여부
고객	고객번호	N
	고객구분코드	Y
	주민등록번호	Y
	사업자번호	Y

(조희 SQL - Oracle 11g 기준)

```
SELECT TABLE_NAME AS 테이블명
, COLUMN_NAME AS 컬럼명
, NULLABLE AS NULL여부
FROM ALL_TAB_COLUMNS
WHERE TABLE_NAME = '고객';
```

1.2 고객 테이블 데이터 구성 확인

고객번호	고객구분코드	주민등록번호	사업자번호
E0006	C01	76****_****	
E0007	CO1		
E0008	CO1	80****_****	
E0009	CO2		
E0010	CO2		20-782-07314

(조희 SQL - Oracle 11g 기준)

```
SELECT 고객번호
, 고객구분코드
```

```

, 주민등록번호
, 사업자번호
FROM 고객 ;
※ 고객구분코드 값 : CO1 - 개인 , CO2- 사업자

```

1.3 고객구분코드 관련 업무규칙 정보 확인

- '고객구분코드' 항목의 값이 '개인'일 경우 '주민등록번호' 항목의 값만 존재해야 하며, '고객구분코드' 항목의 값이 '사업자'일 경우에는 '사업자번호' 항목의 값만 채워져야 하는 규칙에 위배된 경우 오류로 도출할 수 있다.

2. 진단대상 분석 결과 및 진단규칙 도출(업무규칙)

- 업무규칙 정보를 확인한 결과 다른 항목(고객구분코드) 조건에 의해 NULL 값이 물리적으로 허용되는 주민등록번호, 사업자번호 항목은 조건항목(고객구분코드) 값에 따라 NULL 인 데이터가 존재할 수 없는 경우에 위배된 경우 오류로 도출할 수 있다.

(진단규칙 SQL - Oracle 11g 기준)

```

SELECT 고객번호, 고객구분코드, 주민등록번호, 사업자번호
FROM 고객
WHERE (고객구분코드 = 'C01' AND 주민등록번호 IS NULL)
OR (고객구분코드 <> 'C01' AND 주민등록번호 IS NOT NULL)
OR (고객구분코드 = 'C02' AND 사업자번호 IS NULL)
OR (고객구분코드 <> 'C02' AND 사업자번호 IS NOT NULL) ;

```

(예시 6) 병력당뇨내용 : 진단대상 컬럼의 조건에 따라 값이 완전성 업무 요건에 맞게 존재하는지 진단 예시

1. 진단대상 분석

1.1 "병력당뇨여부" 값이 "Y" 일때 "병력당뇨내용" 값이 NOT NULL인지 확인

- "병력당뇨여부" 값이 "Y" 일때 "병력당뇨내용" 값이 NULL이면 오류로 도출하는 진단규칙을 적용한다.
- "유방암_환자_건강정보"의 "병력당뇨여부"의 값이 Y이면 "병력당뇨내용"이 존재하여야 한다.

테이블명	컬럼명	NULL 허용여부
유방암_환자_건강정보	병력당뇨내용	N

(조회 SQL - Oracle 11g 기준)

```

SELECT 환자대체번호, 병력당뇨여부, 병력당뇨내용,
FROM 유방암_환자_건강정보
WHERE 병력당뇨여부 = "Y"
AND 병력당뇨내용 IS NOT NULL

```

(예시 7) 소유권변경일 : 데이터가 다른 데이터 항목의 업무 조건에 따라 항상 채워져 있는지 진단 예시

1. 진단대상 분석

1.1 소유권변동원인코드가 Not Null이면, 소유권변동일자도 Not Null이다.

- 소유권변동원인코드 관련 업무규칙 정보 확인
- 데이터가 다른 데이터 항목의 업무 조건에 따라 항상 채워져 있는지 진단
- 소유권변동원인코드가 Not Null이면, 소유권변동일자도 Not Null이다.

테이블명	컬럼명	NULL 허용여부
전국 지번별 소유자 특성 및 지가정보	소유권변경일	N

(조회 SQL - PostgreSQL 기준)

```

SELECT
TABLE_NAME AS 테이블명
,COLUMN_NAME as 컬럼명
,IS_NULLABLE
FROM INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_CATALOG='DB명'
AND TABLE_SCHEMA='스키마명'
AND TABLE_NAME='전국 지번별 소유자 특성 및 지가정보';

```

(오류내역 SQL - PostgreSQL 기준)

```

select 소유권변경원인코드, 소유권변경일
from 전국 지번별 소유자 특성 및 지가정보
where 소유권변경원인코드 is not null
and 소유권변경일 is null

```

2.4 진단 예시 > 구조 완전성

(예시 8) 물리모델<->DB스키마 : 구조 완전성 품질지표를 기준으로 '물리모델', 'DBMS 물리 스키마' 간 업무규칙 진단기법에 의한 진단규칙 도출 예시

1. 진단대상 분석(업무규칙)

1.1 물리모델 정보 확인

- 물리모델과 DBMS물리스키마 정보를 비교 하여 불일치 하는 항목이 존재하는지 확인 필요(고객 테이블 물리모델 예시)

CUSTOMER (고객) 테이블					
컬럼명	컬럼한글명	타입	PK 여부	길이	NOT NULL 여부
CUST_NO	고객번호	VARCHAR2	Y	10	Y
CUST_DVCD	고객구분코드	VARCHAR2	N	5	N
JUMIN_RGST_NO	주민등록번호	VARCHAR2	N	15	N
BS_RGST_NO	사업자번호	VARCHAR2	N	15	N
CUST_STCD	고객상태코드	VARCHAR2	N	5	N
CUST_RGST_DT	고객등록일자	VARCHAR2	N	8	N
CUST_NM	고객명	VARCHAR2	N	10	N
COMB_RES_YN	통합고객등록여부	VARCHAR2	N	1	N

1.2 DBMS 물리스키마 정보 확인

- 물리모델과 DBMS물리스키마 정보를 비교 하여 불일치 하는 항목이 존재하는지 확인 필요

컬럼명	컬럼한글명	타입	PK 여부	길이	NOT NULL 여부
CUST_ID	고객ID	VARCHAR2	Y	10	Y
CUST_DVCD	고객구분코드	VARCHAR2		5	
JUMIN_RGST_NO	주민등록번호	VARCHAR2		15	
BS_RGST_NO	사업자번호	VARCHAR2		10	

2. 진단대상 분석 결과 및 진단규칙 도출(업무규칙)

- '물리모델', 'DBMS 물리 스키마'를 근거로 물리 모델로 설계 된 내용(타입, 사이즈 등)이 실제 DBMS에 구현되어 있는지 진단한다.

(진단규칙 SQL - Oracle 11g 기준)

```

SELECT  A.TABLE_NAME      AS 테이블명
        , A.COLUMN_NAME    AS 컬럼명
        , B.COMMENT        AS 컬럼한글명
        , A.DATA_TYPE       AS 타입
        , A.DATA_LENGTH     AS 길이
        , A.NULLABLE        AS NULL여부
FROM    ALL_TAB_COLUMNS A
        LEFT OUTER JOIN ALL_COL_COMMENTS(모델정보) B
        ON (A.TABLE_NAME = B.TABLE_NAME
            AND A.COLUMN_NAME = B.COLUMN_NAME)
WHERE   B.COLUMN_NAME IS NULL
OR      A.DATA_TYPE <> B.DATA_TYPE
OR      A.DATA_LENGTH <> B.DATA_LENGTH ;

```

2.5 진단 예시 > 코드 유효성

(예시 9) 주문상태코드 : 코드 유효성 품질지표를 기준으로 '주문상태코드' 항목의 프로파일 진단기법을 활용한 진단규칙 도출 예시

1. 진단대상 분석

1.1 주문 테이블 주문상태코드 항목 프로파일 분석

- '고객구분코드' 항목 프로파일(DISTINCT, COUNT) 수행

테이블명	DISTINCT	COUNT
주문	A	50
	B	1200
	FA	2

(조회 SQL - Oracle 11g 기준)

```

SELECT  주문상태코드
        , COUNT(*) AS CNT
FROM    주문
GROUP BY 주문상태코드 ;

```

1.2 공통 코드 데이터 구성 확인

- 공통코드 테이블에 정의 된 '주문상태코드' 항목의 값은 'A', 'B'로 정의 되어 있다.

공통코드ID	공통코드명	코드값	코드값명
1001	주문상태코드	A	주문
1001	주문상태코드	B	주문취소

(조회 SQL - Oracle 11g 기준)

```

SELECT  공통코드ID
        , 공통코드명
        , 공통코드값
        , 공통코드값명
FROM    공통코드
WHERE   공통코드명 = '주문상태코드' ;

```

2. 진단대상 분석결과 및 진단규칙 도출

- 공통코드 테이블에 정의되어 있지 않은 '주문상태코드' 항목 값은 오류로 도출한다.

(진단규칙 SQL - Oracle 11g 기준)

```

SELECT  주문.주문상태코드 AS 주문상태코드
        , COUNT(*)        AS 코드건수

```

```

FROM      주문 LEFT OUTER JOIN 공통코드
          ON ('1001' = 공통코드.분류코드
              AND 주문.주문상태코드 = 공통코드.코드값)
WHERE     공통코드.코드값 IS NULL
GROUP BY  주문.주문상태코드 ;

```

(예시 10) 대상자성별코드 : 플랫폼의 개별코드 또는 공통코드를 사용하는 컬럼의 유효성을 진단하여 코드 테이블에 목록으로 정의되어 있지 않은 값을 오류 데이터로 도출 예시

1. 진단대상 분석

1.1 개별코드 또는 공통코드 테이블 확인 및 해당 코드를 사용하는 테이블과 컬럼 확인

- 목록으로 정의된 코드별로 해당 코드를 사용하는 테이블 및 컬럼 매칭

코드 테이블명	코드값	코드명	사용 테이블	사용 컬럼
1001	M	남자	InBody 데이터	대상자 성별코드
1001	F	여자		

(조희 SQL1 - PostgreSQL 기준)

```

SELECT 'INBODY 데이터' as TableName, '대상자성별코드' as ColName, count(*) as errCnt
from INBODY 데이터 where 대상자성별코드 not in (select 코드값 from 공통코드테이블 where
코드명 = 'age_group_cd')

```

(예시 11) 시도코드 : 플랫폼의 코드항목에 정의된 유효한 코드값이 아닌 경우 오류로 도출하는 예시

1. 진단대상 분석

1.1 시도단위소득정보 테이블의 시도코드 항목 프로파일 분석

- "공통코드테이블"의 유효한 시도코드값 확인
- 시도코드는 공통코드테이블에 있는 유효한 시도코드값이어야 한다.
- 표준코드정의서 정보를 참고하여 '시도코드' 항목에 정의된 유효한 코드값이 아닌 경우 오류로 도출하는 진단 규칙을 적용한다.
- 시도코드는 유효한 시도코드값으로 존재하여야 한다.

코드 테이블명	사용 컬럼	값
시도단위소득정보	시도코드	코드

(조희 SQL1 - MySQL 기준)

```

select 시도단위소득정보, 기준년월, 시도코드, 시도코드 ERR_VAL from 시도단위소득정보 where 시
도코드 not in
(select 시도코드값 from 공통코드테이블 where 시도코드값 =시도코드) and 시도코드 is not null'

```

(예시 12) 차로유형구분코드 : 플랫폼의 테이블에 정의되어 있는 값이 아니면 오류로 도출하는 진단 규칙을 적용하는 예시

1. 진단대상 분석

1.1 "차로유형구분코드"의 값은 "공통_코드" 테이블에 정의되어 있는 값인지 확인

- "차로유형구분코드"의 값은 "공통_코드" 테이블에 정의되어 있는 값이 아니면 오류로 도출하는 진단규칙을 적용한다.
- "차로유형구분코드"의 값은 "공통_코드" 테이블에 정의된 값이어야 한다.

코드 테이블명	사용 컬럼	비고
고속도로_콘존_차로유형별_교통소통(1일 단위)	차로유형구분코드	

(조희 SQL1 - Oracle 11g 기준)

```

SELECT 집계년월일,차로유형구분코드,차로유형구분코드
FROM 고속도로_콘존_차로유형별_교통소통(1일 단위)
WHERE 차로유형구분_코드 NOT IN
(SELECT 코드값 FROM 공통_코드 WHERE 코드분류명 ='EX' AND 코드명 ='차로유형구분')

```

2.6 진단 예시 > 형식 유효성

(예시 13) 주문번호 : 형식 유효성 품질지표를 기준으로 '주문번호' 항목의 프로파일 진단기법을 활용한 진단규칙 도출 예시

1. 진단대상 분석

1.1 주문 테이블 주문번호 항목 프로파일 분석

- '주문번호' 항목 프로파일(Pattern) 수행

테이블명	컬럼명	PATTERN	길이	COUNT
주문	주문번호	99999XXXXX	10	1500
		9999XXXXXX	10	7

(조희 SQL - Oracle 11g 기준)

```

SELECT      LENGTH(주문번호)      AS 길이
            , TRANSLATE (주문번호, '0123456789ABCDE...abcde...'
            , '9999999999XXXXX...XXXXX...' ) as PATTERN
            , COUNT (*) AS CNT
FROM        주문
GROUP BY    LENGTH(주문번호)
            , TRANSLATE (주문번호, '0123456789ABCDE...abcde...'

```

, '9999999999XXXXX...XXXXX...')

1.2 주문 테이블 데이터 구성 확인

- 주문 테이블의 '주문번호' 항목 값은 아래와 같이 구성 되어 있다.

주문번호	고객번호	주문일자	주문상태코드
00015DAAAL	E0006	20191101	02
0267AAQBG	E0008	20191214	01

(조회 SQL - Oracle 11g 기준)

```
SELECT  주문번호
        , 고객번호
        , 주문일자
        , 주문상태코드
FROM    주문 ;
```

1.3 주문번호 생성 규칙 확인

- ERD, 업무규칙 정의서, 업무담당자 등 다양한 경로의 정보를 참고하여 주문번호 생성 규칙을 확인 할 수 있다.

항목명	데이터 발생 규칙
주문번호	영문5자리+숫자5자리(예시: 99999ZZZZZ)

2. 진단대상 분석 결과 및 진단규칙 도출

- 주문번호 항목의 프로파일 및 생성 규칙 확인 결과 주문번호는 '영문5자리 + 숫자5자리'로 구성되어야 한다. 이에 규칙에 위배되는 값은 오류로 도출한다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT  주문번호
        , 고객번호
        , 주문일자
        , 주문상태코드
FROM    주문
WHERE   REGEXP_REPLACE(주문번호,'[0-9]{5}[a-z|A-Z]{5}',") > '0' ;
```

(예시 14) (채번)일련번호/전화번호/우편번호 : 형식 유효성 진단 대상인 해당 컬럼별로 정해진 포맷에 맞게 값이 입력되어 있는지를 진단하여 포맷에 맞지 않는 데이터를 오류 데이터로 도출 예시

1. 진단대상 분석

1.1 정해진 포맷에 맞게 값이 입력되어 있는지를 진단하여 포맷에 맞지 않는 데이터 분석

- 정해진 입력포맷을 가지고 있는 컬럼의 값 패턴 및 패턴별 오류데이터 도출을 위한 정규식

테이블명	PATTERN	길이	COUNT
첨부파일 정보	첨부파일정보 연결일련번호	D_X(숫자17자리)	^D_{0-9}(17)\$
교육인력 정보 관리	교육인력전화번호	XXX-XXX-XXXX	^[0-9](0,4)-?[0-9](3,4)-?[0-9](4)\$
산림복지 프로그램 수행장소	수행장소우편번호	신우편번호(5자리) 구우편번호(6자리)	^[0-9](5,6)\$

(조회 SQL1 - PostgreSQL 기준)

```
SELECT '첨부파일 정보' as TableName, '첨부파일정보  
연결일련번호' as ColName, count(*) as errCnt from 첨부파일 정보 where 첨부파일정보 연결일련  
번호 !~ '^D_{0-9}(17)$'
```

(조회 SQL2 - PostgreSQL 기준)

```
SELECT '교육인력 정보 관리' as TableName, '교육인력전화번호' as ColName, count(*) as  
errCnt from 교육인력 정보 관리 where 교육인력전화번호 !~ '^[0-9](0,4)-?[0-9](3,4)-?[0-9](4)$'
```

(예시 15) 우편번호 : 우편번호 형식에 맞지 않는 데이터를 오류 데이터로 도출 예시

1. 진단대상 분석

1.1 계약종별 공실현황 주택용 테이블의 '우편번호' 항목의 프로파일 검사

- '우편번호' 항목의 형식 유효성에 대한 프로파일(패턴) 확인
- '우편번호'는 숫자5자리 또는 숫자6자리로 유효한 우편번호 형식이 아닌 경우 오류로 도출하는 진단 규칙을 적용한다.
- 우편번호는 숫자5자리 또는 숫자6자리이어야 한다.

테이블명	컬럼명	값
계약종별 공실현황 주택용	우편번호	우편번호

(조회 SQL1 - MySQL 기준)

```
select 계약종별 공실현황 주택용, 지역명, 우편번호, 우편번호 ERR_VAL from 계약종별 공실현황 주  
택용 where not 우편번호 regexp('^([0-9]{5,6})$')
```

(예시 16) 가맹점전화번호 : 전화번호 형식에 맞지 않는 데이터를 오류 데이터로 도출 예시

1. 진단대상 분석

1.1 가맹점정보 테이블의 '가맹점전화번호' 프로파일 검사

- '우편번호' 항목의 형식 유효성에 대한 프로파일(패턴) 확인
- '가맹점전화번호' 항목의 형식 유효성에 대한 프로파일(패턴) 확인
- '가맹점전화번호'는 999-9999-9999 등 유효한 전화번호 형식이 아닌 경우 오류로 도출하는 진단 규칙을 적용한다.

○ 가맹점전화번호는 전화번호 형식이어야 한다.

테이블명	컬럼명	비고
가맹점정보	가맹점전화번호	

(조회 SQL1 - PostgreSQL 기준)

```
select 가맹점정보, 가맹점명, 가맹점전화번호, 가맹점전화번호 ERR_VAL from 가맹점정보 where
not 가맹점전화번호
~'^((01|016789){1}|02|03-9){1}[0-9]{1})-[0-9]{3,4}-[0-9]{4})|((1577|1588|1899|1544|1644|16
11|1566|1600|1668|1661|1670|1688|1666|1599|1877|1855|1800|1881|1833|1811|1522)-[0
-9]{4})$' AND 가맹점전화번호 !="
```

(예시 17) CAS등록번호 : 등록번호 형식에 맞지 않는 데이터를 오류 데이터로 도출 예시

1. 진단대상 분석

1.1 CAS등록번호 값이 해당 도메인의 데이터 구성 형식(Pattern)을 준수하는지 진단

○ CAS등록번호 항목의 프로파일 및 생성 규칙 확인 결과 CAS등록번호는 "숫자2자리 + '-' + 숫자2자리 + '-' + 숫자"로 구성되어야 한다. 이에 규칙에 위배되는 값은 오류로 도출한다.

테이블명	컬럼명	비고
생태독성	CAS등록번호	

(조회 SQL1 - PostgreSQL 기준)

```
select CAS등록번호, 고유번호, 항목, 연도 from 생태독성 where not CAS등록번호
~'^[0-9]{2}-[0-9]{2}-[0-9]+$'; AND 등록번호 != "
```

2.7 진단 예시 > 여부 유효성

(예시 18) 통합고객등록여부 : 여부 유효성 품질지표를 기준으로 '통합고객등록여부' 항목의 프로파일 진단기법을 활용한 진단규칙 도출 예시

1. 진단대상 분석

1.1 고객 테이블 통합고객등록여부 항목 프로파일 분석

○ '통합고객등록여부' 항목 프로파일링(DISTINCT, COUNT) 수행

테이블명	통합고객등록여부	CNT
고객	Y	650
	N	400
	1	13

(조회 SQL - Oracle 11g 기준)

```
SELECT 통합고객등록여부
, COUNT(*) AS CNT
FROM 고객
GROUP BY 통합고객등록여부 ;
```

1.2 여부 도메인 생성 규칙 확인

○ ERD, 도메인정의서 등 다양한 경로의 정보를 참고하여 통합고객등록여부 항목의 생성 규칙을 확인할 수 있다.

도메인명	TYPE	길이	정의
여부	VARCHAR2	1	유효값은 'Y', 'N'로 정의한다.

2. 진단대상 분석 결과 및 진단규칙 도출

○ 통합고객등록여부 항목의 프로파일 및 생성 규칙 확인 결과 'Y', 'N'으로 구성되어야 한다. 규칙에 위배되는 값은 오류로 도출한다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT 고객번호
, 통합고객등록여부
FROM 고객
WHERE 통합고객등록여부 NOT IN ('Y', 'N') ;
```

(예시 19) 결혼여부명/자녀여부명 : 여부 유효성 진단 대상 컬럼의 여부('Y', 'N' 또는 '0', '1')값 정의를 기준으로 유효성을 진단하여 오류 데이터를 도출 예시

1. 진단대상 분석

1.1 컬럼의 여부 항목 프로파일 분석

○ 여부 유효성 진단 대상 컬럼의 여부('Y', 'N' 또는 '0', '1')값 정의

테이블명	컬럼명	정의 여부 값
귀산촌 이용자 정보	결혼여부명	Y, N
	자녀여부명	Y, N

(조회 SQL - PostgreSQL 기준)

```
SELECT '귀산촌 이용자 정보' as TableName, '결혼여부명' as ColName, count(*) as errCnt
```

```
from 귀산촌 이용자 정보 where 결혼여부명 not in ('Y','N')
```

(예시 20) 주소표준화 : 여부 유효성 진단 대상 컬럼의 여부('0','1')값 정의를 기준으로 유효성을 진단하여 오류 데이터를 도출 예시

1. 진단대상 분석

1.1 주소표준화데이터셋 테이블의 '주소산여부' 항목 프로파일 검사

- '주소산여부' 항목의 생성규칙 확인
- '주소산여부' 항목에 유효값을 확인한 결과 '0' 또는 '1'이 아닌 경우 오류로 도출하는 진단 규칙을 적용한다.

테이블명	컬럼명	여부 값
신차등록지역별대수	주소산여부	0,1

(조회 SQL - MySQL 기준)

```
select 주소표준화데이터셋, 건물관리번호, 주소산여부, 주소산여부 오류건수 from 주소표준화데이터셋 where not 주소산여부 regexp('^(0|1)$')
```

2.8 진단 예시 > 날짜 유효성

(예시 21) 고객등록일자 : 날짜 유효성 품질지표를 기준으로 '고객등록일자' 항목의 프로파일 진단기법을 활용한 진단규칙 도출 예시

1. 진단대상 분석

1.1 고객 테이블 고객등록일자 항목 프로파일 분석

- '고객등록일자' 항목 프로파일(Pattern) 수행

테이블명	컬럼명	PATTERN	COUNT
고객	고객등록일자	YYYY-MM-DD	3
		YYYYMMDD	950

(조회 SQL - Oracle 11g 기준)

```
SELECT TRANSLATE (고객등록일자, '0123-45-67'
, '9999-99-99' ) as PATTERN
, COUNT (*) AS CNT
FROM 고객
GROUP BY TRANSLATE (고객등록일자, '0123-45-67'
, '9999-99-99' ) ;
```

1.2 고객 테이블 데이터 구성 확인

- 고객 테이블의 '고객등록일자' 항목의 값은 아래와 같다.

고객번호	고객구분코드	고객등록일자
E0004	C01	2019-01-03
E0019	C01	20191205
E0250	C01	20190101

(조회 SQL - Oracle 11g 기준)

```
SELECT 고객번호
, 고객구분코드
, 고객등록일자
FROM 고객 ;
```

1.3 등록일자 생성 규칙 확인

- ERD, 업무규칙 정의서, 업무담당자 등 다양한 경로의 정보를 참고하여 등록일자 생성 규칙을 확인 할 수 있다.

항목명	데이터 발생 규칙
등록일자	YYYYMMDD 형식을 준수한다

2. 진단대상 분석 결과 및 진단규칙 도출

- '고객등록일자' 항목의 프로파일 결과 및 생성 규칙 확인 결과 고객등록일자 항목의 값 중 YYYYMMDD 형식에 위배되는 데이터는 오류로 도출한다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT 고객번호
, 고객등록일자
FROM 고객
WHERE (NOT REGEXP_LIKE(고객등록일자, '[0-9]')
OR NOT REGEXP_LIKE(고객등록일자, '^([0-9]([0-9]([0-9]([0-9]([0-2]([0-9])|([1]([0-2]([3]([0-1])$))
OR REGEXP_LIKE(고객등록일자, '[0230]0231|0431|0631|0931|1131)$');
```

※ 본 예시는 윤달에 의한 조건은 예외로 작성

(예시 22) 년도/월/년월/시간(일자시분초) : 날짜 유효성 품질지표를 기준으로 정규식을 활용해서 날짜 형식에 따른 오류 데이터의 진단규칙 도출 예시

1. 진단대상 분석

1.1 날짜 형식 컬럼의 값 패턴 및 패턴별 오류데이터 도출을 위한 정규식

테이블명	컬럼명	PATTERN	정규식
곤충 종	명명년도	YYYY	^(19 20)[0-9]{2}\$
농림산물수출입 통계데이터	기준월	MM	^(0[1-9] 1[012])\$
생산면적통계정보	조사년월	YYYYMM	^(19 20)[0-9]{2}(0[1-9] 1[012])\$
공유자전거 대여소정보	공유자전거대여소 데이 터기준일자	YYYY-MM-DD	^(19 20)[0-9]{2}-(0[1-9] 1[012])-(0 [1-9] 1[0-9] 2[0-9] 3[0-1])\$
식물사진정보	산림식물자원 사진촬영시간	YYYYMMDD HH24MISS	^(19 20)[0-9]{2}(0[1-9] 1[012])(0[1 -9] 1[0-9] 2[0-9] 3[0-1])(0[0-9] 1[0-9] 2[0-3])([0-5][0-9])([0-5][0-9]) \$

(조회 SQL - PostgreSQL 기준)

```
SELECT '곤충 종' as TableName, '명명년도' as ColName, count(*) as errCnt from 곤충
종 where 명명년도 !~ '^(19|20)[0-9]{2}$'
```

(예시 23) 발생년도 : 일자 도메인 데이터의 값이 해당 도메인의 일자 형식을 준수하는지
진단규칙 도출 예시

1. 진단대상 분석

1.1 일자 도메인 데이터의 값이 해당 도메인의 일자 형식을 준수하는지 진단

- 발표일자 항목의 프로파일 및 생성 규칙 확인결과 YYYY-MM-DD 으로 구성되어야 한다. 규칙에 위배
되는 값은 오류로 도출한다.
- YYYY-MM-DD 형식의 값이어야 한다.

테이블명	컬럼명	PATTERN	비고
가뭄예경보정보	발표일자	YYYY-MM-DD	

(조회 SQL - PostgreSQL 기준)

```
SELECT
  발표일자, count(*)
FROM
  가뭄예경보정보 T1
WHERE
  COALESCE(발표일자, '') = ''
  AND CASE
    WHEN LENGTH(CAST(TRIM(CAST(발표일자 AS VARCHAR)) AS varchar)) =
      10
    AND
      CAST(발표일자
      AS
      VARCHAR) ~
```

```
'^[0-9][0-9][0-9][0-9]-[0-9][0-9]-[0-9][0-9]$'
  AND SUBSTR(TRIM(CAST(발표일자 AS VARCHAR)),1,4) BETWEEN '0001'
  AND '9999'
  AND SUBSTR(TRIM(CAST(발표일자 AS VARCHAR)),6,2) BETWEEN '01' AND
  '12' THEN CASE
    WHEN SUBSTR(TRIM(CAST(발표일자 AS VARCHAR)),9,2) BETWEEN '01'
  AND CASE
    WHEN SUBSTR(TRIM(CAST(발표일자 AS VARCHAR)),6,2) IN (
      '01','03','05','07','08','10','12'
    ) THEN '31'
    WHEN SUBSTR(TRIM(CAST(발표일자 AS VARCHAR)),6,2) IN (
      '04','06','09','11'
    ) THEN '30'
    ELSE CASE
      WHEN
        MOD(CAST(SUBSTR(TRIM(CAST(발표일자 AS
  VARCHAR)),1,4) AS DECIMAL), 400) = 0 THEN '29'
      WHEN
        MOD(CAST(SUBSTR(TRIM(CAST(발표일자 AS
  VARCHAR)),1,4) AS DECIMAL), 100) = 0 THEN '28'
      WHEN
        MOD(CAST(SUBSTR(TRIM(CAST(발표일자 AS
  VARCHAR)),1,4) AS DECIMAL), 4) = 0 THEN '29'
      ELSE '28'
    END
    END THEN 1
    ELSE 0
  END
  ELSE 0
END = 0
GROUP BY
  T1.발표일자 ;
```

2.9 진단 예시 > 범위 유효성

(예시 24) 할인가격 : 범위 유효성 품질지표를 기준으로 '할인가격' 항목의 프로파일 진단기법을 활용
한 진단규칙 도출 예시

1. 진단대상 분석

1.1 주문상세 테이블 '할인가격' 항목 프로파일 분석

- '통합고객등록여부' 항목 프로파일(DISTINCT,COUNT) 수행

테이블명	컬럼명	MIN(할인가격)	MAX(할인가격)
주문상세	할인가격	-12000	8770

(조회 SQL - Oracle 11g 기준)

```
SELECT MIN(할인가격) AS MIN할인가격
      ,MAX(할인가격) AS MAX할인가격
FROM   주문상세 ;
```

1.2 할인가격 생성 규칙 확인

- ERD, 업무규칙정의서 등 다양한 경로의 정보를 참고하여 할인가격 항목의 생성 규칙을 확인할 수 있다.

항목명	데이터 발생 규칙
할인가격	할인가격의 범위는 0~150,000 로 정의한다.

1.3 주문상세 테이블 데이터 구성 확인

- 주문상세 테이블의 '할인가격' 항목 값은 아래와 같다.

주문번호	고객번호	판매가격	할인가격
11276AB	E0046	93850	-12000
21987AK	E0070	50000	1000
11387AZ	E0124	45000	450

2. 진단대상 분석 결과 및 진단규칙 도출

- 할인가격 항목의 프로파일 및 데이터 생성 규칙 확인 결과 범위 값 0~150,000으로 구성되어야 하며, 규칙에 위배되는 값은 오류로 도출한다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT 주문번호
      , 고객번호
      , 판매가격
      , 할인가격
FROM   주문상세
WHERE 할인가격 NOT BETWEEN 0 AND 150000 ;
```

(예시 25) 캠핑장화장실수/10미터기온/기관일련번호 : 범위 유효성의 진단 대상인 해당 컬럼별로 입력 값이 유효 범위 내에 존재하는지를 진단하여 존재하지 않는 데이터를 오류 데이터로 도출 예시

1. 진단대상 분석

1.1 컬럼의 데이터 타입이 숫자인 경우 MIN, MAX함수를 활용해 입력 값의 프로파일 분석

- 입력 값이 유효 범위 내에 존재하는지를 진단 수행

테이블명	컬럼명	MIN	MAX
전국캠핑장정보	캠핑장화장실수	0	40
산악기상데이터연별	10미터기온	-6.5	32.3

(조회 SQL - PostgreSQL 기준)

```
SELECT '전국캠핑장정보' as TableName, '캠핑장화장실수' as ColName, MIN(캠핑장화장실수) as
MinValue, MAX(캠핑장화장실수) as MaxValue from 전국캠핑장정보
```

1.2 컬럼의 데이터 타입이 문자인 경우 컬럼의 입력 값이 숫자로만 이뤄졌는지 먼저 확인 후 MIN, MAX 함수를 활용해 프로파일 분석

- 컬럼의 데이터 타입이 문자형인 경우 입력 값이 숫자로만 이뤄졌는지 확인

테이블명	컬럼명	PATTERN	정규식
산림치유지도사 자격증 현황	자격증양성기관 일련번호	123456789	^[0-9]*\$

(조회 SQL - PostgreSQL 기준)

```
SELECT '산림치유지도사 자격증 현황' as TableName, '자격증양성기관
일련번호' as ColName, count(*) as errCnt from 산림치유지도사 자격증 현황 where 자격증양성기
관일련번호 !~ '^[0-9]*$'
```

(예시 26) 시간대별이용자비율 : 시간대이용자비율이 비율값 내에 존재하는지를 진단하여 존재하지 않는 데이터를 오류 데이터로 도출 예시

1. 진단대상 분석

1.1 자동차접이용행태 테이블 '시간대별이용자비율' 항목 프로파일 분석

- '시간대별이용자비율' 항목 프로파일 확인
- '시간대별이용자비율' 항목의 프로파일 및 데이터 생성규칙 확인 결과 데이터값은 0.0~100.0 범위에 있지 않는 경우 오류로 도출하는 진단 규칙을 적용한다.
- 비율값은 0.0~100.0 사이에 존재하여야 한다.

테이블명	컬럼명	비율값
자동차접이용행태	시간대별이용자비율	0.0~100.0

(조회 SQL - MySQL 기준)

```
select 자동차접이용행태, 시간대별이용자비율, 시간대별이용자비율 ERR_VAL from 자동차접이용행태
where 시간대별이용자비율 not between 0.0 and 100.0 and 시간대별이용자비율 !="
```

(예시 27) 휴업폐업1등급비율 : 비율이 숫자 형식으로 존재하는지를 진단하여 존재하지 않는 데이터를 오류 데이터로 도출 예시

1. 진단대상 분석

1.1 휴폐업데이터 테이블 '휴업폐업1등급비율' 항목 프로파일 분석

- '휴업폐업1등급비율' 항목 프로파일(숫자 형식) 확인
- '휴업폐업1등급비율' 항목의 프로파일 및 데이터 생성규칙 확인 결과 데이터값이 숫자가 아닌 문자 형식의 값이 있는 경우 오류로 도출하는 진단 규칙을 적용한다.
- 비율은 숫자형식이어야 한다.

테이블명	컬럼명	값
휴폐업 데이터	휴업폐업1등급비율	수

(조회 SQL - PostgreSQL 기준)

```
select 휴폐업 데이터, 기준일자, 휴업폐업1등급비율, 휴업폐업1등급비율 ERR_VAL from 휴폐업 데이터 where not 휴업폐업1등급비율 ~'^[0-9\.\-]+$', and 휴업폐업1등급비율 !="
```

(예시 28) 생체검사조직학적등급수 : 값이 양수 형식으로 존재하는지를 진단하여 존재하지 않는 데이터를 오류 데이터로 도출 예시

1. 진단대상 분석

1.1 "생체검사조직학적등급수" 값은 양수 값인지 확인

- "생체검사조직학적등급수" 값이 양수가 아니면 오류로 도출하는 진단규칙을 적용한다.
- 값은 양수 값이어야 한다.

테이블명	컬럼명	값
유방암_병리_생검	생체검사조직학적등급수	수

(조회 SQL - Oracle 11g 기준)

```
SELECT 환자대체번호, 병리번호, 생체검사조직학적등급수
FROM 유방암_병리_생검
WHERE 생체검사조직학적등급수 < 0
```

(예시 29) 온도 : 값이 업무요건에 정해진 유효범위 내에 존재하는지를 진단하여 존재하지 않는 데이터를 오류 데이터로 도출 예시

1. 진단대상 분석

1.1 값이 업무 요건에 의해 정해진 유효 범위 내에 존재하는지 진단

- '온도' 항목의 프로파일 및 생성 규칙 확인결과 온도는 -30이상 50이하의 온도이어야 한다.

테이블명	컬럼명	값
환경별실외미세먼지데이터	온도	-30이상 50이하

(조회 SQL - PostgreSQL 기준)

```
SELECT
    온도 AS ERR_DATA,
    COUNT(*) AS ERR_CNT
```

```
from 환경별실외미세먼지데이터
WHERE NOT ( 온도 )= -30
        AND 온도 <= 50 ) and COALESCE(CAST(온도 AS TEXT),) = "
GROUP by 온도 ;
```

2.10 진단 예시 > 선후관계 유효성

(예시 30) 주문일자(-)배송일자 : 선후행 품질지표의 프로파일 진단기법을 활용한 진단규칙 도출 예시

1. 진단대상 분석

1.1 하나의 테이블에 '날짜' 도메인유형 항목이 2개 이상인 대상 확인

테이블명	컬럼명	데이터타입	데이터길이
주문상세	주문일자	VARCHAR2	14
	배송일자	VARCHAR2	8
	취소일자	VARCHAR2	8

(조회 SQL - Oracle 11g 기준)

```
SELECT A.TABLE_NAME, COLUMN_NAME, DATA_TYPE, DATA_LENGTH
FROM ALL_TAB_COLUMNS A
    INNER JOIN (SELECT OWNER, TABLE_NAME, COUNT(*) AS CNT
                FROM ALL_TAB_COLUMNS
                WHERE OWNER='TEST_SVR'
                AND REGEXP_LIKE(COLUMN_NAME,
                                '_DT|_DTM|_DAY|_YM|_MM$')
                GROUP BY OWNER, TABLE_NAME
            ) B
ON (A.OWNER = B.OWNER
    AND A.TABLE_NAME = B.TABLE_NAME)
WHERE B.CNT > 1
AND REGEXP_LIKE(COLUMN_NAME, '_DT|_DTM|_DAY|_YM|_MM$');
```

1.2 주문상세 테이블 데이터 구성 확인

- 배송일자는 주문일자 항목보다 앞서는 값이 존재할 수 없다.

주문번호	고객번호	주문일자	배송일자
00005DAAAB	E0004	20190421	20190101
00011AAABD	E0145	20191210	20191212
000125AOADB	E0009	20191011	20191118

(조회 SQL - Oracle 11g 기준)


```

SELECT   주문번호
        , 고객번호
        , 주문일자
        , 배송일자
FROM     주문상세 ;

```

2. 진단대상 분석 결과 및 진단규칙 도출

- 주문상세 테이블의 '주문일자', '배송일자' 항목의 데이터를 조회한 결과 '배송일자'항목은 '주문일자' 항목보다 앞서는 값이 존재하는 경우 오류로 도출한다.

(진단규칙 SQL - Oracle 11g 기준)

```

SELECT   주문번호
        , 고객번호
        , 주문일자
        , 배송일자
FROM     주문상세
WHERE    배송일자 IS NOT NULL
AND      주문일자 > 배송일자 ;

```

(예시 31) 생체검사판독일자 : 값이 업무요건에 정해진 선후관계 일자보다 크거나 같은지를 진단하여 선후관계 유효성에 벗어나는 데이터를 오류 데이터로 도출 예시

1. 진단대상 분석

1.1 "판독일자"는 "접수일자" 보다 크거나 같은 값인지 확인

- "판독일자"가 "접수일자" 보다 크거나 같지 않으면 오류로 도출하는 진단규칙을 적용한다.
- "유방암_병리_생검"의 "생체검사판독일자"는 "병리생체검사접수일자" 보다 크거나 같아야 한다.

테이블명	컬럼명	값
유방암_병리_생검	생체검사판독일자	일자

(조회 SQL - Oracle 11g 기준)

```

SELECT   환자대체번호,병리번호,생체검사판독일자
FROM     유방암_병리_생검
WHERE    생체검사판독일자 < 병리생체검사접수일자

```

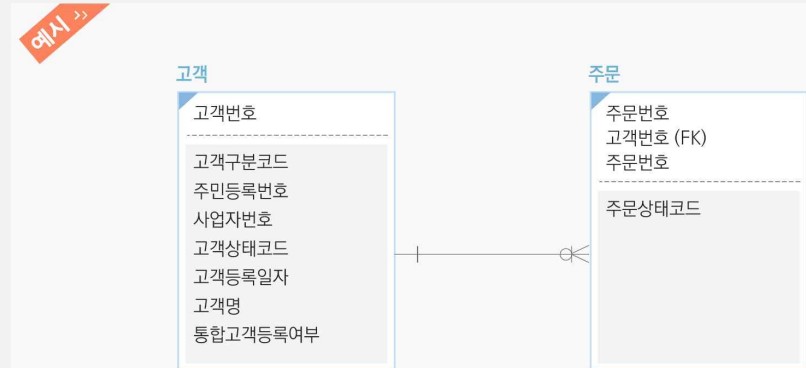
2.11 진단 예시 > 참조 무결성

(예시 32) 고객번호 : 참조 무결성 품질지표를 기준으로 '고객번호(FK)'항목의 업무규칙 진단기법을 활용한 진단규칙 도출 예시

1. 진단대상 분석(업무규칙)

1.1 ERD 모델 및 데이터 규칙정의/업무규칙 정보 확인

- ERD 모델 관계선 및 업무적으로 발생 되는 데이터발생 규칙 등을 참고하여 진단규칙을 도출 할 수 있다.



1.2 주문, 고객 테이블 데이터 확인

주문테이블			고객테이블
주문번호	고객번호	주문일자	고객번호
00005DAAAB	00005	20190503	00005
00006DAAAC	00007	20190803	00007
00003GGGGB	00008	20191123	

(조회 SQL - Oracle 11g 기준)

```

SELECT   주문.주문번호
        , 주문.주문일자
        , 주문.주문고객번호
        , 고객.고객번호
FROM     주문 LEFT OUTER JOIN 고객
        ON (주문.주문고객번호 = 고객.고객번호)

```

2. 진단대상 분석 결과 및 진단규칙 도출

Foreign Key 제약조건이 없는 항목을 대상으로 하여 ERD 논리 모델을 분석하여 진단 규칙을 도출한다.

(진단규칙 SQL - Oracle 11g 기준)

```

SELECT   고객.고객번호
        , 고객.고객등록일자
        , 고객.고객명
        , 주문.주문번호
        , 주문.주문일자
        , 주문.주문고객번호

```

```

FROM      주문 LEFT OUTER JOIN 고객
          ON (주문.주문고객번호 = 고객.고객번호)
WHERE     고객.고객ID IS NULL ;

```

(예시 33) 환자대체번호 : 참조 무결성 품질지표를 기준으로 상속받은 대체번호 값이 아니면 오류로 도출하는 진단규칙 도출 예시

1. 진단대상 분석

1.1 "환자대체번호" 값이 상속받은 부모 테이블에 존재하는 값인지 확인

- "유방암_환자_건강정보" 의 환자대체번호는 "유방암_환자_기본정보"로 부터 상속받은 환자대체번호 값이 아니면 오류로 도출하는 진단규칙을 적용한다.
- "유방암_환자_건강정보"의 "환자대체번호"가 "유방암_환자_기본정보"에 존재하여야 한다.

테이블명	컬럼명	값
유방암_환자_건강정보	환자대체번호	번호

(조회 SQL - Oracle 11g 기준)

```

SELECT 환자대체번호, 기록일자
FROM 유방암_환자_건강정보
WHERE (환자대체번호) NOT IN
      (SELECT 환자대체번호 FROM 유방암_환자_기본정보 )

```

(예시 34) CAS등록번호 : 기본 키와 참조 키 간의 관계가 항상 유지되는지 진단 도출 예시

1. 진단대상 분석

1.1 기본 키와 참조 키 간의 관계가 항상 유지되는지 진단

- 외래 키 컬럼이 데이터 값이 참조하는 Parent 테이블에 존재하지 않는 데이터 값이 있는지 진단
- CAS등록번호 이어야한다.

테이블명	컬럼명	값
화학물질독성정보	CAS등록번호	CAS번호

(조회 SQL - PostgreSQL 기준)

```

SELECT
  T1.CAS등록번호
, T1.성명내용
, T1.계통
, T2.화학물질국문
, T2.화학물질영문
from 화학물질독성정보 T1

```

```

LEFT OUTER JOIN
  화학물질기본물질목록 T2
  ON   T1.CAS등록번호 = T2.CAS등록번호
WHERE T2.CAS등록번호 IS NULL

```

2.12 진단 예시 > 중복 일관성

(예시 35) 통합고객등록여부 : 중복 일관성 품질지표를 기준으로 '통합고객등록여부' 항목의 업무규칙 진단기법에 의한 진단규칙 도출 예시

1. 진단대상 분석(업무규칙)

1.1 ERD 모델 및 정의서/업무규칙 정보 확인

- ERD & 속성(컬럼) 정의서에 정의된 반정규화 관련 내용 및 업무담당자, 운영담당자 등 관련자들의 정보를 참고하여 진단규칙을 도출할 수 있다.
- 업무규칙 정보를 확인한 결과 주문상세 테이블의 '고객번호', '통합고객등록여부' 항목의 원천데이터는 고객 테이블의 '고객번호', '통합고객등록여부' 항목임이 확인 되었다.



1.2 주문상세 및 고객 테이블 데이터 확인

주문상세 테이블			고객 테이블	
주문번호	고객번호	통합고객등록여부	고객번호	통합고객등록여부
00001AAAAA	00001	Y	00001	.
00002CCCCA	00002	Y	00002	Y
00003GGGGB	00003	Y	00003	Y
11236SDFGH	00005	Y	00005	N

(조회 SQL - Oracle 11g 기준)

```

SELECT  주문상세.주문번호

```

```

, 주문상세.주문고객번호
, 주문상세.통합고객등록여부
, 고객.고객번호
, 고객.통합고객등록여부
FROM 주문상세 INNER JOIN 고객
ON (주문상세.주문고객번호 = 고객.고객번호)

```

2. 진단대상 분석 결과 및 진단규칙 도출

ERD & 속성(컬럼) 정의서에 정의 된 반정규화 관련 내용 및 업무담당자, 운영담당자 등 관련자들의 정보 바탕으로 주문 테이블의 '주문고객번호', '주문통합고객등록여부' 항목의 원천은 고객 테이블의 '고객번호', '통합고객등록여부' 항목으로 동일한 시점에 원천데이터와 중복데이터가 같은 값을 가지고 있어야 함을 진단 규칙으로 도출할 수 있다.

(진단규칙 SQL - Oracle 11g 기준)

```

SELECT 주문고객번호
, 주문_통합고객등록여부
, 고객_통합고객등록여부
FROM 주문상세 INNER JOIN 고객
ON (주문상세.주문고객번호 = 고객.고객번호)
WHERE 주문상세.통합고객등록여부 <> 고객.통합고객등록여부 ;

```

(예시 36) 사망일자 : 같은 일자를 관리하는 모든 테이블에 일관성 있는 값으로 존재하는지 진단 도출 예시

1. 진단대상 분석

1.1 같은 환자에 대한 "사망일자"는 "사망일자"를 관리하는 모든 테이블에 일관성 있는 값으로 존재하는지 확인

- "사망일자"를 관리하는 모든 테이블의 값은 "위암_환자_기본정보"의 "사망일자" 값과 일치하지 않으면 오류로 도출하는 진단규칙을 적용한다.
- "위암_사망_정보"의 "사망일자"는 "위암_환자_기본정보"의 "기본환자사망일자" 값과 일치해야 한다.

테이블명	컬럼명	값
위암_사망_정보	사망일자	일자

(조회 SQL - Oracle 11g 기준)

```

SELECT A.환자대체번호,A.사망일자, B.기본환자사망일자
FROM 위암_사망_정보 A INNER JOIN T_YYY B
ON A.환자대체번호 = B.환자대체번호
WHERE A.사망일자 IS NOT NULL
AND A.사망일자 != B.기본환자사망일자
GROUP by T1.CAS등록번호 ;

```

2.13 진단 예시 > 표준 일관성

(예시 37) 표준사전(->)DB스키마 : 표준 일관성 품질지표를 기준으로 '표준관리', 'DBMS 테이블' 간 프로파일 진단기법을 활용한 진단규칙 도출 예시

1. 진단대상 분석

1.1 표준관리(표준용어사전) 문서 확인

표준용어	표준도메인	데이터 타입	데이터 길이
휴대폰번호	전화번호	VARCHAR2	20
집전화번호	전화번호	VARCHAR2	20

1.2 DBMS 테이블 정보

컬럼명	데이터 타입	데이터 길이
핸드폰전화번호	VARCHAR2	20
집전화번호	VARCHAR2	20

2. 진단대상 분석 결과 및 진단규칙 도출(업무규칙)

- 표준관리(표준용어사전) 문서와 DBMS 테이블 정보 분석 결과 표준용어사전에 등록되어 있지 않은 항목을 컬럼명으로 사용하고 있음이 확인되어 오류로 도출한다.

(진단규칙 SQL - Oracle 11g 기준)

```

SELECT B.테이블명
, B.컬럼명
FROM 표준용어사전 A, DBMS테이블정보 B
WHERE A.표준용어 != B.컬럼명 ;
※ 본 예시는 표준관리 문서가 DB에 적재 된 경우를 가정

```

2.14 진단 예시 > 논리관계 정확성

(예시 38) 상품(공급업체ID, 색상코드) : 논리관계 일관성 품질지표를 기준으로 '상품명', '공급업체명' 항목의 업무규칙 진단기법에 의한 진단규칙 도출 예시

1. 진단대상 분석(업무규칙)

1.1 ERD 모델 및 데이터 규칙정의/업무규칙 정보 확인

- ERD & 속성(컬럼) 정의서에 정의 된 테이블/컬럼 내용 및 업무담당자, 운영담당자 등 관련자들의 정보를 참고하여 진단 규칙을 도출 할 수 있다.
- '상품'테이블의 '상품명'이 본체인 상품 중 화이트(WH) 색상인 상품은 MS 사에서만 공급한다 라는 업무규칙이 확인 되었다.

1.2 '공급업체'테이블 정보 확인

공급업체ID	공급업체명
15476	기린전자
18555	MS

(조회 SQL - Oracle 11g 기준)

```
SELECT   공급업체ID
        , 공급업체명
FROM     공급업체 ;
```

1.3 '상품'테이블 정보 확인

상품ID	공급업체ID	상품명	상품색상코드
A549	15476	키보드	BL
A550	15476	본체	WH
A551	18555	본체	WH

(조회 SQL - Oracle 11g 기준)

```
SELECT   상품ID
        , 공급업체ID
        , 상품명
        , 상품색상코드
FROM     상품 ;
```

2. 진단대상 분석 결과 및 진단규칙 도출

'상품'테이블의 정보를 확인한 결과 '상품명'이 본체인 상품 중 화이트(WH) 색상인 상품에서 'MS'사가 아닌 '기린전자'로 등록 된 값이 존재함을 확인하였다. 따라서 아래와 같은 진단규칙을 도출할 수 있다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT   상품ID
        , 공급업체ID
        , 상품명
        , 상품색상코드
        , 기준단가
        , 생산년도
FROM     상품 A
        LEFT OUTER JOIN 공급업체 B
        ON (A.공급업체ID = B.공급업체ID)
WHERE   A.상품명 = '본체'
AND     A.상품색상코드 = 'WH'
WHERE   B.공급업체명 <> 'MS';
```

(예시 39) 배송비여부/배송비 : 논리관계 일관성 품질지표를 기준으로 '배송비여부' 항목과 '배송비' 항목 사이의 업무규칙 논리관계 정확성을 확인하는 진단 도출 예시

1. 진단대상 분석

1.1 일별가전상품가격현황 테이블의 '배송비여부'와 '배송비' 항목 사이의 업무규칙 진단

- "배송비여부" 항목과 "배송비" 항목간의 논리관계 정확성을 확인
- "배송비여부" 항목이 'Y'인 경우 배송비는 '0'이상이고 'N'인 경우 배송비는 '0'이어야하며 이와 다른 경우에는 오류로 도출하는 진단 규칙을 적용한다.

테이블명	컬럼명	값
일별가전상품가격현황	배송비여부	여부
	배송비	수

(조회 SQL - MySQL 기준)

```
SELECT   배송비여부, 배송비
FROM     일별가전상품가격현황
WHERE   (배송비여부 = 'Y' AND 배송비 = 0)
        OR (배송비여부 = 'N' AND 배송비 <> 0)
```

(예시 40) 음주기간년수 : '음주여부' 항목이 'Y' 일 때 '음주기간년수' 값이 0 이상인지 논리관계 정확성을 확인하는 진단 도출 예시

1. 진단대상 분석

1.1 "음주여부" 항목의 값이 "Y" 일때 "음주기간년수" 의 값이 0이상 인지 확인

- "음주여부" 항목의 값이 "Y" 일때 "음주기간년수" 의 값이 0이상이 아니면 오류로 도출하는 진단규칙을 적용한다.

테이블명	컬럼명	값
유방암_환자_건강정보	음주여부	여부
	음주기간년수	수

(조회 SQL - Oracle 11g 기준)

```
SELECT   환자대체번호, 기록일자, 음주기간년수, 음주여부
FROM     유방암_환자_건강정보
WHERE   음주여부 = "Y"
        AND   음주기간년수 < 0
```

2.15 진단 예시 > 파생항목 정확성

(예시 41) 판매가격 : 파생항목 정확성 품질지표를 기준으로 '판매가격' 항목의 업무규칙 진단기법에 의한 진단규칙 도출 예시

1. 진단대상 분석(업무규칙)

1.1 ERD 모델 및 데이터 규칙정의/업무규칙 정보 확인

- 수치(수,율,량)도메인 컬럼 중 연산에 의해 데이터가 발생하는 컬럼을 대상으로 업무담당자, 운영담당자 등 관련자들의 정보를 참고하여 진단 규칙을 도출 할 수 있다.
- 주문상세 테이블의 '판매가격'항목의 연산규칙은 아래 예시와 같다.
(판매가격 연산규칙 예시 - 기준단가 X 주문개수 - 할인가격 = 판매가격)

1.2 주문상세 및 상품 테이블 정보 확인

주문상세 테이블					상품 테이블
주문번호	상품ID	판매가격	주문개수	할인가격	기준단가
11279IJIIH	A547	141230	50	8770	3000
11265SDFGH	A547	50200	17	800	3000
11260SDFGH	A547	96000	32	300	3000
11266SDFGH	A785	1200000	14	900	85000
11281IJIIH	B732	52766	54	1234	1000

(조회 SQL - Oracle 11g 기준)

```
SELECT  A.주문번호
        , A.상품ID
        , A.판매가격
        , A.주문개수
        , A.할인가격
        , B.기준단가
FROM    주문상세 A LEFT OUTER JOIN 상품 B
        ON (A.상품ID = B.상품ID
AND A.공급업체ID = B.공급업체ID) ;
```

2. 진단대상 분석 결과 및 진단규칙 도출

- 주문상세 테이블의 '판매가격'항목의 데이터 생성 규칙(연산규칙)을 기반으로 아래와 같은 진단규칙을 도출할 수 있다.

(진단규칙 SQL - Oracle 11g 기준)

```
SELECT  A.주문번호
        , A.상품ID
        , A.판매가격
        , B.기준단가 * A.주문개수 - A.할인가격 AS 연산결과
FROM    주문상세 A LEFT OUTER JOIN 상품 B
        ON (A.상품ID = B.상품ID
AND A.공급업체ID = B.공급업체ID)
WHERE   A.판매가격 <> B.기준단가 * A.주문개수 - A.할인가격 ;
```

(예시 42) 방사선치료총선량 : 파생항목 정확성 품질지표를 기준으로 '방사선치료총선량' 항목의 업무규칙 진단기법에 의한 진단규칙 도출 예시

1. 진단대상 분석

1.1 ("방사선치료1회선량" x "방사선치료총횟수")는 "방사선치료총선량"과 같은지 확인

- ("방사선치료1회선량" x "방사선치료총횟수")는 "방사선치료총선량"과 같지 않으면 오류로 도출하는 진단규칙을 적용한다.
- ("유방암_치료_방사선"의 "방사선치료1회선량" x "유방암_치료_방사선"의 "방사선치료총횟수")는 "유방암_치료_방사선"의 "방사선치료총선량"과 같아야 한다.

테이블명	컬럼명	값
유방암_치료_방사선	방사선치료총선량	수

(조회 SQL - Oracle 11g 기준)

```
SELECT  환자대체번호,방사선치료시작일자,방사선치료종료일자,(방사선치료1회선량*방사선치료총횟수),
        방사선치료총선량
FROM    유방암_치료_방사선
WHERE   방사선치료총선량 IS NOT NULL
AND ASCII(방사선치료1회선량) <> 0
AND (방사선치료1회선량*방사선치료총횟수) != 방사선치료총선량
```

(예시 43) 대중교통이용정보 : 파생되어 관리되는 항목에 대해 원천 항목의 값과 파생 된 항목의 값이 정확하게 관리되고 있는지 진단규칙 도출 예시

1. 진단대상 분석

1.1 파생되어 관리되는 항목에 대해 원천 항목의 값과 파생 된 항목의 값이 정확하게 관리되고 있는지 진단

- 00시대~23시대 승객수의 합계는 총승객수와 같다.
- 총승객수=(승객수00시+승객수01시+승객수02시+승객수03시+승객수04시+승객수05시+승객수06시+승객수07시+승객수08시+승객수09시+승객수10시+승객수11시+승객수12시+승객수13시+승객수14시+승객수15시+승객수16시+승객수17시+승객수18시+승객수19시+승객수20시+승객수21시+승객수22시+승객수23시)

테이블명	컬럼명	값
대중교통 이용정보	승객수	수

(조회 SQL - PostgreSQL 기준)

```
SELECT  (승객_수_00시+승객_수_01시+승객_수_02시+승객_수_03시+승객_수_04시+승객_수_05시+승객_수_06시+승객_수_07시+승객_수_08시+승객_수_09시+승객_수_10시+승객_수_11시+승객_수_12시+승객_수_13시+승객_수_14시+승객_수_15시+승객_수_16시+승객_수_17시+승객_수_18시+승객_수_19시+승객_수_20시+승객_수_21시+승객_수_22시+승객_수_23시) AS CALC_SUM, T1.총_승객_수 AS ERR_DATA
FROM    대중교통이용정보
```

WHERE (승객_수_00시+승객_수_01시+승객_수_02시+승객_수_03시+승객_수_04시+승객_수_05시+승객_수_06시+승객_수_07시+승객_수_08시+승객_수_09시+승객_수_10시+승객_수_11시+승객_수_12시+승객_수_13시+승객_수_14시+승객_수_15시+승객_수_16시+승객_수_17시+승객_수_18시+승객_수_19시+승객_수_20시+승객_수_21시+승객_수_22시+승객_수_23시) < > 총_승객_수

1 데이터 품질관리 단계별 산출물 서식

- 서식-1 품질관리 계획서
- 서식-2 품질진단 대상목록(데이터셋-테이블)
- 서식-3 품질진단 대상목록(항목-컬럼)
- 서식-4 품질진단 수행계획서
- 서식-5 프로파일링 결과내역
- 서식-6 품질진단 규칙 정의 내역
- 서식-7 품질진단 결과내역
- 서식-8 품질진단 결과서
- 서식-9 오류원인 분석 내역
- 서식-10 품질진단 결과 분석서
- 서식-11 품질개선 계획서
- 서식-12 품질개선 결과내역
- 서식-13 품질개선 결과서
- 서식-14 품질 통제 계획서

2 데이터 품질관리 기준 산출물 서식

- 서식-21 데이터 분류체계 기준 정의서
- 서식-22 핵심 품질 항목(CTQ) 정의서
- 서식-23 업무규칙 정의서
- 서식-24 품질 이슈 목록

3 데이터 연계키 서식

- 서식-25 연계표준인덱스 신청서

4 연계표준사전 서식

- 서식-26 연계표준용어 신청서(신규등록)
- 서식-27 연계표준용어 신청서(변경)
- 서식-28 연계표준용어 신청서(폐기 및 신규등록)
- 서식-29 연계표준단어 신청서(신규등록)
- 서식-30 연계표준단어 신청서(변경)
- 서식-31 연계표준단어 신청서(폐기 및 신규등록)
- 서식-32 연계표준도메인 신청서(신규등록)
- 서식-33 연계표준도메인 신청서(변경)
- 서식-34 연계표준도메인 신청서(폐기 및 신규등록)

[품질관리 단계별 산출물 목록]

DMAIC 5단계를 확장(Planning 단계 추가)한 데이터 품질관리 방법론의 16개 세부 절차별로 작성해야 하는 14종의 품질진단 산출물 목록은 다음과 같다.

품질진단 단계		품질진단 세부단계	산출물
Planning	품질관리 계획	D0. 품질관리 계획	1. 품질관리 계획서
Define	품질진단 대상 정의	D1. 데이터셋(테이블) 단위 선정	2. 품질진단 대상목록 (데이터셋-테이블)
		D2. 항목(컬럼) 단위 선정	3. 품질진단 대상목록 (항목-컬럼)
Measure	품질진단 수행	M1. 품질진단 수행 계획	4. 품질진단 수행계획서
		M2. 품질진단 규칙 정의	5. 프로파일링 결과내역 6. 품질진단 규칙 정의내역
		M3. 품질진단 실시	7. 품질진단 결과내역
		M4. 품질진단 결과	8. 품질진단 결과서
Analyze	진단결과 분석	A1. 오류원인 분석	9. 오류원인 분석내역
		A2. 개선기회 도출	10. 품질진단 결과 분석서
Improve	개선 수행	I1. 품질개선과제정의	11. 품질개선 계획서
		I2. 품질개선계획수립	
		I3. 품질개선환경준비	
		I4. 품질개선수행	12. 품질개선 결과내역 13. 품질개선 결과서
Control	품질 통제	C1. 개선결과 평가	14. 품질통제 계획서
		C2. 품질목표 관리	
		C3. 품질통제 실시	

01

데이터 품질진단 산출물 서식



서식-1 품질관리 계획서

품질관리 계획서

2021.09

기관명

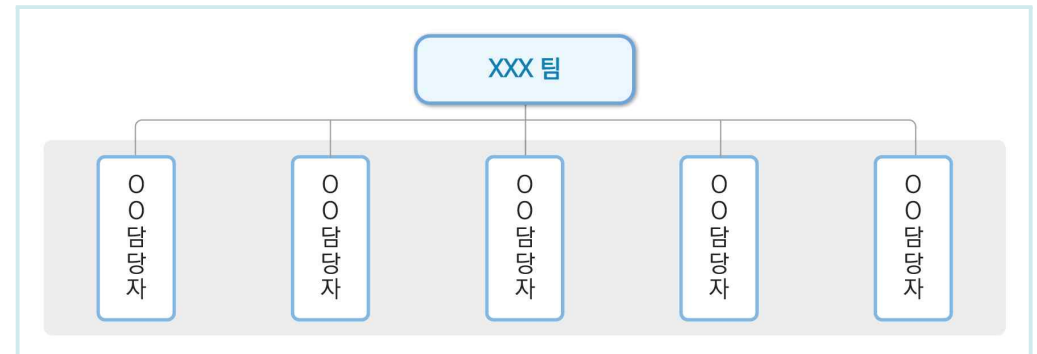
1 품질진단 및 개선 목적

작성방법 품질진단 및 개선 계획에 대한 목적을 기술한다.

예) 품질진단 정기 1차 점검에 따른 품질진단 수행 및 오류 데이터 개선을 수행하고자 한다.

2 품질관리 조직

작성방법 품질관리 수행 조직에 대한 조직도/역할 내역을 기술한다.



수행주체	역할
품질관리 담당자	OOO 품질관리 계획수립, 추진, 기획·조정
품질진단 담당자	OOO 관련 시스템 운영 등 OOO 품질진단 업무
품질개선 담당자	OOO 품질개선 업무
...	...

3 품질진단 대상

수행주체	DBMS	DATA SIZE (GB)	전체 테이블수	진단대상영역 여부
수집/저장	HBASE	100	250	Y
가공/융합	ORACLE	50	100	Y
분석/활용	ORACLE	20	60	N
...

4 품질진단 일정

수행 TASK	수행 일정	수행 내용
1. 품질진단대상선정	2021/09/13 ~ 2021/09/17	품질진단 대상 테이블 및 컬럼 선정
2. 품질진단환경구성	2021/09/27 ~ 2021/10/01	1.진단 데이터 적재 점검 2.진단DB, DQ TOOL 권한 구성
3. 품질진단규칙정의	2021/10/04 ~ 2021/10/08	* 지표 별 품질진단 규칙 정의
4. 품질진단수행	2021/10/11 ~ 2021/10/15	* 지표 별 품질진단 수행
5. 품질진단 결과	2021/10/18 ~ 2021/10/22	* 품질진단 결과 보고서 작성
6. 품질진단 결과 분석	2021/10/25 ~ 2021/10/29	* 품질진단 결과 원인분석 및 개선기회 도출

5 품질개선 계획

작성방법 품질개선 범위 및 일정을 기술한다.

▶ 품질개선 범위

예) 품질진단 정기 1차 품질진단 수행결과에 따라 최우선과제에 대한 개선을 수행하며, 단기 및 중장기 과제는 “품질개선 계획서”에 별도 일정을 수립 할 예정이다.

▶ 품질개선 일정

수행TASK	수행일정	수행내용
1. 품질개선 과제 정의	2021/10/29~2021/10/29	품질개선 과제에 대한 우선순위 및 수행 시기 결정
2. 품질개선환경구성	2021/11/01~2021/11/01	1. 품질개선 수행 프로그램 구성 2. 협조체계 구성(원천데이터 재구성)
3. 품질개선 수행(우선)	2021/11/02~2021/11/05	* 개선 과제 별 품질개선 수행
4. 품질개선 결과서 작성	2021/11/08~2021/11/08	* 품질개선 결과 보고서 작성
5. 품질개선 계획 (단기 및 중장기 과제)	2022/03~	* 단기 및 중장기 과제 개선 수행

서식-2 품질진단 대상목록(데이터셋-테이블)

▶ 품질진단 대상목록(데이터셋-테이블)

분류		테이블			
저장영역	DBMS	주제영역	테이블명	테이블 한글명	유형분류명
수집	ORACLE	고객	T_DD_ACDT_RT_LT	일일속박대여내역	거래
...

테이블	테이블 단위 선정		업무담당자	비고
전수/샘플	진단대상선정여부	제외사유		
전수	Y		홍길동	
...

▶ 주요 항목 설명

항목명	설명
저장영역	• 데이터 저장영역 명 (예 - 수집, 가공/분석, 활용)
DBMS	• DBMS(Oracle, MySql, DB2 등), NoSQL(MongGo DB, Neo4j, Redis) 등
주제영역	• 테이블의 주제영역
유형분류명	• 데이터 유형 분류체계에 따른 데이터셋(테이블)의 유형분류명 (예 - 메타, 기준, 거래, 로그, 분석, GIS, 문서, 미디어, 센서, 활용)
전수/샘플	• 데이터셋(테이블)에 대한 표본 추출 기준 참조(품질진단 및 개선 가이드) (예 - 전수, 샘플)
진단대상선정여부	• 품질진단 대상 여부 (예- Y, N)
제외사유	• 품질진단 대상 제외 사유 (예 - 그래프, 문서구조(JSON,XML), GIS데이터 등)

(# 유형분류명 항목의 상세 설명은 “[서식-21] 데이터 분류체계 기준 정의서” 참고)

서식-3 품질진단 대상목록(항목-컬럼)

▶ 품질진단 대상목록(항목-컬럼)

분류		테이블				
저장영역	DBMS	주제영역	테이블명	테이블 한글명	유형분류명	
수집	ORACLE	고객	T_DD_ACDT_RT_LT	일일숙박대여내역	거래	
...		

컬럼							
컬럼명	컬럼 한글명	컬럼타입	컬럼길이	도메인명	PK여부	FK여부	NOT NULL여부
D_RT_CNT	일일대여진수	NUMBER	5				
...

컬럼 단위 선정		비고
CTQ여부	진단대상선정여부	
Y	Y	
...

▶ 주요 항목 설명

항목명	설명
저장영역	<ul style="list-style-type: none"> 데이터 저장영역 명 (예 - 수집 , 가공/분석, 활용)
DBMS	<ul style="list-style-type: none"> DBMS(Oracle, MySql, DB2 등), NoSQL(MongGo DB, Neo4j, Redis) 등
주제영역	<ul style="list-style-type: none"> 테이블의 주제영역
유형분류명	<ul style="list-style-type: none"> 데이터 유형 분류체계에 따른 데이터셋(테이블)의 유형분류명 (예 - 메타, 기준, 거래, 로그, 분석, GIS, 문서, 미디어, 센서, 활동)
CTQ여부	<ul style="list-style-type: none"> CTQ 항목 여부 (예 - Y, N)
진단대상선정여부	<ul style="list-style-type: none"> 품질진단 대상 여부 (예 - Y, N)

(# 유형분류명 항목의 상세 설명은 “[서식-21] 데이터 분류체계 기준 정의서” 참고)

서식-4 품질진단 수행계획서

품질진단 수행계획서

2021.09

기관명

1 품질진단 목적

작성방법 품질진단을 수행하는 목적에 대하여 기술한다.

예) 품질진단 정책에 따른 월 정기 1차 데이터 품질진단을 수행한다.

2 품질관리 조직

작성방법 품질진단 대상 현황에 대해 기술한다.

▶ 품질진단 대상 현황

저장영역	DBMS	주제영역	전체		진단 대상		제외 대상	
			테이블수	컬럼수	테이블수	컬럼수	테이블수	컬럼수
수집	ORACLE	고객	115	800	55	480	60	320
수집	HBASE	고객	50	600	30	350	20	250
가공/분석	HBASE	고객	100	400	60	150	40	250
활용	MYSQL	고객	90	300	60	190	30	110
...

(# 상세 내역: [서식-2] 품질진단대상목록(데이터셋-테이블), [서식-3] 품질진단 대상목록(항목-컬럼))

3 품질진단 환경구성

작성방법 품질진단 환경 구성에 대해 기술한다.

예) 저장영역 별, 진단의 환경(원천, 복제), 품질진단도구/SQL 사용 등 설명을 기술한다.

저장영역	DBMS	진단 대상 데이터 정보	품질진단방식	환경구성 필요내역
수집/저장	HBASE	복제 데이터 진단 (DQ전용저장소-ORACLE)	SQL 작성, DQ TOOL	1. 진단DB 권한 구성 2. DQ TOOL 설치 및 환경구성
가공/융합	HBASE	복제 데이터 진단 (DQ전용저장소-ORACLE)	DQ TOOL	1. 진단DB 권한 구성 2. DQ TOOL 설치 및 환경구성
분석/활용	MYSQL	운영 데이터 진단	SQL 작성	진단DB 권한 구성

4 품질진단 일정

작성방법 품질진단 환경 구성에 대해 기술한다.

수행TASK	수행일정	수행내용
1. 품질진단환경구성	2021/09/27 ~ 2021/1-01	1. 진단 데이터 적재 점검 2. 진단DB, DQ TOOL 권한 구성
2. 품질진단규칙정의	2021/11/04 ~ 2021/11/08	* 지표 별 품질진단 규칙 정의
3. 품질진단수행	2021/10/11 ~ 2021/10/15	* 지표 별 품질진단 수행
4. 품질진단결과확인	2021/10/18 ~ 2021/10/22	* 품질진단 결과 보고서 작성

서식-5 프로파일링 결과내역

▶ 프로파일링 결과내역

분류				테이블			
저장영역	DBMS	주제영역	유형분류명	테이블명		테이블 한글명	
수집	ORACLE	숙박	거래	T_DD_ACDT_RT_LT		일일숙박대여내역	
...	
컬럼							
컬럼명	컬럼 한글명	컬럼타입	컬럼길이	도메인명	PK여부	FK여부	NOT NULL 여부
D_RT_CNT	일일대여건수	NUMBER	5				
프로파일링 결과							비고
패턴 값	건수	MIN 값	MAX 값	총건수			
999	300	0	550	10,000,000,000			
...			

▶ 주요 항목 설명

항목명	설명
저장영역	• 데이터 저장영역 명 (예 - 수집, 가공/분석, 활용)
DBMS	• DBMS(Oracle, MySql, DB2 등), NoSQL(MongGo DB, Neo4j, Redis) 등
주제영역	• 테이블의 주제영역
유형분류명	• 데이터 유형 분류체계에 따른 데이터셋(테이블)의 유형분류명 (예 - 메타, 기준, 거래, 로그, 분석, GIS, 문서, 멀티미디어, 센서, 활동)
(프로파일링 결과)	• 패턴값, 패턴별 건수, MIN, MAX, 총건수 등 프로파일링 의도에 맞게 항목 변경 가능
패턴 값	• 패턴 분석이 의미 있는 항목에 대해 DISTINCT 등을 수행한 결과의 값 종류 (예 - 여부 항목인 경우, Y, N, 0, 1)

(# 유형분류명 항목의 상세 설명은 “[서식-21] 데이터 분류체계 기준 정의서” 참고)

서식-6 품질진단 규칙 정의 내역

▶ 품질진단 규칙 정의 내역

분류				테이블	
저장영역	DBMS	주제영역	유형분류명	테이블명	테이블 한글명
수집	ORACLE	숙박	거래	T_DD_ACDT_RT_LT	일일숙박대여내역
...

분류		테이블				
저장영역	DBMS	주제영역	테이블명	테이블 한글명	전수/샘플	유형분류명
활용	ORACLE	숙박	T_DD_ACDT_RT_LT	일일숙박대여내역	전수	거래

컬럼							
컬럼명	컬럼 한글명	컬럼타입	컬럼길이	도메인명	PK 여부	FK 여부	NOT NULL여부
D_RT_CNT	일일대여건수	NUMBER	5				

품질진단 규칙 정의						
CTQ여부	진단대상선정여부	품질지표(DQI)	품질상세지표	진단규칙내용	진단기법	업무규칙ID
Y	Y	정확성	논리관계 정확성		업무규칙	BR0001

품질진단 규칙 정의 (업무규칙인 경우 생략)				비고
총건수		오류건수	오류내역	
SELECT COUNT(*) ...		SELECT COUNT(*)	SELECT 숙박대여일자 ...	

▶ 주요 항목 설명

항목명	설명
저장영역	<ul style="list-style-type: none"> 데이터 저장영역 명 (예 - 수집, 가공/분석, 활용)
DBMS	<ul style="list-style-type: none"> DBMS(Oracle, MySql, DB2 등), NoSQL(MongGo DB, Neo4j, Redis) 등
주제영역	<ul style="list-style-type: none"> 테이블의 주제영역
유형분류명	<ul style="list-style-type: none"> 데이터 유형 분류체계에 따른 데이터셋(테이블)의 유형분류명 (예 - 메타, 기준, 거래, 로그, 분석, GIS, 문서, 미디어, 센서, 활동)
CTQ여부	<ul style="list-style-type: none"> CTQ 항목 여부 (예 - Y, N)
품질지표(DQI)	<ul style="list-style-type: none"> 데이터 품질관리 지표(유효성, 정확성, 완전성, 일관성 등)
품질상세지표	<ul style="list-style-type: none"> 데이터 품질관리 상세지표(참조무결성, 중복일관성, 범위유효성, 논리관계정확성, 단독완전성 등)
진단기법	<ul style="list-style-type: none"> 품질진단의 기법 종류 (예 - 메타정보, 프로파일링, 업무규칙)
진단규칙내용	<ul style="list-style-type: none"> 품질진단을 수행할 내용을 기반으로 진단스크립트(sql)를 작성
업무규칙ID	<ul style="list-style-type: none"> 업무규칙 생성시 업무규칙정의서에 정의된 ID명
총건수	<ul style="list-style-type: none"> 총건수를 구할 수 있는 스크립트(sql문)
오류건수	<ul style="list-style-type: none"> 오류건수를 구할 수 있는 스크립트(sql문)
오류내역	<ul style="list-style-type: none"> 품질진단규칙 내용에 해당하는 오류내역을 구하는 스크립트(sql문)

(# 유형분류명 항목의 상세 설명은 “[서식-21] 데이터 분류체계 기준 정의서” 참고)

서식-7 품질진단 결과내역

▶ 품질진단 결과내역

분류		테이블				
저장영역	DBMS	주제영역	테이블명	테이블 한글명	전수/샘플	유형분류명
분석	ORACLE	숙박	T_DD_ACDT_RT_LT	일일숙박대여내역	전수	거래
컬럼						
컬럼명	컬럼 한글명	컬럼타입	컬럼길이	도메인명	PK 여부	FK 여부
D_RT_CNT	일일대여건수	NUMBER	5			
품질진단 규칙 정의						
CTQ여부	진단대상선택여부	품질지표(DQI)	품질상세지표	진단규칙내용	진단기법	업무규칙ID
Y	Y	정확성	논리관계 정확성		업무규칙	BR0001
품질진단 규칙 정의 (업무규칙인 경우 생략)				품질진단 결과		
총건수	오류건수	오류내역	전체건수	오류건수	실질 오류율(%)	
SELECT COUNT(*) ...	SELECT COUNT(*)	SELECT 숙박대여일자 ...	10,000,000	98	0.001	

▶ 주요 항목 설명

항목명	설명
저장영역	• 데이터 저장영역 명 (예 - 수집, 가공/분석, 활용)
DBMS	• DBMS(Oracle, MySql, DB2 등), NoSQL(MongGo DB, Neo4j, Redis) 등
주제영역	• 테이블의 주제영역
유형분류명	• 데이터 유형 분류체계에 따른 데이터셋(테이블)의 유형분류명 (예 - 메타, 기준, 거래, 로그, 분석, GIS, 문서, 미디어, 센서, 활동)
CTQ여부	• CTQ 항목 여부 (예 - Y, N)
품질지표(DQI)	• 데이터 품질관리 지표(유효성, 정확성, 완전성, 일관성 등)
품질상세지표	• 데이터 품질관리 상세지표(참조무결성, 중복일관성, 범위유효성, 논리관계정확성, 단독완전성 등)
진단규칙내용	• 품질진단을 수행 할 내용을 기반으로 진단스크립트(sql)를 작성
진단기법	• 품질진단의 기법 종류 (예 - 메타정보, 프로파일링, 업무규칙)
업무규칙ID	• 업무규칙 생성시 업무규칙정의서에 정의 된 ID명

(# 유형분류명 항목의 상세 설명은 “[서식-21] 데이터 분류체계 기준 정의서” 참고)

서식-8 품질진단 결과서

품질진단 결과서

2021.11

기관명

1 품질진단 수행 현황

작성방법 품질진단의 계획대비 수행 현황을 기술한다.

저장영역	DBMS	주제 영역	전체		진단 대상(계획)		진단 대상(수행)	
			테이블수	컬럼수	테이블수	컬럼수	테이블수	컬럼수
수집	ORACLE	고객	115	800	55	480	54	470
수집	HBASE	고객	50	600	30	350	30	350
가공/분석	HBASE	고객	100	400	60	150	60	150
활용	MYSQL	고객	90	300	60	190	60	190
...

(# 상세 내역: [서식-2]품질진단대상목록(데이터셋-테이블), [서식-3] 품질진단 대상목록(항목-컬럼))

2 품질진단 수행 결과내역

작성방법 품질진단결과내역을 요약 기술한다.

▶ 품질진단 오류 종합현황

○ 정량진단(값) 오류율

데이터영역	완전성	유효성	일관성	정확성
수집	1.4%	1.5%	0.8%	0.3%
가공/분석	0.9%	0.3%	0.4%	0.2%
활용	0.1%	0.09%	0.001%	0.02%

○ 정량진단(구조표준) 오류율

데이터영역	완전성(구조)	일관성(표준)
수집	0.3%	0.1%
가공/분석
활용

▶ 품질진단 오류 상세현황

○ 정량진단(값)

품질지표	품질상세지표	진단 내용	테이블수	컬럼수	전체건수	오류 건수	실질 오류율 (%)	비고
유효성	여부	여부유효값진단	54	470	1000000000	350	0.000035	
	범위	금액유효범위진단	30	350	1200000000	1000	0.0001	
	날짜	...	60	150	2000000000	200	0.00001	
	코드	...	60	190	50000000	50	0.0001	
정확성	파생항목 정확성	
...
총계								

○ 정량진단(구조, 표준)

품질지표	품질상세지표	진단내용	테이블수	컬럼수	오류건수	실질 오류율(%)	비고
완전성	구조 완전성	테이블중복 진단	100	-	2	2	
		테이블 현행화 진단	100	-	1	1	
		컬럼 현행화진단	-	1000	4	0.4	
		식별자 누락 진단	-	1000	3	0.3	
일관성	표준 일관성	표준 도메인 진단	-	1000	10	1	
		엔터티명명규칙 진단	100	-	2	2	
		테이블 명명규칙 진단	100	-	2	2	
총계							

(# 상세 내역: [서식-7] 품질진단 결과내역)

3 진단결과 총평

작성방법 품질진단 이후 진행되어야 할 사항에 대한 간단한 요약 및 일정을 기술한다.

예) 품질진단 결과의 전반적인 오류율을 비교하였을 때 유효성 지표에 대한 오류율이 높아 향후 유효성 지표 관련 품질진단 대상의 지속적인 관리가 필요하다.

예) 품질진단 계획대비 미수행 1개 테이블 10개 컬럼에 대한 진단실시는 2차 품질진단에 적용하여 수행 예정이다.

서식-9 오류원인 분석 내역

▶ 오류원인 분석 내역

분류		테이블				
저장영역	DBMS	주제영역	테이블명	테이블 한글명	전수/샘플	유형분류명
수집	ORACLE	숙박	T_DD_ACDT_RT_LT	일일숙박대여내역	전수	거래

컬럼							
컬럼명	컬럼 한글명	컬럼타입	컬럼길이	도메인명	PK여부	FK여부	NOT NULL
D_RT_CNT	일일대여건수	NUMBER	5				

품질진단 규칙 정의					
CTQ여부	품질지표(DQI)	품질상세지표	진단규칙내용	진단 기법	업무규칙ID
Y	정확성	논리관계 정확성		업무규칙	BR0001

품질진단 규칙 정의(업무규칙인 경우 SQL 생략)			품질진단 결과		
총건수	오류건수	오류내역	전체건수	오류건수	실질 오류율(%)
SELECT COUNT(*)	SELECT COUNT(*)	SELECT 숙박대여일자	10,000,000	98	0.001

품질진단결과 분석(오류원인 분석)					
오류발생원인	값		구조	표준	관리체계
원천 데이터 오류	기 적재 데이터는 센터와 협의후 직접수정				

품질진단결과 분석		
관리체계	응용프로그램	개선 시기 (우선순위)
	일일숙박대여건수 수신 프로그램 체크로직 추가	

▶ 주요 항목 설명

항목명	설명
CTQ여부	• CTQ 항목 여부 (예 - Y, N)
품질지표(DQI)	• 데이터 품질관리 지표(유효성, 정확성, 완전성, 일관성 등)
품질상세지표	• 데이터 품질관리 상세지표(참조무결성, 중복일관성, 범위유효성, 논리관계정확성, 단독완전성 등)

항목명	설명
진단규칙내용	• 질진단을 수행 할 내용을 기반으로 진단스크립트(sql)를 작성
진단기법	• 품질진단의 기법 종류 (예 - 메타정보, 프로파일링, 업무규칙)
업무규칙ID	• 업무규칙 생성시 업무규칙정의서에 정의 된 ID명
오류발생원인	• 진단분석결과 오류가 발생한 원인(예 - 응용프로그램 검증 부재)
값	• 개선해야 할 내용이 값이면 개선 내용 기재
구조	• 개선해야 할 부분이 구조이면 개선 내용 기재
표준	• 개선해야 할 부분이 표준이면 개선 내용 기재
관리체계	• 개선해야 할 부분이 관리체계이면 개선 내용 기재
응용프로그램	• 개선해야 할 부분이 응용프로그램이면 개선 내용 기재
개선 시기 (우선순위)	• 개선해야 할 내용에 대해 업무중요도, 시스템영향도 등을 파악하여 개선시기 부여 (예 - 최우선 , 단기, 중장기)

(# 유형분류명 항목의 상세 설명은 "[서식-21] 데이터 분류체계 기준 정의서" 참고)

서식-10 품질진단 결과 분석서

품질진단 결과 분석서

2021.11

기관명

1 품질진단 결과내역

작성방법 품질진단결과내역을 기술한다.

▶ 지표 별 오류 현황

○ 정량진단(값)

품질지표	품질상세지표	진단 내용	테이블수	컬럼수	전체건수	오류건수	실질오류율(%)	비고
유효성	여부	여부유효값진단	54	470	1000000000	350	0.000035	
	범위	금액유효범위진단	30	350	1200000000	1000	0.0001	
	날짜	...	60	150	2000000000	200	0.00001	
	코드	...	60	190	500000000	50	0.0001	
정확성	파생항목 정확성	
...
총계								

○ 정량진단(구조, 표준)

품질지표	품질상세지표	진단내용	테이블수	컬럼수	오류건수	실질오류율(%)	비고
완전성	구조완전성	테이블중복 진단	100	-	2	2	
		테이블 현행화 진단	100	-	1	1	
		컬럼 현행화진단	-	1000	4	0.4	
		식별자 누락 진단	-	1000	3	0.3	
일관성	표준일관성	표준 도메인 진단	-	1000	10	1	
		엔터티명명규칙 진단	100	-	2	2	
		테이블 명명규칙 진단	100	-	2	2	
총계							

(# 상세 내역: [서식-7]품질진단 결과내역)

2 오류 원인 현황

작성방법 품질진단결과 분석내역의 오류발생유형을 요약 기술한다.

▶ 지표 별 발생 유형 현황

저장영역	DBMS	오류발생유형 수				
		값 오류	구조 오류	표준 오류	관리체계 오류	응용프로그램 오류
수집						
가공/분석						
활용						

(# 상세 내역: [서식-9]오류원인 분석 내역)

3 개선 기회 도출

작성방법 오류발생원인에 따른 개선 기회에 대해 기술한다.

오류발생유형	오류발생원인	개선 기회
값	원천데이터 오류	원천데이터 개선
구조	<ul style="list-style-type: none"> 논리/물리 모델의 현행화 미흡 데이터 모델관리 가이드 및 지침서 부재 	<ul style="list-style-type: none"> 논리/물리 모델 현행화 수행 데이터 모델관리 가이드 및 지침 마련
표준	<ul style="list-style-type: none"> 표준(용어, 도메인) 지침의 변경관리 미흡 	<ul style="list-style-type: none"> 표준(용어, 도메인) 정비 수행
관리체계	<ul style="list-style-type: none"> 데이터 품질관리 지침 가이드 부재 	<ul style="list-style-type: none"> 데이터 품질관리 지침 마련
응용프로그램	<ul style="list-style-type: none"> 연계/수집/생성 값의 형식 검증기능의 미흡/부재 	<ul style="list-style-type: none"> 형식 검증기능 보완 및 추가

(# 상세 내역: [서식-9] 오류원인 분석 내역)

서식-11 품질개선 계획서

품질개선 계획서

2021.11

기관명

1 품질개선 목적

작성방법 품질개선 수행의 목적을 기술한다.

예) 시스템 품질진단 정기 1차 점검에 따른 오류에 대한 품질개선을 실시하고자 한다.

2 품질개선 대상

작성방법 품질오류 개선 대상에 대해 기술한다.

▶ 저장영역별 오류 발생 유형 현황

저장영역	DBMS	오류발생유형 수				
		값 오류	구조 오류	표준 오류	관리체계 오류	응용프로그램 오류
수집	HBASE					
가공/분석	ORACLE					
활용	ORACLE					

(# 상세 내역: [서식-7] 품질진단결과 분석결과 내역, [서식-10] 품질진단 결과 분석서)

▶ 저장영역별 오류 발생 현황

작성방법 오류발생유형 별 오류가 발생된 테이블 컬럼의 개수를 기술한다.

저장영역	DBMS	오류 발생 테이블 수	오류 발생 컬럼 수
수집	HBASE	10	18
가공/분석	ORACLE	5	9
활용	ORACLE	3	10

3 개선과제 정의

작성방법 개선 우선순위 및 개선과제를 정의하여 기술한다.

▶ 개선과제별 우선순위 선정

우선순위	개선 수행방식	개선과제
최우선 개선과제 (품질관리계획서의 일정)	유지보수 중 수행	기본적 오류 데이터 개선
	유지보수 중 수행	정제된 데이터 일괄 재수신
	유지보수 중 수행	형식 검증 응용프로그램 개선
단기 개선과제 (반기 단위)	유지보수 중 수행	데이터 모델 현행화
	유지보수 중 수행	표준(용어, 도메인) 현행화 수행
중장기 개선과제 (년 단위)	품질개선 사업 발주	데이터 품질/모델 관리 지침 마련

▶ 저장영역별 우선순위 도출

작성방법 저장영역, DBMS 별 오류발생유형에 대한 개선 과제 별 우선순위를 기술한다.

예) 우선순위 유형 : ① 최우선 과제, ② 단기 과제, ③ 중장기 과제

저장영역	DBMS	오류발생유형				
		값	구조	표준	관리체계	응용프로그램
수집	HBASE	① 최우선 과제	① 최우선 과제	① 최우선 과제	② 단기 과제	① 최우선 과제
가공/분석	ORACLE	① 최우선 과제	② 단기 과제	② 단기 과제	② 단기 과제	② 단기 과제
활용	ORACLE	① 최우선 과제	③ 중장기 과제	③ 중장기 과제	③ 중장기 과제	③ 중장기 과제

4 품질개선 수행 일정

작성방법 품질개선 수행일정을 기술한다.

▶ 최우선 과제

수행TASK	수행일정	수행내용
1. 품질개선환경구성	2021/11/01~2021/11/01	1. 품질개선 수행 시스템 (프로그램, 데이터) 구성 2. 관련 이해관계자와의 협조체계 구성
2. 품질개선 수행 (최우선)	2021/11/02~2021/11/05	• 개선과제별 품질개선 수행
3. 품질개선 결과서 작성	2021/11/08~2021/11/08	• 품질개선 결과 보고서 작성

▶ 단기 및 중장기 과제

우선순위	과제명	수행일정	수행방법	수행예산
② 단기 과제	데이터 모델 현행화	2021/12/27~2021/12/31	운영인력 자체수행	해당없음
② 단기 과제	표준(용어, 도메인) 현행화 수행	2022/01/03~2022/01/07	별도 사업 추진	xxx만원
③ 중장기 과제	데이터 품질/모델 관리 지침 마련

서식-12 품질개선 결과내역

▶ 품질개선 결과내역

분류		테이블				
저장영역	DBMS	주제영역	테이블명	테이블 한글명	전수/샘플	유형분류명
수집	ORACLE	숙박	T_DD_ACDT_RT_LT	일일숙박대여내역	전수	거래

컬럼							
컬럼명	컬럼 한글명	컬럼타입	컬럼길이	도메인명	PK여부	FK여부	NOT NULL 여부
D_RT_CNT	일일대여건수	NUMBER	5				

품질진단 규칙 정의						
CTQ여부	진단대상선정여부	품질지표 (DQI)	품질상세지표	진단내용	진단기법	업무규칙ID
Y		정확성	논리관계 정확성		업무규칙	BR0001

품질진단 규칙 정의			품질진단 결과		
총건수	오류건수	오류내역	전체건수	오류건수	실질 오류율(%)
SELECT COUNT(*)	SELECT COUNT(*)	SELECT 숙박대여일자	10,000,000	98	0.001

품질진단결과 분석			
오류발생원인	값	구조	표준
원천 데이터 오류	기 적재 데이터는 센터와 협의후 직접수정		

품질진단결과 분석		
관리체계	응용프로그램	개선 시기 (우선순위)
	일일숙박대여건수 수신 프로그램 체크로직 추가	

개선 수행							
값	구조	표준	관리 체계	응용프로그램	개선상태	최종개선일자	업무담당자
N				Y	개선중	2021-11-05	홍길동

▶ 주요 항목 설명

항목명	설명
CTQ여부	• CTQ 항목 여부 (예 - Y, N)
품질지표(DQI)	• 데이터 품질관리 지표(유효성, 정확성, 완전성, 일관성 등)
품질상세지표	• 데이터 품질관리 상세지표(참조무결성, 중복일관성, 범위유효성, 논리관계정확성, 단독완전성 등)
진단규칙내용	• 진단단을 수행 할 내용을 기반으로 진단스크립트(sql)를 작성
진단기법	• 품질진단의 기법 종류 (예 - 메타정보, 프로파일링, 업무규칙)
오류발생원인	• 진단분석결과 오류가 발생한 원인(예 - 응용프로그램 검증 부재)
품질진단결과분석 (값/구조/표준/관리체계/응용)	• 진단분석결과 개선내용을 해당 영역에 기재 (여러 영역에 개선이 필요하면 복수 기재)
개선시기(우선순위)	• 개선해야 할 내용에 대해 업무중요도, 시스템영향도 등을 파악하여 개선시기 부여 (예 - 최우선, 단기, 중장기)
개선수행 (값/구조/표준/관리체계/응용)	• 개선을 수행한 대상 영역 항목에 체크
개선상태	• 개선이 수행되는 과정(예 - 미개선, 개선중, 개선완료)
최종개선일자	• 최종개선 일자

(# 유형분류명 항목의 상세 설명은 “[서식-21] 데이터 분류체계 기준 정의서” 참고)

서식-13 품질개선 결과서

품질개선 결과서

2021.11

기관명

1 품질 개선 대상

작성방법 품질개선의 계획대비 수행 현황을 기술한다.

2 우선 개선과제 정의

작성방법 품질개선의 우선 개선과제에 대해 기술한다.

우선순위	개선 수행방식	개선과제
최우선 개선 과제 (품질관리계획서의 일정)	유지보수 중 수행	기본적 오류 데이터 개선
	유지보수 중 수행	정제 된 데이터 일괄 재수신
	유지보수 중 수행	형식 검증 응용프로그램 개선

3 품질개선 수행 결과

작성방법 품질개선 결과를 기술한다.

▶ 오류유형 및 지표 별 개선 결과

○ 영역별 오류유형 개선 수행 현황

저장영역	DBMS	오류유형 수									
		값 오류		구조 오류		표준 오류		관리체계 오류		응용프로그램 오류	
		개선 대상	개선 완료	개선 대상	개선 완료	개선 대상	개선 완료	개선 대상	개선 완료	개선 대상	개선 완료
수집	HBASE										
가공/분석	ORACLE										
활용	ORACLE										

(# 상세 내역: [서식-10] 품질진단 결과 분석서, [서식-12] 품질개선 결과내역)

○ 지표별 개선 전후 비교 현황

- 정량진단(값)

품질지표	품질상세지표	개선전			개선후	
		전체건수	오류건수	실질 오류율(%)	개선건수	실질 개선율 (%)
완전성	단독 완전성					
	조건 완전성					
	구조 완전성					
	공간객체 완전성					
유효성	범위 유효성					
	날짜 유효성					
	형식 유효성					
	코드 유효성					
	선후관계 유효성					
일관성	참조무결성					
	중복일관성					
	위상구조 일관성					
	개념적 일관성					
	연계일관성					
정확성	논리관계 정확성					
	파생항목 정확성					
	메타 정확성					
	그리드 정확성					
	위치 정확성					

- 정량진단(구조, 표준)

품질지표	품질상세지표	개선전			개선후	
		전체건수	오류건수	실질 오류율(%)	개선건수	실질 개선율 (%)
일관성	표준 일관성					
완전성	구조 완전성					

(# 상세 내역: [서식-7] 품질개선 결과내역)

서식-14 품질통제 계획서

품질통제 계획서

2021.09

기관명

1 품질개선 대상

작성방법 품질통제 수행의 목적을 기술한다.

예) 데이터품질 정기진단 결과에 따른 지표별 오류율 개선 결과와 주기적인(반기·분기별) 지표별 성숙도 측정을 통해 품질개선 효과를 평가한다. 그리고 데이터 품질을 유지 또는 향상시키기 위한 품질목표 수준을 설정하고 지속인 개선활동과 모니터링을 수행하기 위함이다.

2 품질수준 현황

작성방법 개선 전·후의 오류율을 비교하고 주기적인 성숙도 평가를 통해 현 수준의 데이터 품질수준을 평가한다.

▶ 지표별 오류율 개선 결과

평가지표		오류율(전체 건수 기준)	
품질지표	상세지표	개선 전	개선 후
일관성	식별자 정확성	1.5%	0%
	테이블간 일관성	2.0%	0%

유효성	범위 유효성	1.2%	0%
	날짜 유효성	2.4%	0%
	형식 유효성	1.3%	0%
정확성	코드 정확성	0.5%	0.2%
	표준 정확성	0.5%	0.1%
완전성	...	0.5%	0%

▶ 지표별 성숙도 평가

품질지표	성숙도 평가			
	19-01	19-02	19-03	19-04
일관성				
유효성				
정확성				
완전성				
...				

3 품질 수준 목표설정

작성방법 현재 시점의 품질수준을 고려하여 상세지표별 목표 오류율과 품질지표별 목표 성숙도를 결정함으로써 정량적인 데이터 품질수준 목표를 설정한다.

평가지표		품질목표	
품질지표	상세지표	오류율	성숙도
일관성	식별자 정확성		
	테이블간 일관성		
	...		
유효성	범위 유효성		
	날짜 유효성		
	형식 유효성		
정확성	코드 정확성		
	표준 정확성		
완전성	...		

4 품질통제 활동 계획

작성방법 품질 수준 목표 달성을 위한 품질 활동 및 모니터링 계획을 수립한다.

예) 지속적인 품질개선을 위한 품질교육 및 문화정착, 시스템 품질점검 활동, 데이터 품질 이슈 및 품질수준 모니터링 활동, 품질오류 신고 접수 및 처리 활동 등

02

데이터 관리 산출물 서식



서식-21 데이터 분류체계 기준 정의서

1 데이터 분류체계 기준 정의

품질진단 대상 선정을 위해서는 “데이터 유형” 및 “데이터 구조”의 분류체계가 정의되어 있어야 하며, 전체 테이블 별로 제시된 분류체계를 참고하여 “데이터 유형” 및 “데이터 구조” 관점으로 테이블을 분류한다.

2 데이터 분류체계 기준

▶ 데이터 유형

데이터 유형	데이터 유형 설명	데이터 유형(예시)
메타 데이터 (Meta data)	<ul style="list-style-type: none"> 데이터에 대한 표준화 정보 및 데이터 특성을 설명하는 요약 정보로 정보명, 주제영역, 위치, 중요도, 품질수준, 다른 데이터와 연관성 등을 관리하는 정보 	<ul style="list-style-type: none"> 표준사전, 데이터셋의 특성정보(정보명, 위치, 주제영역, 출처, 중요도, 품질수준, 연관성 등)
기준 데이터 (Master data)	<ul style="list-style-type: none"> 업무 프로세스의 중심이 되는 공통 개념의 데이터로 정보 값이 자주 변하지 않는 정보 및 참조정보(Reference data : 데이터의 값을 참조하는 코드성 정보)를 포함 	<ul style="list-style-type: none"> 제품명세, 고객정보, 사업자정보 등 지역코드, 성별코드, 학력코드 등
로그데이터 (Log data)	<ul style="list-style-type: none"> 시스템이 생성한 Log,웹상의 모든 흔적들의 Raw file 형식 데이터 	<ul style="list-style-type: none"> 로그기록, WebLog (Raw file) 등

데이터 유형	데이터 유형 설명	데이터 유형(예시)
거래 데이터 (Transaction data)	<ul style="list-style-type: none"> 기업이나 기관의 고유한 업무활동 및 서비스에 의하여 생성되는 정보 	<ul style="list-style-type: none"> 신용/직불카드 및 금융 거래 자료, 유통업체 자료
분석 데이터 (Analytics data)	<ul style="list-style-type: none"> 집계 데이터 및 분석을 통하여 결과로 생성된 정보 	<ul style="list-style-type: none"> 지역별/업종별 매출현황, 서울 인구이동 분포, 종로 상권분석 등
GIS 데이터 (GIS data)	<ul style="list-style-type: none"> GIS 시스템에 의하여 생성한 공간 정보로 벡터정보와 속성정보로 구성된 정보 	<ul style="list-style-type: none"> 행정구역도, 지하매설물도, 농경지지도, 산림도 등
문서 데이터 (Document data)	<ul style="list-style-type: none"> 문서작성기로 생성한 문서 정보(고유의 저장형식을 가짐) 	<ul style="list-style-type: none"> 일반문서, 논문, 보고서 등
멀티미디어 데이터 (Multimedia data)	<ul style="list-style-type: none"> 사진, 영상, 음성 등 멀티미디어 정보 	<ul style="list-style-type: none"> 사진, 위성사진, 항공사진, 영상, 음성 등
센서 데이터 (Sensor data)	<ul style="list-style-type: none"> 추적장치정보(Tracking device data)를 포함한 각종 센서를 통하여 생성되는 정보 	<ul style="list-style-type: none"> 위치, 기상, 수질, 대기, IoT 등
활동 데이터 (Online behavior data)	<ul style="list-style-type: none"> 의견정보(Opiniondata), 웹 검색정보 등을 포함한 온라인상에서 생성된 것으로 분석을 위하여 가공된 정보 	<ul style="list-style-type: none"> 인터넷 검색 및 페이지뷰 인덱싱 정보(WebLog) 등 이메일, 카카오톡, 트위터 등 공개 자료

▶ 데이터 구조

데이터 구조	데이터 구조 설명	데이터 구조(예시)
주제영역 (Subject Area)	<ul style="list-style-type: none"> 기업/기관이 사용하는 데이터의 최상위 집합 <ul style="list-style-type: none"> 하나의 주제영역으로 정의되는 데이터 간의 관계는 밀접하고, 다른 주제영역에 포함되는 데이터 간의 상호작용은 최소화할 수 있도록 정의 비즈니스 목표를 달성하기 위해 전사적으로 관리되어야 하는 데이터를 업무에 따른 분류가 아닌 데이터 자체 고유 속성을 기준으로 분류한 데이터의 그룹을 의미함 <ul style="list-style-type: none"> 주제영역은 데이터의 정체성, 유일성, 독립성을 기준으로 분류함 	<ul style="list-style-type: none"> 고객(기업고객, 개인고객, 회원고객...), 제품(완제품, 반제품, 가공품...), 회원(법인회원, 일반회원...), 주문(일반주문, 계획주문, 수시주문, 특별주문,...) 하나의 주제영역은 하위의 주제영역 또는 하나 이상의 엔티티로 구성됨

데이터 구조	데이터 구조 설명	데이터 구조(예시)
주제영역 (Subject Area)	<ul style="list-style-type: none"> 전사 데이터 주제영역은 전사 업무를 위한 전체 데이터 구성에 대한 청사진 <ul style="list-style-type: none"> 데이터 구성 및 통합에 대한 방향 제시 (선언적 성격) 효율적 데이터 관리를 위한 기준 제공 	<ul style="list-style-type: none">
엔티티 (Entity)	<ul style="list-style-type: none"> 데이터베이스 분야에서 개체 또는 엔티티(Entity)라고 하는 것은 데이터베이스에 표현하려고 하는 유형, 무형의 객체로서 서로 구별되는 것을 뜻함 이 개체는 현실 세계에 대해 사람이 생각하는 개념이나 정보의 단위로서 의미를 가지고 있음 하나의 개체는 하나 이상의 속성, 즉 애트리뷰트(attribute)로 구성되고 각 속성은 그 개체의 특성이나 상태를 기술해 줌 	<ul style="list-style-type: none"> 학생이라는 개체 또는 엔티티는 학번, 이름, 학과라는 3개의 속성들로 구성되어 있음. 이때 학번, 이름, 학과는 학생이라는 개체가 가지고 있는 특성, 즉 값을 나타내고 있는 것임 하나의 엔티티는 하나 이상의 속성을 가짐
테이블 (Table)	<ul style="list-style-type: none"> 분석 단계 논리 데이터 모델에서 사용된 엔티티는 설계 단계 물리 데이터 모델에서 테이블이 됨 	<ul style="list-style-type: none"> M E M B E R (회 원) , PRODUCT(제품)
테이블 스페이스 (Table Space)	<ul style="list-style-type: none"> 디스크를 포맷팅한 후 데이터베이스를 생성할 때, 업무적으로 연관된 테이블들끼리 클러스터링하여 동일한 데이터베이스에 저장해야 성능 및 유지관리 등의 측면에서 유리한데, 일반적으로 주제영역명이 테이블 스페이스명이 됨 	<ul style="list-style-type: none"> TS_MEM, TS_USR

▶ 테이블 분류

테이블	엔티티명	데이터 유형	데이터 구조
MEMBER	회원	기준 데이터	회원
MEMBER_DETAIL	회원상세	기준 데이터	회원
PRODUCT	제품	기준 데이터	제품

▶ 주요 항목 설명

항목명	설명
테이블	<ul style="list-style-type: none"> 물리 데이터 모델 혹은 데이터베이스 객체(Object)로써 존재함
엔티티명	<ul style="list-style-type: none"> 일반적으로 테이블의 한글명
데이터 유형	<ul style="list-style-type: none"> 기준 데이터, 거래 데이터 등 테이블에 저장된 데이터들의 성격
데이터 구조	<ul style="list-style-type: none"> 분류체계내의 위치명 예) (상위)주제영역-(하위)주제영역-엔티티(테이블)-속성(칼럼) 테이블의 상위 구조는 주제영역에 해당함

서식-22 핵심 품질 항목(CTQ) 정의서

▶ 핵심품질항목(CTQ) 정의서

분류		테이블			
저장영역	DBMS	주제영역	테이블명	테이블 한글명	유형분류명
수집	ORACLE	숙박	T_DD_ACDT_RT_LT	일일숙박대여내역	거래

컬럼				CTQ 항목					비고
컬럼명	컬럼 한글명	도메인명	데이터 타입	활용도수준	점수	순위	선정 여부	우선대상여부	
D_RT_CNT	일일대여건수		NUMBER	중	90	1	Y	Y	

▶ 주요 항목 설명

항목명	설 명
저장영역	<ul style="list-style-type: none"> 데이터 저장영역 명 (예 - 수집/저장, 가공/융합, 분석/활용 영역)
유형분류명	<ul style="list-style-type: none"> 데이터 유형 분류체계에 따른 데이터셋(테이블)의 유형분류 명 (예 - 메타, 기준, 로그, 거래, 분석, GIS, 문서, 미디어, 센서 활동 등)
도메인명	<ul style="list-style-type: none"> 도메인정의서에 정의된 도메인 명
활용도 수준	<ul style="list-style-type: none"> 해당 대상이 업무적으로 미치는 영향도(예 - 상, 중, 하)
선정여부	<ul style="list-style-type: none"> CTQ 선정 및 제외 여부 (예 - Y, N)
점수	<ul style="list-style-type: none"> CTQ 평가항목(영향도, 가중치) 합산 점수 (예- 0~100)
순위	<ul style="list-style-type: none"> 총점 기준 순위 (예 - 1~3순위)
선정여부	<ul style="list-style-type: none"> CTQ 항목 선정 여부
우선대상여부	<ul style="list-style-type: none"> CTQ 평가항목(영향도, 가중치)중 상위 우선여부(예 - Y, N)

(# 유형분류명 항목의 상세 설명은 “[서식-21] 데이터 분류체계 기준 정의서” 참고)

서식-23 업무규칙 정의서

▶ 업무규칙 정의서

분류		테이블			
저장영역	DBMS	주제영역	대상 테이블명	대상 테이블 한글명	유형분류명
수집	ORACLE	숙박	T_DD_ACDT_RT_LT	일일숙박대여내역	거래

컬럼								
컬럼명	컬럼 한글명	데이터 타입	컬럼길이	도메인명	PK 여부	FK 여부	NOT NULL 여부	CTQ 여부
D_RT_CNT	일일대여건수	NUMBER	6					Y

품질 지표 및 업무규칙						
품질지표 (DQI)	품질상세지표	업무규칙ID	업무규칙 내용	총건수	오류건수	오류내역
정확성	논리관계 정확성	BI_00001	일일숙박 대여건수는...	SELECT COUNT(*)	SELECT COUNT(*)	SELECT 숙박 대여일자

담당자		비고
업무규칙 작성일자	업무규칙 담당자	
20190930	홍길동	

▶ 주요 항목 설명

항목명	설명
저장영역	• 데이터 저장영역 명 (예 - 수집/저장, 가공/융합, 분석/활용)
DBMS	• DBMS(Oracle, MySql, DB2 등), NoSQL(MongGo DB, Neo4j, Redis) 등
주제영역	• 테이블의 주제영역
유형분류명	• 데이터 유형 분류체계에 따른 데이터셋(테이블)의 유형분류명 (예 - 메타, 기준, 거래, 로그, 분석, GIS, 문서, 미디어, 센서, 활동)
CTQ여부	• CTQ 항목 여부 (예 - Y, N)
품질지표(DQI)	• 데이터 품질관리 지표(유효성, 정확성, 완전성, 일관성 등)
품질상세지표	• 데이터 품질관리 상세지표(참조무결성, 중복일관성, 범위유효성, 논리관계정확성, 단독완전성 등)
업무규칙내용	• 품질진단을 수행 할 내용을 기반으로 진단스크립트(sql)를 작성
총건수	• 총건수를 구할 수 있는 스크립트(sql문)
오류건수	• 오류건수를 구할 수 있는 스크립트(sql문)
오류내역	• 업무규칙내용에 해당하는 오류내역을 구하는 스크립트(sql문)
업무규칙 작성일자	• 업무규칙을 작성한 일자
업무규칙 담당자	• 업무규칙을 도출한 담당자

(# 유형분류명 항목의 상세 설명은 “[서식-21] 데이터 분류체계 기준 정의서” 참고)

서식-24 품질 이슈 목록

▶ 품질 이슈 목록

이슈 번호	이슈 유형	이슈명	이슈 내용	시스템명	요청 부서	요청자	요청 일자	검토 부서	검토 담당자

▶ 주요 항목 설명

항목명	설명
이슈번호	• 품질이슈의 일련번호
이슈유형	• 품질이슈의 유형(예시- 데이터 값, 프로세스 등)
이슈명	• 품질이슈의 명칭
이슈내용	• 품질이슈의 상세내용
시스템명	• 품질이슈가 발생하는 시스템명
요청부서	• 품질이슈 요청부서
요청자	• 품질이슈 요청자 명
요청일자	• 품질이슈 요청일자
검토부서	• 품질이슈 검토부서
검토담당자	• 품질이슈 검토담당자 명

03

데이터 연계키 서식



서식-26 연계표준용어 신청서(신규등록)

연계표준용어 신청서(신규등록)		
신청기관	플랫폼명	
	소속부서	성명
	직급	전화번호/E-Mail
신청사유		
관련근거		
신청내역	용어명(한글명)	용어구성
	용어영문약어명	용어영문정식명
	데이터타입	데이터길이
	도메인유형	도메인한글명
	도메인영문명	

※ 신청내역은 [연계표준용어 정의서]를 참조하여 작성하시기 바랍니다.

[빅데이터 플랫폼 및 센터 품질관리 가이드]에 따라 상기와 같이 연계표준용어의 신규 등록을 신청합니다.

년 월 일

신 청 인 : (서명 또는 인)

04.

연계표준사전 서식



서식-26 연계표준용어 신청서(신규등록)

연계표준용어 신청서(신규등록)		
신청기관	플랫폼명	
	소속부서	성명
	직급	전화번호/E-Mail
신청사유		
관련근거		
신청내역	용어명(한글명)	용어구성
	용어영문약어명	용어영문정식명
	데이터타입	데이터길이
	도메인유형	도메인한글명
	도메인영문명	

※ 신청내역은 [연계표준용어 정의서]를 참조하여 작성하시기 바랍니다.

[빅데이터 플랫폼 및 센터 품질관리 가이드]에 따라 상기와 같이 연계표준용어의 신규 등록을 신청합니다.

신 청 인 : _____ 년 월 일
(서명 또는 인)

서식-27 연계표준용어 신청서(변경)

연계표준용어 신청서(변경)			
신청기관	플랫폼명		
	소속부서	성명	
	직급	전화번호/E-Mail	
변경신청사유			
관련근거			
변경신청내역	해당 연계표준용어 (용어명:한글명)		
	변경항목	변경 전	변경 후
<p>※ 신청내역은 연계표준용어 정의서를 참조하여 작성하시기 바랍니다.</p> <p>※ 변경은 DB에 직접적인 반영이 없는 항목이 변경되는 것을 의미합니다.</p> <p>[빅데이터 플랫폼 및 센터 품질관리 가이드]에 따라 상기와 같이 연계표준용어의 변경을 신청합니다.</p> <p style="text-align: right;">년 월 일</p> <p>신 청 인 : (서명 또는 인)</p>			

서식-28 연계표준용어 신청서(폐기 및 신규등록)

연계표준용어 신청서(폐기 및 신규등록)			
신청기관	플랫폼명		
	소속부서	성명	
	직급	전화번호/E-Mail	
변경신청사유			
관련근거			
신청내역 (폐기 및 신규등록)	해당 연계표준용어 (용어명:한글명)		
	변경항목	변경 전(폐기)	변경 후(신규등록)
	용어명		
	용어구성		
	용어영문약어명		
	용어영문정식명		
	데이터타입		
	데이터길이		
	도메인유형		
	도메인한글명		
	도메인영문명		
<p>※ 신청내역은 연계표준용어 정의서를 참조하여 작성하시기 바랍니다.</p> <p>※ DB에 직접적인 반영이 필요한 항목이 변경되는 것을 의미하며, 폐기와 신규등록을 동시에 신청해야 합니다.</p>			

[빅데이터 플랫폼 및 센터 품질관리 가이드]에 따라 상기와 같이
연계표준용어의 변경을 신청합니다.

신 청 인 : _____ 년 월 일 (서명 또는 인)

서식-29 연계표준단어 신청서(신규등록)

연계표준단어 신청서(신규등록)		
신청기관	플랫폼명	
	소속부서	성명
	직급	전화번호/E-Mail
신청사유		
관련근거		
신청내역	단어명(한글명)	
	신규등록 신청 연계표준용어 관련 여부 <input type="checkbox"/> 해당(관련 신청 용어명 :) <input type="checkbox"/> 해당없음	
	단어영문명	영문약어명
	단어정의	

※ 신청내역은 [연계표준단어 정의서]를 참조하여 작성하시기 바랍니다.

[빅데이터 플랫폼 및 센터 품질관리 가이드]에 따라 상기와 같이 연계표준단어의 신규 등록을 신청합니다.

신 청 인 : (서명 또는 인)

서식-30 연계표준단어 신청서(변경)

연계표준단어 신청서(변경)			
신청기관	플랫폼명		
	소속부서	성명	
	직급	전화번호/E-Mail	
변경신청사유			
관련근거			
변경신청내역	해당 연계표준단어 단어명(한글명)		
	변경항목	변경 전	변경 후
	연계표준단어 정의		
<p>※ 신청내역은 [연계표준단어 정의서]를 참조하여 작성하시기 바랍니다.</p> <p>[빅데이터 플랫폼 및 센터 품질관리 가이드]에 따라 상기와 같이 연계표준단어의 변경을 신청합니다.</p> <p style="text-align: right;">년 월 일</p> <p>신 청 인 : (서명 또는 인)</p>			

서식-31 연계표준단어 신청서(폐기 및 신규등록)

연계표준단어 신청서(폐기 및 신규등록)			
신청기관	플랫폼명		
	소속부서	성명	
	직급	전화번호/E-Mail	
변경신청사유			
관련근거			
신청내역 (폐기 및 신규등록)	해당 연계표준단어 단어명(한글명)		
	폐기 및 신규등록 용어 관련 여부	[] 해당 (관련 신청용어명 :) [] 해당없음	
	변경항목	변경 전	변경 후
	단어명		
	단어영문명		
	영문약어명		
<p>※ 신청내역은 [연계표준단어 정의서]를 참조하여 작성하시기 바랍니다.</p> <p>※ DB에 직접적인 반영이 필요한 항목이 변경되는 것을 의미하며, 폐기 및 신규 등록을 동시에 신청해야 합니다.</p> <p>[빅데이터 플랫폼 및 센터 품질관리 가이드]에 따라 상기와 같이 연계표준단어의 변경을 신청합니다.</p> <p style="text-align: right;">년 월 일</p> <p>신 청 인 : (서명 또는 인)</p>			

서식-32 연계표준도메인 신청서(신규등록)

연계표준도메인 신청서(신규등록)		
신청기관	플랫폼명	
	소속부서	성명
	직급	전화번호/E-Mail
신청사유		
관련근거		
신청내역	도메인·대분류명 (신규등록 여부 : <input type="checkbox"/> 기등록, <input type="checkbox"/> 신규등록)	
	도메인·중분류명 (신규등록 여부 : <input type="checkbox"/> 기등록, <input type="checkbox"/> 신규등록)	
	도메인 한글명	
	신규등록 신청 표준용어 관련 여부 <input type="checkbox"/> 해당 (관련 신청용어명 :) <input type="checkbox"/> 해당없음	
	도메인영문명	데이터타입
	데이터길이	
	정의설명	
<p>※ 신청내역은 [연계표준도메인 정의서]를 참조하여 작성하시기 바랍니다.</p> <p>[빅데이터 플랫폼 및 센터 품질관리 가이드]에 따라 상기와 같이 연계표준도메인의 신규 등록을 신청합니다.</p> <p style="text-align: right;">년 월 일</p> <p>신 청 인 : (서명 또는 인)</p>		

서식-33 연계표준도메인 신청서(변경)

연계표준도메인 신청서(변경)			
신청기관	플랫폼명		
	소속부서	성명	
	직급	전화번호/E-Mail	
신청사유			
관련근거			
신청내역	변경 도메인한글명		
	변경항목	변경 전	변경 후
	도메인 대분류		(신규등록 여부 : <input type="checkbox"/> 기등록, <input type="checkbox"/> 신규등록)
	도메인 중분류		(신규등록 여부 : <input type="checkbox"/> 기등록, <input type="checkbox"/> 신규등록)
	정의설명		
<p>※ 신청내역은 [연계표준도메인 정의서]를 참조하여 작성하시기 바랍니다.</p> <p>[빅데이터 플랫폼 및 센터 품질관리 가이드]에 따라 상기와 같이 연계표준도메인의 변경을 신청합니다.</p> <p style="text-align: right;">년 월 일</p> <p>신 청 인 : (서명 또는 인)</p>			

94 빅데이터 플랫폼 및 센터 데이터 품질관리 가이드

01

개요



1.1 데이터 연계기 개요

1.1.1 데이터 연계기 등장 배경

오늘날의 비즈니스 환경은 서로 다른 업종 간의 데이터를 융·복합 분석하여 새로운 비즈니스 기회를 창출하고자 하는 요구들이 지속해서 증가하고 있다. 이를 위해 빅데이터 플랫폼 및 센터에서 수집된 다양한 유형의 데이터를 플랫폼 간 또는 센터들과 연계하여 데이터 융·복합 서비스를 제공하기 위해서는 데이터를 연계 할 수 있는 ‘데이터 연계기’가 필요하다. 서로 다른 플랫폼의 다양한 데이터를 연계·융합하여 빅데이터를 발굴 확산할 필요성이 대두되었으며 이를 위하여 데이터 간 연계를 위한 ‘데이터 연계기’ 필요성이 대두되었다.

그림 1 상권분석을 위해 은행 및 통신자료의 주소기반 연계기 필요

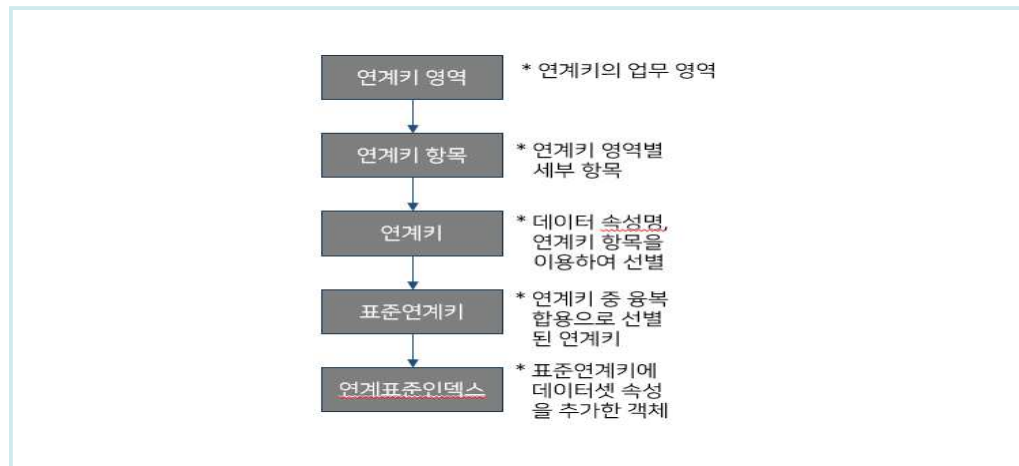


1.1.2 데이터 연계기 정의

- ▶ **연계기 영역** : 다양한 분야에서 활용될 연계기의 업무 영역으로 기본 영역은 사람, 기업, 공간, 시간 영역으로 정의하며 업무의 활용도 등을 고려하여 연계기 영역은 확장하여 사용한다. 연계기가 속하는 일종의 업무영역의 대분류명이다.
- ▶ **연계기 항목** : 연계기 영역별로 세부 특성을 식별할 수 있는 항목으로 연계기 영역별 세부 항목이다.
- ▶ **연계기** : 데이터셋간의 상호 연계를 위한 속성(컬럼)이다.
- ▶ **표준 연계기** : 서로 다른 데이터셋을 연계하는 데 이용되는 데이터셋의 논리속성명으로 연계기에서 선별하며, 빅데이터 유통 플랫폼의 다양한 분야에서 공통으로 활용 할 연계표준용어 기반 연계기이다.
- ▶ **표준 연계기 사전** : 표준연계기를 관리하는 사전, 일종의 표준 연계기 목록이다.
- ▶ **연계표준인덱스** : 표준연계기사전에 등록된 표준연계기에 매핑된 각 플랫폼의 데이터셋 속성을 데이터 레코드 형태로 저장된 데이터, 연계표준인덱스 속성에는 플랫폼명, 데이터셋명, 속성명 등이 있고 이들은 데이터셋의 융복합을 위한 인덱스 역할을 수행한다.

1.1.3 연계기와 표준연계기 관계

그림 2 연계기와 표준연계기 구조



02

데이터 연계기 발굴



2.1 표준연계기 발굴방안

2.1.1 표준연계기 필요성

빅데이터 유통 플랫폼 환경에서는 각 플랫폼이 자체 용어사전을 적용하여 다른 플랫폼과 상이한 속성(컬럼)명을 사용하고 있다. 이와 같은 환경에서는 서로의 데이터셋 정보를 파악하기 쉽지 않아 데이터 연계가 어렵다. 이에 빅데이터 유통 플랫폼에서는 데이터를 공통으로 활용할 연계표준사전 기반 연계기의 발굴이 필요하며, 이를 기반으로 표준연계기를 추출하고 연계표준인덱스를 구축할 필요성이 있다. 그 결과 서로 다른 플랫폼별 용어사전의 호환성을 확보할 수 있고, 플랫폼 간 데이터를 연계할 수 있다.

그림 3 연계기, 표준연계기, 연계표준인덱스 관계



2.1.2 표준연계기 발굴범위

다양한 분야에서 공통으로 활용할 표준연계기는 먼저 ‘사람’, ‘기업’, ‘공간’ 및 ‘시간’ 영역에서 발굴한다. 이들 영역은 그동안 활용도가 높았던 분야로, 이들을 중심으로 우선 표준연계기가 발굴되어 활용하도록 한다. 예를 들어 사람영역에서는 표준연계기를 성, 연령, 연령대, 이메일, 거주지 주소, 우편번호 등으로 볼 수 있다. 이와 같이 각 영역에서 활용도를 기초로 표준연계기를 발굴하도록 한다.

표 1 표준연계기 발굴을 위한 연계기 기본영역

연계기 영역	표준 연계기 발굴을 위한 연계기 항목
사람	• 성, 연령, 연령대, 이메일, 거주지 주소, 우편번호
기업	• 사업체 명칭, 사업체 전화번호, 사업체 주소, 법인번호, 업종
시간	• 시간, 시각, 날짜, 연도
공간	• 주소, 좌표, 행정구역, 국가기초구역, 건물 (집합건물, 빌딩)

표준연계기에 적용할 영역은 그동안 활용도가 높았던 ‘사람’, ‘기업’, ‘공간’, ‘시간’ 등의 범위에서 정의한다.

- ▶ ‘사람’ 영역에 해당하는 표준연계기는 사람의 특성을 식별할 수 있는 성별코드, 연령, 연령대 코드 등의 항목을 이용한다.
- ▶ ‘기업’ 영역에 해당하는 표준연계기는 기업을 식별할 수 있는 사업체 명칭, 전화번호 등의 항목을 이용한다.
- ▶ ‘시간’ 영역에 해당하는 표준 연계기는 시간을 구분하는 연도, 연/월/일 등의 항목을 이용한다.
- ▶ ‘공간’ 영역에 해당하는 표준연계기는 지역을 식별할 수 있는 주소, 좌표, 행정구역 등의 항목을 이용한다.

2.1.3 표준연계기 발굴항목

표준연계기 영역별 항목은 플랫폼 및 센터에서 데이터 연계·융합을 위하여 활용 할 것을 권고하며, 필요한 항목을 선택하여 활용하되 개인정보 이슈가 있는 항목에 대해서는 반드시 익명화 등의 조치를 거쳐 활용하여야 한다.

표 2 사람영역에서 활용을 권장하는 표준 연계기 항목

연계기 항목	연계기 데이터 포맷/코드값	적용예시(설명)
성별코드 (SEXSTN_CODE)	M/F	M(남자), F(여자)
연령(AGE)	99	23, 45
연령대코드 (AGRDE_CODE)	10- 10,20,30,40,50,60,70,80,90, 90+	10-(10대 미만), 50 (50대), 90+(90대 초과)
연령구간코드 (AGE_SCTN_CODE)	00~04,05~09,10~14,15~19, 20~24,25~29,30~34,35~39, 40~44,45~49,50~54,55~59, 60~64,65~69,70~74,75~79, 80~84,85~89,90~94,95~99, 99+	20~24, 99+(99세 초과)
생일년도 (BRTHDY_YEAR)	9999	1966, 2013
생일년월 (BRTHDY_YM)	YYYY-MM	2019-01

표 3 기업영역에서 활용을 권장하는 표준 연계기 항목

연계기 항목	연계기 데이터 포맷/코드값	적용예시(설명)
사업체명(BSNES_NM) 업체명(ENTRPS_NM) 기업명(ENTRPRS_NM)	(A)AAAA	
전화번호	999-9999-9999 9999-9999-9999	032-6879-1029 0594-0862-9495
법인등록번호 (JURIRNO)	999999-99999999	110111-3430040
사업자등록번호 (BIZRNO)	999-99-99999	211-87-00928
홈페이지	aaa.aaaaaaa.aa.aa aaa.aaaaaaa.aaa	www.dgs.co.kr www.raphas.com
표준산업업종분류코드	99999	57112

표 4 시간영역에서 활용을 권장하는 표준 연계기 항목

연계기 항목	연계기 데이터 포맷/코드값	적용예시(설명)
년, 년도	YYYY	2019
년월	YYYY-MM	2019-08
일자	YYYY-MM-DD	2019-10-12
일시	YYYY-MM-DD HH24:MI	2019-10-12 20:18
일시분초	YYYY-MM-DD HH24:MI:SS	2019-10-12 20:18:33
일시분초밀리초	YYYY-MM-DD HH24:MI:SS.SSS	2019-10-12 20:18:33.001
월	MM	01 ~ 12
주	WWW	W01 ~ W52
일	DD	01 ~ 31
시간	HH24:MI~HH24:MI	20:18~20:58
시각	HH24:MI	20:18
시분초	HH24:MI:SS	20:18:33
년주	YYYY-WWW	2019-W01

표 5 공간영역에서 활용을 권장하는 표준 연계기 항목

연계기 항목	연계기 데이터 포맷/코드값	적용예시(설명)
시도코드	99	11
시도명	AAAAAA	부산광역시
시군구코드	99999	110
시군구명	AAAAAA	종로구, 광산구
법정동코드	9999999999	4884025021
행정동코드	9999999999	1111051500

연계기 항목	연계기 데이터 포맷/코드값	적용예시(설명)
법정동명	AAAAAA	전라남도 구례군 산동면 원촌리
행정동명	AAAAAA	서울시 성동구 성수2가3동
지번상세주소번지	9999-9999	141-51번지
지번상세주소건물명	AAAAAA	신창빌딩
지번상세주소동호	9999-9999, 9999	103-3044호, 394호
도로상세주소도로명	AAAAAA길, AAAA99길, AAA99A길, AAAAAA로, AAAAAA대로	지명로, 이산12로, 세종로3가길, 서평로, 세종대로
도로상세주소건물번호	9999, 9999-9999	1293, 13-45
도로명상세주소부가명	(AAAAAA,AAAAAA 9999-9999)	(명륜1가, 렉스빌 아파트 4동1023호)
좌표X	999999	322622, 430192
좌표Y	999999	522195, 596021
위도	999.9999	127.3049
경도	999.9999	37.4958
신우편번호	99999	53446, 85674
집합건물명	AAAAA	대림이편한세상아파트
집합건물동호	9999-9999	245-1209 (245동 1209호)

2.2 데이터 연계기 발굴

2.2.1 연계표준인덱스 발굴

표준 연계기 영역 및 표준 연계기 항목을 이용하여 빅데이터 연계표준사전(용어사전 등)에서 표준연계기를 발굴하고 빅데이터 융·복합 서비스를 개발할 수 있도록 연계표준인덱스를 구현한다.

빅데이터 플랫폼 및 센터의 데이터를 이용한 융·복합 서비스를 제공하기 위하여 연계표준인덱스를

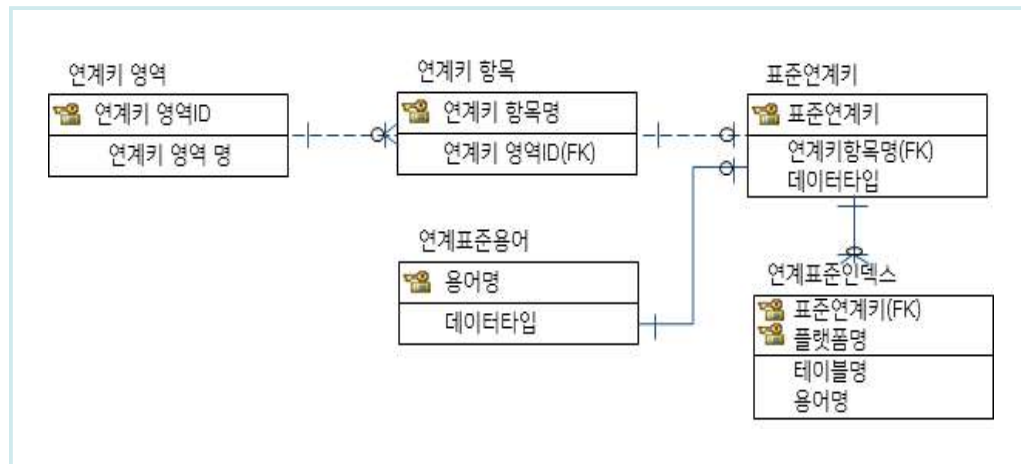
발굴하되, 그동안 활용도가 높았던 분야로 평가되는 '사람', '기업', '공간' 및 '시간' 영역에서 우선적으로 발굴한다.

사람, 기업 등 연계기 영역에서 활용도가 높다고 선정되어 있는 연계기 항목을 활용하여 표준연계기를 발굴하며, 연계표준사전(용어사전 등)에 정규표현식 등을 적용하여 표준연계기를 발굴할 수 있다. 발굴된 표준연계기에 플랫폼의 데이터셋 속성 정보 등을 추가하여 연계표준인덱스를 발굴한다.

2.2.2 논리 다이어그램

연계기 영역, 연계기 항목 및 연계표준인덱스를 위한 논리 다이어그램은 다음과 같다.

그림 4 연계표준인덱스 발굴을 위한 논리 다이어그램(예시)



2.2.3 물리 다이어그램

연계표준인덱스를 위한 물리 다이어그램은 빅데이터 융·복합 서비스 발굴에 적합하게 비정규화 등을 수행한 물리 모델링 후 연계표준인덱스를 구현한다.

그림 5 물리 다이어그램(예시1)

표준연계기 (용어명)	연계기항목	데이터타입	데이터길이	플랫폼01		플랫폼02	
				연계기(용어명)	테이블ID	연계기(용어명)	테이블ID
시작일자	일자	VARCHAR2	10	해당값 X	해당값 X	해당값 X	해당값
종료일자	일자	VARCHAR2	10	해당값 X	해당값 X	해당값 X	해당값
연령대코드	연령대코드	VARCHAR2	3	AGRDE_CD	KC_SMGU_RSD T_SUM_TB	AGGR_CD	TBBP10_CC_
상품명	-	VARCHAR2	255	해당값 X	해당값 X	해당값 X	해당값

그림 6 물리 다이어그램(예시2)

표준연계기 (용어명)	연계기항목	데이터타입	데이터길이	표준연계기 (용어명)	플랫폼명	테이블ID	테이블명	컬럼명
시작일자	일자	VARCHAR2	10	시작일자	플랫폼A	TABLE-A-1	테이블가1	영업시작일자
종료일자	일자	VARCHAR2	10	시작일자	플랫폼A	TABLE-B	테이블나	작업시작일자
연령대코드	연령대코드	VARCHAR2	3	시작일자	플랫폼B	TABLE-B-1	테이블나1	시작일자
상품명	-	VARCHAR2	255	시작일자	플랫폼C	TABLE-C-1	테이블다	검사시작일자

2.2.3 연계표준인덱스 발굴

연계표준인덱스 발굴을 위하여 연계기 항목에 해당하는 정규표현식 등을 연계표준사전(연계표준용어, 컬럼정의서 등)의 각 속성에서 찾는다.

연계표준인덱스 발굴을 위해 그동안 활용도가 높았던 분야로 평가되는 '사람', '기업', '공간' 및 '시간'의 연계기 영역에서 우선적으로 발굴하고, 융·복합서비스의 제공 범위에 따라 연계기 영역의 확대와 정규표현식 이외에 추가적인 방식을 도출하여 연계표준인덱스를 발굴한다.

연계표준인덱스 발굴유형 및 기준에 대한 예시는 다음과 같으며, 발굴유형은 (1) 용어사전의 용어영문약어명을 이용한 정규식, (2) 용어구성을 이용한 정규식, (3) 용어영문약어명을 이용한 정규식과 도메인한글명을 이용한 정규식, (4) 용어영문약어명을 이용한 정규식과 도메인유형을 이용한 정규식을 이용했다.

예시의 발굴유형은 연계표준용어만을 이용했으며 업무의 활용에 따라 데이터 값을 이용한 발굴유형을 개발할 필요가 있다.

표 6 연계표준인덱스 발굴 방법(예시)

표준연계기 항목			발굴방법		
영역	연계기 항목	연계기 데이터 포맷/코드값	발굴유형	발굴기준1	발굴기준2
사람	성별코드	M/F	2	성별_코드	
사람	연령	99	1	[^A-Z]AGE[^A-Z]	
사람	연령대코드	10-,10,20,30,40,50,60,70,80,80+	1	[^A-Z]AGE[0-9]*[^A-Z]	
사람	연령구간코드	00~04,05~09,10~14,15~19,-,99+	1	[^A-Z]AGE[0-9]*[^A-Z]	
사람	생일연도	9999	1	BRY Y	
사람	생일연월	YYYY-MM	1	BRY M	
사람	이메일	AAAA@AAAA	1	EMAIL\$	
사람	주소	공간영역 주소항목과 동일	1	_ADRES\$	
시간	년	YYYY	1	YY	
시간	년도	YYYY	3	YR YY	년도_
시간	년월	YYYY-MM	4	YM	시간
시간	일자	YYYY-MM-DD	4	DD	시간
시간	일시	YYYY-MM-DD HH24:MI	4	DTM	시간
시간	월	MM	3	MN	월_C2
시간	일	DD	3	_DAY	일_V2 수_V2
시간	시간	HH24:MI~HH24:MI	3	TM	시간_
시간	시분초	HH24:MI:SS	4	HMS	시간
기업	업체명	(A)AAAA	2	업체_명	
기업	기업명	(A)AAAA	2	기업_명	
기업	전화번호	999-9999-9999, 9999-9999-9999	1	TELNO\$	
기업	법인등록번호	999999-99999999	1	(CPR).{0,}(CD)	
기업	사업자등록번호	999-99-999999	1	BIZRNO	
기업	홈페이지		1	HMPG_URL	
공간	시도코드	99	2	시도_코드	
공간	시도명	AAAAAA	2	시도_명	
공간	시군구코드	99999	1	SIGNGU_CD	
공간	시군구명	AAAAAA	1	SIGNGU_NM	
공간	법정동코드	1E+10	2	법정동_코드	
공간	행정동코드	1E+10	2	행정동_코드	

표준연계기 항목			발굴방법		
영역	연계기 항목	연계기 데이터 포맷/코드값	발굴유형	발굴기준1	발굴기준2
공간	법정동명	AAAAAA	2	법정동_명	
공간	행정동명	AAAAAA	2	행정동_명	
공간	지번상세주소번지	9999-9999	2	_상세주소	
공간	지번상세주소건물명	AAAAAA	2	건물_명	
공간	도로상세주소도로명	AAAAAA길, AAAA99길, AAAAAA로, AAAAAA대로	2	_도로_명	
공간	도로상세주소건물번호	9999, 9999-9999	2	_건물번호	
공간	좌표X	999999	1	XCRD	
공간	좌표Y	999999	1	YCRD	
공간	위도	999.9999	4	LTD	수 값 길이
공간	경도	999.9999	4	[^A-Z]LO[^A-Z]	수 값 길이
공간	우편번호	AAAAA	1	ZIP	
공간	우편번호주소	9999-9999	1	ZIP_AD[A-Z]*	

그림 7 연계표준인덱스 발굴방법을 이용한 발굴 결과(예시)

용어명	연계기, 영역	연계기, 항목명	A 플랫폼	B 플랫폼	C 플랫폼	D 플랫폼
가맹점우편번호	공간	우편번호	TBBP10_CC_30	해당값 X	해당값 X	해당값 X
가맹점주소	사람	주소	TBBP10_CC_09	해당값 X	TB_DPIS_KC_604_CLT_NURI_CARD_MAP_STATN_BIZAE_A	해당값 X
건물명	공간	지번상세주소 건물명	TBBP10_CC_09	해당값 X	TB_DPIS_KC_BCYCL_RECO_MEND_STRET_LIST	해당값 X
건축년도	시간	년도	TBBP19_RE_12	해당값 X	해당값 X	해당값 X
관리기관전화번호	기업	전화번호	해당값 X	rx_fusion_purification_reservoir_water_quality_info	TB_DPIS_BC_WNTY_AREA_SPCLIZ_DSTNC_STD_DATA	해당값 X
광역시도명	공간	시도명	TBBP18_FP_04	해당값 X	해당값 X	해당값 X
기업명	기업	기업명	해당값 X	해당값 X	해당값 X	해당값 X
기준년도	시간	년도	TBBP14_ID_03	nc_road_cfr_pcind_in fo_stats	TB_DPIS_BC_BSL_AGE_TYP_E	해당값 X
기준년월	시간	년월	TBBP10_CC_01	rx_fusion_fine_dust_s ales_info	TB_DPIS_KCB_ADSTRD_DAT A1_02	해당값 X
년도	시간	년도	TBBP19_RE_23	ke_population_ivstg	TB_DPIS_AK_ART_CENTER_MEDIA_DATA_LIST	해당값 X

03

데이터 연계기 관리체계



3.1 데이터 연계기 관리조직

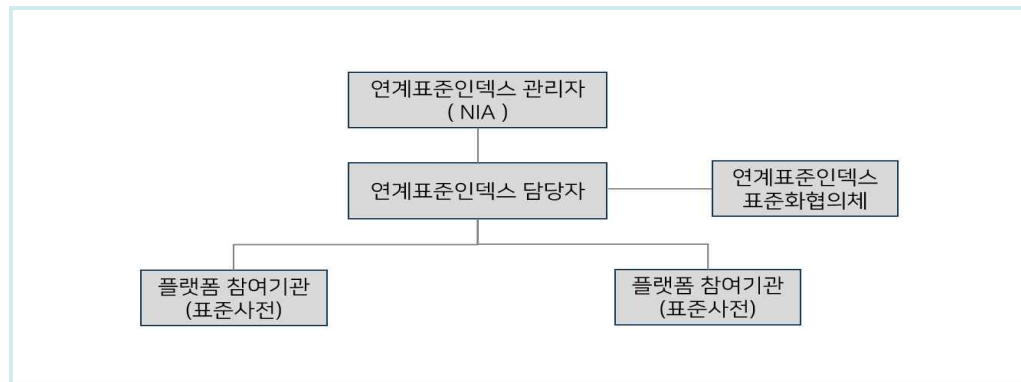
3.1.1 연계표준인덱스 관리조직

한국지능정보사회진흥원은 연계표준인덱스 정책을 총괄하며 연계표준인덱스의 추가, 수정, 폐기 등 업무를 수행하며, 필요시 연계표준인덱스 담당자와 관리자를 지정하여 연계표준인덱스를 관리한다.

연계표준인덱스 표준화 사안 협의·조정 등을 위해 필요시 연계표준인덱스 표준화 협의체를 운영한다.

빅데이터 플랫폼 및 센터는 확정된 연계표준인덱스를 이용하여 빅데이터 융·복합서비스를 발굴하고 연계표준인덱스의 추가, 수정 등 보완이 필요한 사항은 연계표준인덱스 담당자에게 연계기 영역, 연계기 항목, 표준연계기, 연계표준인덱스 등의 변경을 신청한다.

그림 8 연계표준인덱스 발굴 및 관리를 위한 조직



3.1.2 연계표준인덱스 조직별 역할

표 7 연계표준인덱스 조직별 역할

연계기 영역	표준 연계기 발굴을 위한 연계기 항목
연계표준인덱스 관리자	<ul style="list-style-type: none"> 연계표준인덱스의 최종 승인 및 확정
연계표준인덱스 담당자	<ul style="list-style-type: none"> 연계표준인덱스의 연구개발 및 관리 등 업무 수행 연계기 영역, 연계기 항목, 표준연계기 발굴 및 관리기준 수립 연계표준인덱스 발굴 및 전파 연계표준인덱스 협의체 관리
연계표준인덱스 표준화 협의체	<ul style="list-style-type: none"> 연계표준인덱스 발굴 및 연계표준인덱스 활성화를 주도적으로 추진하고, 연계표준인덱스와 관련된 이슈를 협의 조정
플랫폼 참여기관	<ul style="list-style-type: none"> 최종 확정된 연계표준인덱스를 이용하여 빅데이터 융복합서비스 제공 추가적인 빅데이터 융복합서비스를 제공하기 위한 연계기 항목, 표준연계기 신규 발굴 및 연계표준인덱스로 등록 요청 <p>[플랫폼 및 센터 담당자]</p> <ul style="list-style-type: none"> 확정된 연계표준인덱스를 이용하여 플랫폼 및 센터의 서비스 개발 추가적인 빅데이터 서비스 발굴을 위한 연계표준인덱스 발굴 및 등록 신청

3.2 연계표준인덱스 등록 및 폐기

3.2.1 연계표준인덱스 신규 등록

빅데이터 융·복합서비스 개발을 위한 표준연계기를 연계표준인덱스로 신규등록하는 것을 의미하며, 플랫폼 참여기관의 신청과 한국지능정보사회진흥원의 자체 발굴에 의하여 등록된다.

각 플랫폼 참여기관은 플랫폼 표준용어로 정의된 용어가 빅데이터 융·복합서비스 개발에 공동·활용이 된다고 판단되면 연계표준인덱스로 등록 신청할 수 있다. 연계표준인덱스를 운영하는 한국지능정보사회진흥원은 플랫폼의 표준용어, 테이블정의서 등을 분석하여 연계표준인덱스 등록 대상으로 발굴할 수 있다.

연계표준인덱스 신규 등록 시 연계기 영역, 연계기 항목의 신규등록이 필요한지 검토하여, 신규등록이 필요한 경우 다음과 같이 등록한다.

그림 9 연계표준인덱스 신규등록 절차

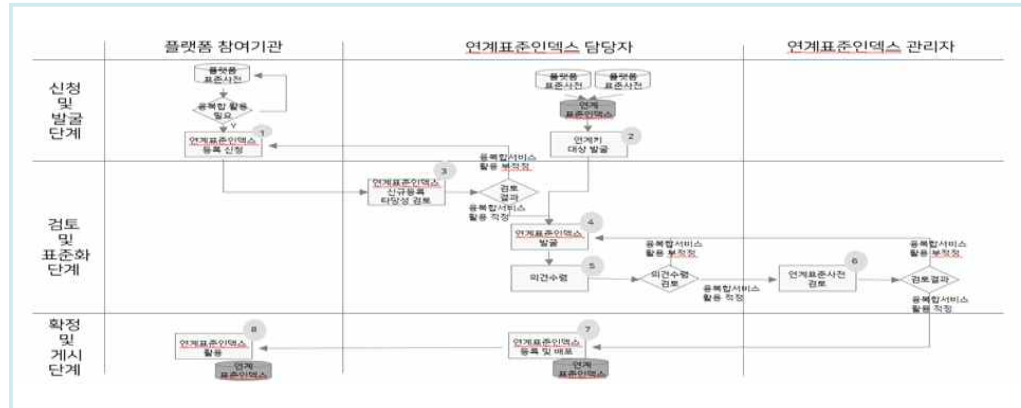


표 8 연계표준인덱스 신규등록 세부절차

구분	절차	절차내용	수행주체
신청 및 발굴 단계	1. 연계표준인덱스 등록 신청	<ul style="list-style-type: none"> 플랫폼 참여기관은 전 플랫폼이 융복합서비스 발굴에 필요하다고 판단되는 연계표준용어를 선정하여 연계표준인덱스에 등록을 신청 해당 신청 연계표준인덱스와 관련하여 연계기 영역 및 연계기 항목의 신규 생성이 필요한 경우 같이 신청 	플랫폼 참여기관
	2. 연계표준인덱스 대상 발굴	<ul style="list-style-type: none"> 연계표준인덱스 담당자는 플랫폼 표준용어 및 빅데이터 융복합서비스 현화를 정기적으로 수집·분석하여 연계표준인덱스 추가 등록 대상을 발굴 	연계표준인덱스 담당자

구분	절차	절차내용	수행주체
검토 및 확정 단계	3. 연계표준인덱스 신규등록 타당성 검토	<ul style="list-style-type: none"> 연계표준인덱스 담당자는 연계표준인덱스 등록 신청내역을 검토하여 타당성 분석 실시 기존 연계표준인덱스와의 중복 여부 검토 연계표준인덱스 신청시 표준연계사전 등록 여부 및 빅데이터 융복합서비스 발굴에 활용 가능성 등을 검토하며, 연계기 영역과 연계기 항목 검토 	연계표준인덱스 담당자
	4. 연계표준인덱스 발굴	<ul style="list-style-type: none"> 연계표준인덱스 담당자는 플랫폼에서 신청한 연계표준인덱스 및 자체적으로 발굴한 연계표준인덱스를 대상으로 연계표준인덱스를 발굴 연계표준인덱스 발굴 내용 <ul style="list-style-type: none"> 빅데이터 융복합서비스 활용 가능성 점검 기존 연계표준인덱스에 등록 여부 확인 기존 연계기 영역, 연계기 항목 등록 여부 확인 	연계표준인덱스 담당자
	5. 의견수렴	<ul style="list-style-type: none"> 연계표준인덱스 담당자는 필요시, 외부전문가 및 협의체, 플랫폼에 의견수렴하고 연계표준인덱스 관리자에게 연계표준인덱스의 확정을 요청 	연계표준인덱스 담당자
	6. 연계표준인덱스 검토 및 확정	<ul style="list-style-type: none"> 연계표준인덱스 관리자는 신규 연계표준인덱스를 검토 한 후 연계표준인덱스를 최종 확정 	연계표준인덱스 관리자
	7. 연계표준인덱스 등록 및 배포	<ul style="list-style-type: none"> 연계표준인덱스 담당자는 확정된 연계표준인덱스(연계기 영역, 연계기 항목 포함)를 전 플랫폼에서 적용할 수 있도록 배포 및 관리 	연계표준인덱스 담당자
배포 및 활용 단계	8. 연계표준인덱스 활용	<ul style="list-style-type: none"> 모든 플랫폼은 확정된 연계표준인덱스(연계기 영역, 연계기 항목 포함)를 이용하여 빅데이터 융복합서비스 발굴 	플랫폼 및 센터 담당자

3.2.2 연계표준인덱스 폐기

연계표준인덱스(영역, 항목 포함)의 폐기는 기존에 사용하던 연계표준인덱스를 더 이상 사용하지 않는 것을 의미하며, 사용 중인 연계표준인덱스가 전 플랫폼의 데이터베이스에서 사용하지 않게 되는 경우와 연계표준인덱스가 빅데이터 융·복합서비스를 발굴하는데 적합하지 않은 경우 폐기절차에 따라 연계표준인덱스를 폐기할 수 있다.

그림 10 연계표준인덱스 폐기 절차

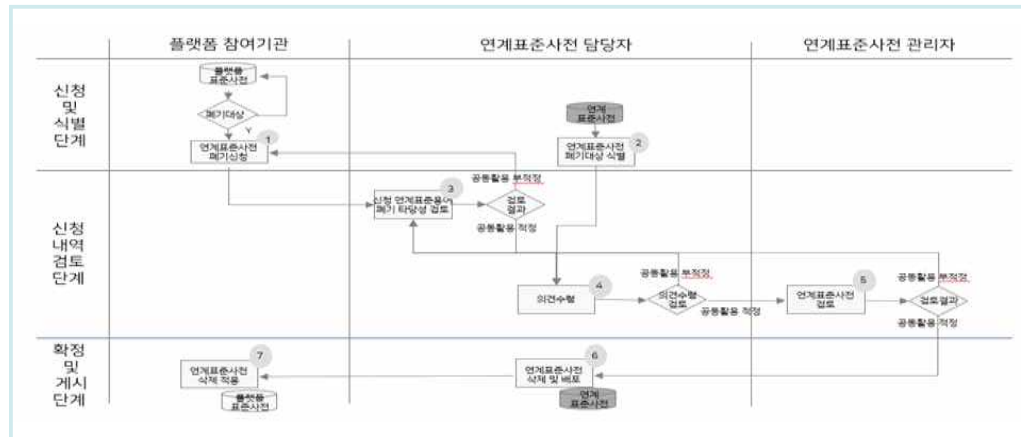


표 9 연계표준인덱스 폐기 세부절차

구분	절차	절차내용	수행주체
신청 및 식별 단계	1. 연계표준인덱스 폐기 신청	<ul style="list-style-type: none"> 플랫폼 참여기관은 연계표준인덱스 중 용·복합서비스 발굴에 적합하지 않은 경우 해당 연계표준인덱스에 대한 폐기를 신청 연계키 영역, 연계키 항목, 연계표준인덱스의 변경이 발생하는 경우 	플랫폼 참여기관
	2. 연계표준인덱스 폐기대상 식별	<ul style="list-style-type: none"> 용·복합서비스 개발에 부적합 하거나, 시스템 폐기 등으로 연계표준인덱스(연계키 영역, 연계키 항목)가 더 이상 사용되지 않는 경우 연계표준인덱스 담당자는 연계표준인덱스 폐기 대상을 식별 	연계표준인덱스 담당자
검토 및 확정 단계	3. 연계표준인덱스 폐기 타당성 검토	<ul style="list-style-type: none"> 연계표준인덱스 담당자는 연계표준인덱스 폐기 신청내역을 검토 연계표준인덱스, 연계키 영역, 연계키 항목의 폐기 타당성 검토 	연계표준인덱스 담당자
	4. 의견수렴	<ul style="list-style-type: none"> 연계표준인덱스 담당자는 필요시, 외부전문가 및 협의체, 플랫폼의 의견을 수렴하고 연계표준인덱스 관리자에게 연계표준인덱스의 확정을 요청 	연계표준인덱스 담당자
	5. 연계표준인덱스 검토	<ul style="list-style-type: none"> 연계표준인덱스 관리자는 폐기 연계표준인덱스를 검토 한 후 폐기 대상으로 최종 확정 	연계표준인덱스 관리자
배포 및 활용 단계	6. 연계표준인덱스 삭제 및 배포	<ul style="list-style-type: none"> 연계표준인덱스 담당자는 폐기된 연계표준인덱스를 전 플랫폼에서 적용할 수 있도록 배포 및 관리 	연계표준인덱스 담당자
	7. 연계표준인덱스 활용	<ul style="list-style-type: none"> 모든 플랫폼은 삭제된 연계표준인덱스를 이용하여 빅데이터 융복합서비스 발굴 	플랫폼 및 센터 담당자

01

개요



1.1 연계표준사전 개요

1.1.1 연계표준사전 필요성

빅데이터 플랫폼 및 센터 상호운용성(interoperability) 확보, 타 기관이나 플랫폼 간의 데이터 연계를 위해 데이터 표준화에 기초한 연계표준사전의 필요성이 대두 되었다.

각 빅데이터 플랫폼이 데이터 용·복합서비스를 제공·활용하기 위해서는 플랫폼별 독자적인 표준을 수립하여 적용하면 동일한 의미를 갖는 용어가 플랫폼 및 센터별로 다양한 명칭과 형식으로 정의되어 데이터의 활용에 어려움을 겪게 된다.

또한 데이터 활용이 증가함에 따라 데이터 수요자는 빅데이터 플랫폼 이외에 민간 플랫폼 등 여러 기관의 데이터를 융·복합 및 연계하여 활용하고자 하는데 이는 데이터 표준화에 기초한 연계표준사전의 중요성이 강조되는 상황이다.

연계표준사전은 데이터 표준화에 기초하여 빅데이터 플랫폼 및 센터의 데이터를 용·복합서비스에 활용될 수 있도록 [공공기관의 데이터 표준화 지침]을 근거로 하여 빅데이터 플랫폼의 표준사전을 이용하여 작성된 표준자료(연계표준용어, 연계표준단어, 연계표준도메인)이다.

1.1.2 연계표준사전 구성요소

빅데이터 플랫폼의 데이터 표준화는 연계표준사전의 표준화이며, 연계표준사전은 [공공기관의 데이터 표준화 지침]의 공통표준사전 및 빅데이터 플랫폼의 개별사전을 이용하여 용·복합서비스에 활용될 수 있는 표준용어를 선별하여 생성하며, 연계표준사전의 구성요소에 대한 정의는 다음과 같다.

표 10 연계표준사전 구성요소

구성요소	설명
연계표준단어	<ul style="list-style-type: none"> 표준단어명, 표준단어영문명, 표준단어영문약어명, 표준단어정의 등으로 구성 업무용어에 기반한 원자단위의 명사형 단어
연계표준도메인	<ul style="list-style-type: none"> 표준도메인대분류, 표준도메인중분류, 도메인한글명, 도메인영문명, 데이터타입, 데이터길이 등으로 구성
연계표준용어	<ul style="list-style-type: none"> “연계표준단어”와 “연계표준도메인”의 조합으로 생성 연계표준단어의 조합으로 용어명이 생성되고, 연계표준도메인에서 정의한 데이터 형식 반영 용어명, 용어영문약어명, 도메인한글명 등으로 구성

1.1.3 공통표준사전과 연계표준사전 관계

연계표준사전은 [공공기관의 데이터 표준화 지침]을 준용하고, [공공기관의 데이터 표준화 지침]의 공통표준용어(단어, 도메인)를 포함하며, 연계표준사전에서 정의되지 않고 플랫폼 고유의 독자적인 표준사전(용어, 단어, 도메인)은 개별표준으로 정의·관리 한다. 플랫폼 참여기관은 개별표준으로 관리되는 표준용어 중 타 플랫폼과 연계가 필요하다고 판단되는 항목은 연계표준사전에 등록하여 융·복합서비스에 활용될 수 있도록 한다.

빅 데이터 구축 시 적용되는 표준 유형은 공통표준용어, 연계표준사전, 플랫폼 개별사전으로 구분하며 표준 간의 관계 및 내용은 다음과 같다.

그림 11 표준간의 관계 및 내용

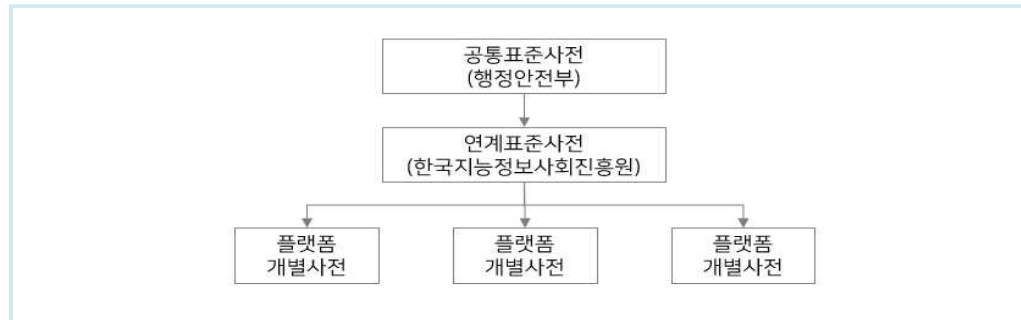


표 11 표준화 유형별 대상범위

표준유형	내용
공통표준사전	공공데이터베이스 표준화 지침의 공통표준단어, 공통표준용어, 공통표준도메인
연계표준사전	공통표준사전을 준수하며, 빅데이터 플랫폼 및 센터에서 일관되게 사용하는 연계표준용어, 연계표준도메인 및 연계표준단어
플랫폼 개별사전	연계표준사전에 기반하며, 각 플랫폼, 센터에서 해당 업무를 수행하기 위해서 정의한 용어, 도메인 및 단어 사전

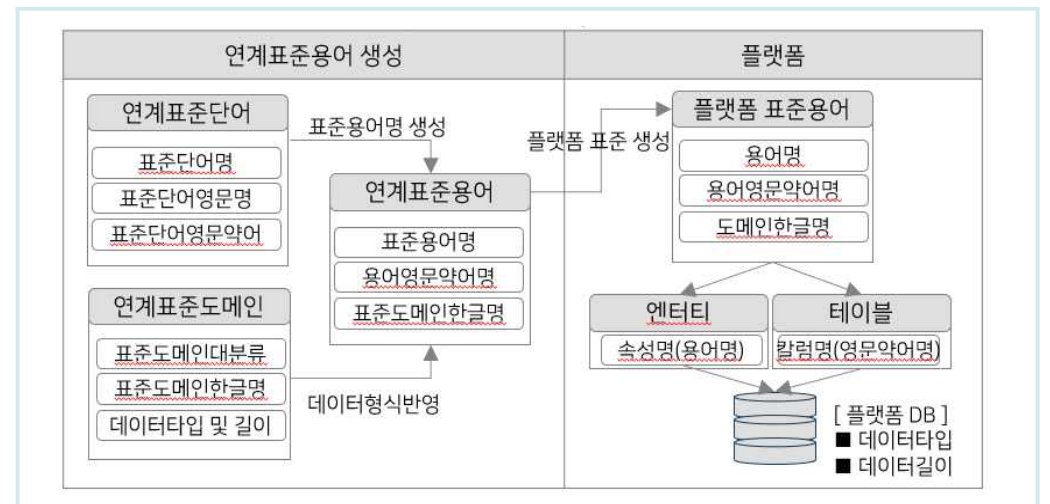
02

연계표준사전 표준화



연계표준사전은 연계표준단어, 연계표준도메인, 연계표준용어로 구성되며, 연계표준사전 구성요소 별로 표준을 정의하고 있으며 구성요소의 주요 관리항목 간의 관계는 다음과 같다.

그림 12 연계표준사전 구성요소 관계



2.1 연계표준용어

연계표준용어는 빅데이터 플랫폼 및 센터 내에서 사용하는 용어 중 공통적으로 사용하거나, 플랫폼 및 센터별로 상이하게 사용하는 동일한 의미의 유사한 컬럼들에 대해 공통의 표준용어를 적용하여 컬럼을 용이하게 식별하고, 여러 플랫폼에 걸친 데이터에 대해 통계산출 및 융·복합서비스를 위해 일관된 기준을 제공한다.

또한, 연계표준용어 표준화란 빅데이터 플랫폼 및 센터 내의 정보시스템에서 사용하는 개별용어의 적용 및 관리원칙을 정하는 것을 의미한다. 연계표준용어는 공공기관의 데이터베이스 표준화 지침을

준수한다.

연계표준용어의 관리항목은 필수요소인 용어명, 용어영문약어명, 도메인한글명 등이 있으며, 관리를 위해 도메인유형, 데이터타입, 데이터길이, 도메인영문명 등으로 구성된다.

표 12 연계표준용어 관리항목 및 예시

용어명	용어구성	용어영문약어명	용어영문정식명	데이터타입	데이터길이	도메인유형	도메인한글명	도메인영문명	등록일
시작일자	시작_일자	BGN_DT	BEGIN_DATE	VARCHAR2	10	시간	일자_V10	DT_V10	2021-01-01
종료일자	종료_일자	END_DT	END_DATE	VARCHAR2	10	시간	일자_V10	DT_V10	2021-01-01
연령대코드	연령대_코드	AGRDE_CD	AGE GRADE_C ODE	VARCHAR2	3	코드	연령대코드_V3	AGRDE_CD_V3	2021-01-01

용어명은 단어명의 조합으로 구성되고, 용어명의 마지막 단어는 형식단어로 구성된다. 용어명은 [공공기관의 데이터베이스 표준화 지침]의 공통표준용어를 포함하고, 빅데이터 플랫폼 및 센터 내에서 공통적으로 사용하거나, 유사한 컬럼들에 대해 공통의 표준용어를 적용하여 컬럼을 용이하게 식별할 수 있는 용어의 집합을 의미한다.

동음이의어는 가급적 연계표준용어에서는 정의하지 않고 플랫폼에서 필요시에 정의하여 사용한다.

용어명 정의 기준은 다음과 같다

표 13 용어명 정의 기준

표준유형	내용	예시
용어구성	<ul style="list-style-type: none"> 용어는 공공기관의 데이터베이스 표준화 지침의 공통표준용어와 플랫폼 및 센터에서 공통적으로 사용하거나, 유사한 컬럼으로 구성된다. 	<ul style="list-style-type: none"> 공통표준용어 : 가산_금액->가산금액 플랫폼 개별용어 : 시작_일자->시작일자
용어 글자 수	<ul style="list-style-type: none"> 용어는 가급적 6개 단어 이하로 조합하여 구성한다. 	<ul style="list-style-type: none"> 개인정보+수집+동의+여부->개인정보 수집동의여부
용어생성 급칙 문자	<ul style="list-style-type: none"> 용어생성 시 띄워쓰기, 특수문자는 사용하지 않는다. 용어는 한글사용을 권장하며, 불가피한 경우 숫자와 영문자 사용을 허용한다. 영문자를 사용할 경우에는 대문자를 사용한다. 	<ul style="list-style-type: none"> 거래 일자(X)->거래일자(O) 1등급기준점수(O) html사용여부(X)-> HTML사용여부(O)
용어의미	<ul style="list-style-type: none"> 하나의 용어가 다수의 의미를 갖지 않도록 사용함을 원칙으로 함 	<ul style="list-style-type: none"> 시작및종료일자(X) ->시작일자(O), 종료일자(O)

표준유형	내용	예시
숫자가 포함된 경우	<ul style="list-style-type: none"> 숫자의 사용은 가능한 용어의 직관적 의미 파악이 가능하도록 사용을 최소화한다. 숫자 자체는 단일어와 조합하여 등록하여 사용한다. 의미없는 일련번호를 부여하기 위한 숫자는 사용하지 않는다. 	<ul style="list-style-type: none"> 1등급기준일자(O) 전화번호1(X) -> 회사전화번호(O)
동음이의어 및 이음동어	<ul style="list-style-type: none"> 동음이의어는 가급적 사용하지 않으나, 플랫폼의 상황에 따라 플랫폼의 개별 용어사전으로 정의하여 사용할 수 있다. 	<ul style="list-style-type: none">

용어구성은 용어명을 구성하는 단어를 [단어+”_”+단어]의 형식으로 정의한 속성을 말하며, 용어구성의 개별단어는 표준단어사전에 정의되어 있어야 한다.

용어영문약어명은 용어명의 영문약어명을 의미하며, 용어명을 구성하는 개별단어에 대하여 단어사전에 등록되어 있는 영문약어명으로 구성한다. 용어영문약어명은 [단어영문약어명 + “_” + 단어영문약어명] 형식으로 구성한다.

용어영문정식명은 용어명을 구성하는 개별단어에 대하여 단어사전에 등록되어 있는 영문정식명으로 구성한다. 용어영문정식명은 [단어영문정식명 + “_” + 단어영문정식명] 형식으로 구성한다.

데이터타입은 개별 용어명의 도메인한글명에 해당하는 데이터 타입을 의미한다. 데이터타입은 일반적인 데이터타입을 사용하며, 개별 플랫폼의 DBMS 등 시스템 형식에 맞게 변환하여 적용해야 한다. 표준화 가이드에서 적용한 데이터타입은 문자를 위해 CHAR, VARCHAR2, 숫자를 위해서는 NUMERIC, DOUBLE, FLOAT, 대용량의 데이터 저장을 위해서는 CLOB와 BLOB 타입이 있다.

데이터길이는 개별 용어명의 해당 도메인한글명에 해당하는 데이터 길이르 의미한다. 데이터길이 표현은 [전체길이·소수점이하길이]이며, 전체길이는 소수점이하 길이를 포함한 전체길이를 표현한다.

도메인한글명은 개별 용어명의 데이터 저장형식을 표현하는 인포타입을 의미한다. 도메인한글명 표현은 [도메인분류명+”_”+데이터타입약어명+데이터길이]형식으로 정의한다.

도메인영문명은 개별 용어명의 도메인한글명에 해당하는 영문명을 의미한다. 도메인영문명 표현은 [도메인분류명의 영문약어명+”_”+데이터타입약어명+데이터길이]형식으로 정의한다.

연계표준용어에 대한 정의서는 다음과 같다.

표 14 연계표준용어 정의서

표준유형	내용	비고
용어명	<ul style="list-style-type: none"> 모든 플랫폼에서 공통적으로 사용하는 표준용어의 한글명 공공데이터 표준화 지침의 공통표준용어명 포함 용어구성의 “_”를 제외한 용어를 기재 한글을 원칙으로 하며 특수문자와 공백은 사용 불가 관용적으로 널리 알려진 영문약어에 한하여 대문자로 사용 	필수

표준유형	내용	비고
용어구성	<ul style="list-style-type: none"> 용어명을 “업무단어”+ “_”+ “형식단어”로 구성된 한글 명칭 업무단어 및 형식단어는 단어사전에 등재 	
용어영문약어명	<ul style="list-style-type: none"> 용어구성의 단어에 대하여 단어사전에 등록 되어 있는 영문약어명을 기재 “영문약어명” + “_” + “영문약어명” 형식으로 기재 대문자로 작성하며 특수문자는 사용하지 않음 	필수
용어영문정식명	<ul style="list-style-type: none"> 용어구성의 단어에 대하여 단어사전에 등록되어 있는 영문명을 기재 “단어영문명” + “_” + “단어영문명” 형식으로 대문자 기재 	
데이터타입	<ul style="list-style-type: none"> 해당 용어명이 가지는 데이터 타입을 기재 예) BLOB, CLOB, CHAR, VARCHAR2, NUMBER, DOUBLE, FLOAT, DATE, DATETIME 	
데이터길이	<ul style="list-style-type: none"> 해당 용어명이 가지는 데이터 길이를 기재 [전체길이.소수점이하길이]이며, 전체길이는 소수점이하 길이를 포함한 전체길이를 표현 	
도메인유형	<ul style="list-style-type: none"> 용어명의 형식단어에 해당하는 도메인사전의 도메인.중분류 명칭을 기재 	
도메인한글명	<ul style="list-style-type: none"> 용어명의 데이터 저장형식을 표현하는 인포타입을 기재 [도메인유형+ “_”+ 데이터타입약어명+ 데이터길이] 도메인사전의 도메인.한글명을 기재 	필수
도메인영문명	<ul style="list-style-type: none"> 도메인한글명에 해당하는 영문명을 기재 도메인사전의 도메인.한글명을 기재 	
등록일	<ul style="list-style-type: none"> 용어명 등록일자 	필수

2.2 연계표준도메인

연계표준도메인은 빅데이터 플랫폼 및 센터에서 공통으로 사용하는 연계표준용어의 데이터 속성을 의미하며, 도메인 분류체계를 수립하고 도메인별 데이터 타입, 데이터 길이 등 구성요소를 정의한다.

연계표준도메인은 공공데이터베이스 표준화 관리원칙을 준수하며, 관리항목으로는 필수 항목인 도메인·대분류, 도메인·중분류, 도메인·한글명, 데이터타입, 데이터길이 등이 있으며, 관리를 위해 도메인·영문명, 도메인설명 등으로 구성 된다.

표 15 연계표준도메인 관리항목 및 예시

도메인·대분류	도메인·중분류	도메인·한글명	도메인·영문명	데이터 타입	데이터 길이	코드체계 및 데이터포맷	정의설명
시간	일자	일자_V10	DT_V10	VARCHAR2	10	-	-

도메인·대분류	도메인·중분류	도메인·한글명	도메인·영문명	데이터 타입	데이터 길이	코드체계 및 데이터포맷	정의설명
코드	성별코드	성별코드_V1	SEX_CD_V1	VARCHAR2	1	-	-
코드	연령대코드	연령대코드_V3	AGRDE_CD_V3	VARCHAR2	3	-	-

도메인 대분류는 연계표준도메인의 상위 구분을 정의하며, 도메인 분류체계의 최상위로 해당 도메인의 특성을 잘 나타내고 사용자가 도메인을 쉽게 인식·분류할 수 있도록 명칭을 정한다. 업무에 따라 도메인 대분류가 달라지지 않도록 데이터가 가질 수 있는 데이터 형식을 기준으로 정의한다.

표 16 도메인 대분류

도메인 대분류	설명
코드	<ul style="list-style-type: none"> 데이터를 분류, 구분하거나 긍정·부정, 유·무로 구분할 수 있는 속성을 정의하기 위한 도메인 기존의 구분·여부 포함
수	<ul style="list-style-type: none"> 객체의 개수나 양을 수로써 표현하기 위한 도메인 일반적인 측량 단위 및 기존의 올·무게·길이·면적 포함
시간	<ul style="list-style-type: none"> 특정사건이 일어난 시점, 시점과 시점 간의 정해진 기간을 표현하기 위한 도메인 기존의 기간 포함
번호	<ul style="list-style-type: none"> 일정한 체계를 가지거나 특정 자리의 의미가 존재하는 속성을 정의할 때 사용하는 도메인 숫자·문자 등을 포함한 식별번호의 의미로 사용
금액	<ul style="list-style-type: none"> 화폐의 가치를 수로 표현하기 위한 도메인 기존의 값 포함
명	<ul style="list-style-type: none"> 문자 형식으로 객체에 대한 식별을 표현하기 위한 도메인 아이디는 객체를 식별하고, 주소는 거리명, 행정구역명으로 나타냄으로 명 그룹에 해당 기존의 ID·경로·지명·주소·연락처·위치 포함
내용	<ul style="list-style-type: none"> 말·글·그림 따위의 모든 매체 속에 들어 있는 것을 표현한 그룹 기존의 제목·SQL·결과·데이터·비고·설명·인덱스 포함

도메인 중분류는 연계표준도메인의 세부구분을 정의한 것으로, 도메인 대분류에서 정한 상위 기준의 포괄적인 명칭이 아닌 데이터타입, 데이터길이 등을 정의하는 구체적인 명칭을 사용한다. 도메인 중분류명은 데이터 속성으로 활용 가능한 상세화 수준 까지를 분류하여 제공한다. 예시) 시간도메인은 연월일시분초로 모두 표현할 수 있으나 연도·연월·연월일 등의 상세화된 분류가 데이터베이스에서 활용도가 높아 세분화하여 제공한다.

도메인한글명은 “도메인.중분류+ “_”+ “데이터타입의약어”+ “데이터길이” 로 구성한다. 데이터타입 및 데이터길이의 표현형식은 연계표준도메인의 데이터타입 및 데이터길이 속성에서

정한다.

도메인·영문명은 알파벳 대문자 및 _와 .(점) 및 숫자로만 구성하며, 괄호 등 특수문자는 사용할 수 없다. 도메인영문명은 “도메인·중분류명+”_”+“데이터타입약어”+“데이터길이”로 구성한다. 도메인·중분류의 영문약어가 연계표준단어에 미등록시에는 도메인·대분류의 영문약자를 대신 사용한다.

연계표준도메인이 나타내는 문자, 숫자, 날짜 등의 데이터 형식을 식별하는 데이터타입과 데이터가 가질 수 있는 값의 범위인 데이터길이를 정의한다. 사용하는 데이터타입은 [연계표준도메인 정의서]에 정의하며, 빅 플랫폼 및 센터에서는 사용하는 데이터베이스의 환경에 맞게 변환하여 사용 가능하나, 제공하는 연계표준도메인은 변경하여 사용할 수 없고 별도로 매핑기능이나 독자적인 사전을 이용하여야 한다. 데이터길이 중 소수점이하를 사용하는 표시하는 경우 “전체데이터길이”+“(점)”+“소수점이하자리수” 형식으로 정의한다.

표 17 사람영역에서 활용을 권장하는 표준 연계키 항목

데이터타입	도메인명의 길이 표현	내 용	비고
BLOB	BL	최대4GB 이진파일	TEXT
CLOB	CL	최대4GB 텍스트파일	
CHAR	C+길이	최대2KB 고정길이 문자	
VARCHAR2	V+길이	최대4KB 가변길이 문자	N V A R C H A R 2 , VARCHAR, STRING
NUMBER	N+전체길이+“.”+소수점길이	숫자 1*10~130~1*10+126	DECIMAL, INT, INTEGER, NUMERIC
DOUBLE	D+전체길이+“.”+소수점길이	숫자 -1.7*10308~1.7*10308	REAL
FLOAT	F+전체길이+“.”+소수점길이	숫자 -3.4*1038~+3.4*1038	
DATE	DT	년월일시분초	

연계표준도메인의 정의서는 다음과 같다.

표 18 연계표준도메인 정의서

표준유형	내용	비고
도메인 대분류	• 도메인의 가장 큰 분류를 기재(금액·내용·명·번호·수·시간·코드)	필수
도메인 중분류	• 도메인으로 관리할 필요가 있는 단어 또는 단어들의 조합을 기재하며, 도메인 한글명에서 사용	필수
도메인 한글명	• 도메인중분류명 + “_” + 데이터타입 + 데이터길이	필수
도메인 영문명	• 도메인 한글명의 영문약어를 기재	
데이터 타입	• 해당 용어명이 가지는 데이터 타입을 기재	필수

표준유형	내용	비고
데이터 길이	<ul style="list-style-type: none"> 도메인 값을 표현하기 위한 데이터의 최대 길이를 지정 숫자로 기재 날짜, 시각처럼 길이가 정해져 있거나, 최대 길이를 알 수 없거나, 최대 필수 길이가 의미가 없는 경우 생략 가능 	
코드체계 및 데이터 포맷	<ul style="list-style-type: none"> 코드의 경우 코드값 목록, 코드가 아닌 경우 데이터 포맷 또는 비고 사항을 기재 	
정의 설명	<ul style="list-style-type: none"> 도메인에 대한 정의 및 설명을 기재 	

2.3 연계표준단어

연계표준단어는 빅데이터 플랫폼 및 센터에서 동일한 의미의 단어들 가운데 공통으로 사용되는 표준단어를 선정하고 관리항목을 정의한 것이다. 빅데이터 플랫폼 및 센터 차원의 일관성 있는 데이터 생성 및 활용기반을 마련하기 위해 연계표준단어를 정의하였으며, 공공데이터베이스 표준화 관리원칙을 준수한다.

연계표준단어는 필수요소인 단어명, 단어영문명, 영문약어명 등이 있으며, 관리를 위한 단어 정의 등으로 구성된다.

표 19 연계표준도메인 관리항목 및 예시

구분	단어명	단어영문명	영문약어명	정의	등록일
표준단어	가계	FAMILY LINE	FMLN	家系 대대로 이어온 한집안의 계통 집안살림을 꾸려나가는 수지의상태	기존
표준단어	가구	HOUSEHOLD	HSULD	家口 현실적으로 주거 및 생계를 같이하는 사람의 집단.	기존
표준단어	가로	WIDTH	WDTH	왼쪽에서 오른쪽으로 나 있는 방향. 또는 그 길이	기존

연계표준단어의 구분은 표준단어, 분류단어(형식단어)로 구분하여 관리한다. 연계표준단어는 연계표준용어의 용어명을 분할 후 선정된 단어와 [공공기관의 데이터베이스 표준화 지침]의 공통표준단어로 구성된다. 분류단어는 연계표준단어 중에서 데이터값, 데이터타입, 데이터길이 등 특정한 데이터 형식 식별이 가능한 단어로 연계표준용어의 용어명 구성시 일반적으로 가장 뒤에 위치하여야 하며 연계표준도메인의 한글도메인의 명칭으로 사용되는 단어(도메인 대분류, 도메인 중분류)를 의미한다.

연계표준단어의 단어명은 빅데이터 플랫폼에서 적용 중인 단어 중에서 연계표준용어의 용어명을 분할한 개별단어와 [공공기관의 데이터베이스 표준화 지침]의 단어로 구성된다.

▶ 단어 정의 시 주요 기준

- 단어는 한글, 영문, 숫자로 구성하며 영문자는 대문자를 사용한다. 예) html→HTML
- 한글문자 수는 제한을 두지 않으나, 한글자의 단어는 사용을 최소화하고 구체적인 단어로 정의한다. 예) 집→주택/주택
- 단어는 띄어쓰기를 하지 않는다. 예) 주민 등록 번호→주민등록번호
- 특수문자(_ - + [] () 등)는 사용하지 않는다.
- 동음이의어는 사용하지 않는다. 예) 다리(leg)→다리, 다리(bridge)→교량
- 이음동어는 사용하지 않는 것을 원칙으로 하나 필요시는 사용 가능. 예) 연도/년도→연도(표준단어)/년도(이음동어)

연계표준단어의 단어영문명은 단어명을 영문으로 표기한 단어영문명(Full Name)을 의미하며 단어명의 정의에 부합하는 단어영문명을 정의한다. 단어영문명은 영문약어명을 만드는 기준이 된다. 빅데이터 플랫폼 및 센터는 단어영문명을 생성시 분야별 전문용어사전(정보통신용어사전, 한국산업표준용어, 법령표준용어 등)을 참조하며 이외에 출판사 영어사전 등을 활용한다.

연계표준단어의 영문약어명은 단어영문명에 대한 약어로 연계표준용어에 대한 영문명과 일치하며, 컬럼정의서의 컬럼명과 일치해야한다. 빅데이터 플랫폼 및 센터는 영문약어명 정의시 [공공기관의 데이터베이스 표준화 지침]의 영문약어명 정의 기준을 준수하여 작성한다.

▶ 영문약어명 정의 시 주요 기준

- 국내외 표준으로 사용하는 영문명의 경우 해당 약어를 사용한다. 예) 수요일→WED
- 영문명에서 어두 이외의 모음을 제거하고 각 음절의 대표자음을 사용하여 구성하며, 두개의 자음이 연속할 경우 한 개의 자음은 생략한다.
- 숫자가 필요한 경우 영문 뒤에 위치하며, 영문명이 4자 이하인 경우 그대로 사용

연계표준단어의 정의는 연계표준단어의 단어 이해도를 높이기 위해 작성하며 영문약어명 선정 시 참조한다. 빅데이터 플랫폼 및 센터는 정의 작성시 표준국어대사전, 기술·산업표준 등의 정의를 활용한다.

연계표준단어 정의서는 다음과 같다.

표 20 연계표준단어 정의서

관리항목	항목 내용	비고
구분	<ul style="list-style-type: none"> 표준단어, 분류단어(형식단어)로 구분하여 기재 표준단어는 표준용어사전의 용어명을 분할한 후 선정된 단어와 [공공기관의 데이터베이스 표준화 지침]의 공통표준단어를 기재 분류단어는 표준단어 중에서 데이터값의 데이터타입, 데이터길이 등 특정한 데이터 형식 식별이 가능한 단어로 표준용어사전의 용어명 구성 시 일반적으로 가장 뒤에 위치하여야 하며 표준도메인사전의 한글도메인의 명칭으로 사용되는 단어(도메인 대분류, 도메인 중분류)를 의미 	

관리항목	항목 내용	비고
단어명	<ul style="list-style-type: none"> 빅데이터 플랫폼 및 센터에서 적용 중인 단어 중에서 표준용어사전의 용어명을 분할한 개별단어와 [공공기관의 데이터베이스 표준화 지침]의 단어를 기재 단어는 한글, 영문, 숫자로 구성하며 영문자는 대문자를 기재. 예) html -> HTML 단어는 띄어쓰기를 하지 않는다. 예) 주민 등록 번호→주민등록번호 특수문자(_ - + [] () 등)는 사용하지 않는다. 이음동어는 표준에서 정의하지 않으며, 필요시 이음동어와 동음이의어를 플랫폼에서 개별표준으로 정의하여 기재 	
단어영문명	<ul style="list-style-type: none"> 단어명을 영문으로 표기한 단어영문명(Full Name)을 의미하며 단어명의 정의에 부합하는 단어영문명을 기재 단어영문명 생성시 분야별 전문용어사전(정보통신용어사전, 한국산업표준용어, 법령표준용어 등)을 참조하며 이외에 출판사 영어사전 등을 활용하여 기재 	
영문약어명	<ul style="list-style-type: none"> 단어영문명에 대한 약어로 연계표준용어의 영문명과 일치하며, 컬럼정의서의 컬럼명과 일치 국내외 표준으로 사용하는 영문명의 경우 해당 약어 기재. 예) 수요일→WED 단어영문명에서 어두 이외의 모음을 제거하고 각 음절의 대표자음을 사용하여 구성하며, 두개의 자음이 연속할 경우 한 개의 자음은 생략하여 기재 숫자가 필요한 경우 영문 뒤에 위치하며, 영문명이 4자 이하인 경우 그대로 기재 	
정의	<ul style="list-style-type: none"> 단어의 이해도를 높여 영문약어명 선정 시 참조 가능하도록 기재 빅데이터 플랫폼 및 센터는 정의 작성 시 표준국어대사전, 기술·산업표준 등의 정의를 활용 	
등록일	단어명 등록시점을 기재	

03

연계표준사전 관리체계



3.1 연계표준사전 관리조직

연계표준사전 관리조직은 행정안전부, 한국지능정보사회진흥원, 플랫폼 참여기관으로 구성된다. 행정안전부는 범정부 차원의 공공데이터베이스 표준화 추진을 총괄 수행하고, 공통표준용어의 제·개정 업무를 관리한다. 한국지능정보사회진흥원은 연계표준사전 정책을 총괄하며 연계표준사전의 추가·수정·폐기 등 업무를 수행한다.

연계표준사전 표준화 사안 협의·조정 등을 위해 연계표준사전 표준화 협의회를 운영하며, 플랫폼 참여기관은 최종 확정된 연계표준사전을 플랫폼 및 센터 내에 적용하고, 전 플랫폼에서 공동 활용할 수 있는 연계표준용어를 신규 발굴하여 연계표준사전에 등록을 신청한다.

연계표준사전의 관리조직과 세부 역할은 다음과 같다.

그림 13 연계표준사전 관리조직

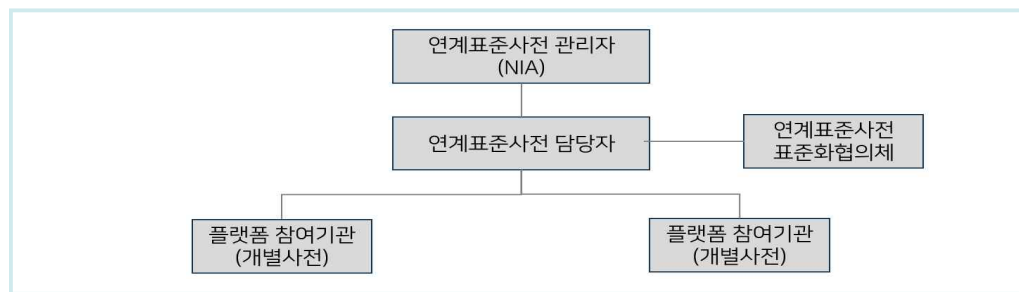


표 21 연계표준사전 조직별 역할

관리항목	항목 내용
연계표준사전 관리자(NIA)	<ul style="list-style-type: none"> 연계표준사전 최종 승인 및 확정
연계표준사전 담당자	<ul style="list-style-type: none"> 연계표준사전 연구개발 및 관리 등 업무 수행 연계표준사전 발굴 및 표준화, 플랫폼 개별사전 관리기준 수립
연계표준사전 표준화협의체	<ul style="list-style-type: none"> 특정 플랫폼의 표준화를 주도적으로 추진하고, 플랫폼 간 표준화 이슈를 협의 조정
플랫폼 참여기관	<ul style="list-style-type: none"> 최종 확정된 연계표준사전을 플랫폼 개별사전에 등록 및 관리하고 플랫폼의 정보시스템에 적용 전 플랫폼 공동활용 대상 용어 신규 발굴 및 연계표준사전으로 등록 요청 [플랫폼 담당자] 확정된 연계표준사전을 플랫폼 내 정보시스템에 적용 확정된 연계표준사전을 플랫폼 개별사전에 반영 및 센터에 전달 및 교육, 관리 전 플랫폼 공동활용이 가능한 연계표준사전 발굴 및 등록 신청 [센터담당자] 소관 정보시스템에 표준사전 적용 및 관리

3.2 연계표준사전 관리원칙

연계표준사전은 다음 원칙에 따라 관리한다.

- ▶ **일관성** : 연계표준사전(용어, 단어, 도메인)은 각 플랫폼에서 구축하여 운영 중인 데이터베이스에 일관되게 적용한다.
- ▶ **체계성** : 연계표준용어는 표준으로 정한 단어와 도메인의 조합으로 이루어지며, 이를 체계적으로 관리하기 위해 연계표준단어와 연계표준도메인을 정의한다.
- ▶ **범용성** : 모든 플랫폼은 [표준화 가이드] 및 [공공기관의 데이터베이스 표준화 지침] 을 참조하여 소관 데이터에 일관된 적용이 가능하도록 자체적인 데이터 표준 검토 체계를 마련한다.
- ▶ **표준성** : 모든 플랫폼은 플랫폼의 개별사전(용어, 단어, 도메인)을 정의할 경우, [한국지능정보사회진흥원]에서 제공하는 연계표준사전을 우선적으로 적용한다.
- ▶ **효율성** : 모든 플랫폼은 메타시스템을 활용하여 플랫폼 개별사전(용어, 단어, 도메인)을 구현하고, 연계표준사전을 등록 관리해야 하며, 공동활용이 필요한 경우 연계표준사전(용어, 단어, 도메인)으로 등록 할 수 있도록 상호 협력해야 한다.

3.3 연계표준사전 신규등록

빅 데이터 플랫폼 간 공동 활용이 가능한 플랫폼 개별용어를 연계표준용어로 신규 등록하는 것을 의미하며, 플랫폼 참여기관의 신청과 한국지능정보사회진흥원의 자체 발굴에 의하여 신규로 등록된다.

플랫폼 참여기관이 플랫폼 개별용어로 정의된 용어 중에서 플랫폼의 공동 활용이 필요하다고 판단되면 연계표준용어로 등록 신청할 수 있다. 또한, 연계표준용어를 운영하는 한국지능정보사회진흥원은 플랫폼의 개별용어, 메타데이터 등을 조사·분석하여 연계표준용어 등록 대상으로 발굴할 수 있다.

플랫폼의 개별용어를 연계표준용어에 신규 등록 시 연계표준단어, 연계표준도메인의 신규 등록이 필요한지 검토하여, 필요한 경우 같이 등록해야한다.

연계표준사전 신규 등록 절차 및 세부 절차는 다음과 같다.

그림 14 연계표준사전 신규 등록 절차

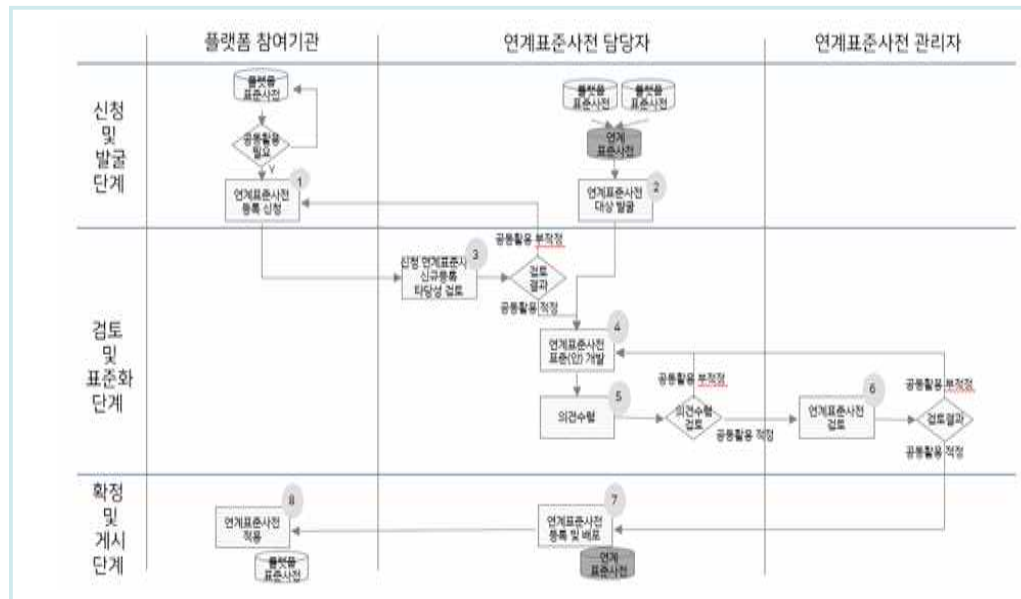


표 22 연계표준사전 신규등록 세부절차

구분	절차	절차 내용	수행주체
신청 및 발굴 단계	1. 연계표준사전 등록 신청	<ul style="list-style-type: none"> 플랫폼 참여기관은 전 플랫폼 공동 활용이 필요하다고 판단되는 연계표준사전(용어, 단어, 도메인)을 선정하여 연계표준사전에 등록을 신청 해당 신청 용어와 관련하여 연계표준단어 및 연계표준도메인의 신규 생성이 필요한 경우 같이 신청 	플랫폼 참여기관
	2. 연계표준사전 대상 발굴	<ul style="list-style-type: none"> 연계표준사전 담당자는 플랫폼 표준용어를 정기적으로 수집·분석하여 연계표준용어 추가 등록 대상을 발굴 여러 플랫폼에서 사용하거나, 공동활용이 필요하다고 판단하는 용어 발굴 	연계표준사전 담당자
검토 및 표준화 단계	3. 신청 연계표준사전 신규등록 타당성 검토	<ul style="list-style-type: none"> 연계표준사전 담당자는 연계표준사전 등록 신청내역을 검토하여 타당성 분석 실시 기존 연계표준사전(용어, 단어, 도메인)와의 중복 여부 검토 연계표준용어 신청시 용어를 구성하는 연계표준단어, 연계표준도메인 존재 유무 검토(미 등록시 신규 등록 필요 검토) 	연계표준사전 담당자
	4. 연계표준사전 표준(안) 개발	<ul style="list-style-type: none"> 연계표준사전 담당자는 플랫폼에서 신청한 연계표준용어 및 자체 발굴한 연계표준용어를 대상으로 연계표준사전 표준(안)을 개발 표준(안) 내용 <ul style="list-style-type: none"> 연계표준용어를 의미를 갖는 최소단위인 연계표준단어 수준으로 분할 기존 연계표준용어에 기 등록 여부확인 기존 연계표준단어, 연계표준도메인에 미 등록시 신규 정의 	연계표준사전 담당자
	5. 의견수렴	연계표준사전 담당자는 필요시, 외부전문가 및 표준화협의체, 플랫폼에 의견수렴하고 연계표준사전 관리자에게 표준(안)의 확정을 요청	연계표준사전 담당자
	6. 연계표준사전 검토	연계표준사전 관리자는 신규 연계표준사전 표준(안)을 검토 한 후 연계표준용어로 최종 확정	연계표준사전 관리자
확정 및 게시 단계	7. 연계표준사전 등록 및 배포	연계표준사전 담당자는 확정된 연계표준사전(용어, 단어, 도메인 포함)을 전 플랫폼에서 적용할 수 있도록 배포 및 관리	연계표준사전 담당자
	8. 연계표준사전 적용	모든 플랫폼은 확정된 연계표준사전(용어, 단어, 도메인 포함)을 적용하고, 관련 센터도 적용할 수 있도록 배포 및 관리	플랫폼 및 센터 담당자

3.4 연계표준사전 변경절차

연계표준사전 변경절차는 사용 중인 연계표준사전(용어, 단어, 도메인)의 관리항목 중 데이터베이스에 직접적 영향이 없는 항목이 변경되는 것을 대상으로 한다. 데이터베이스에 직접적인 영향이 있는 용어명, 용어영문약어명 등은 표준사전(용어, 단어, 도메인) 자체를 “폐기” 후 “신규등록” 하여야 한다.

데이터베이스에 직접적 영향이 없는 항목은 단어정의, 도메인정의설명, 등록일 등이다. 데이터베이스에 직접적 영향이 있는 항목은 용어명, 용어영문약어명, 도메인.중분류, 도메인한글명, 데이터타입, 데이터길이 등이다.

연계표준사전의 변경절차 및 세부절차는 다음과 같다.

그림 15 연계표준사전 변경절차

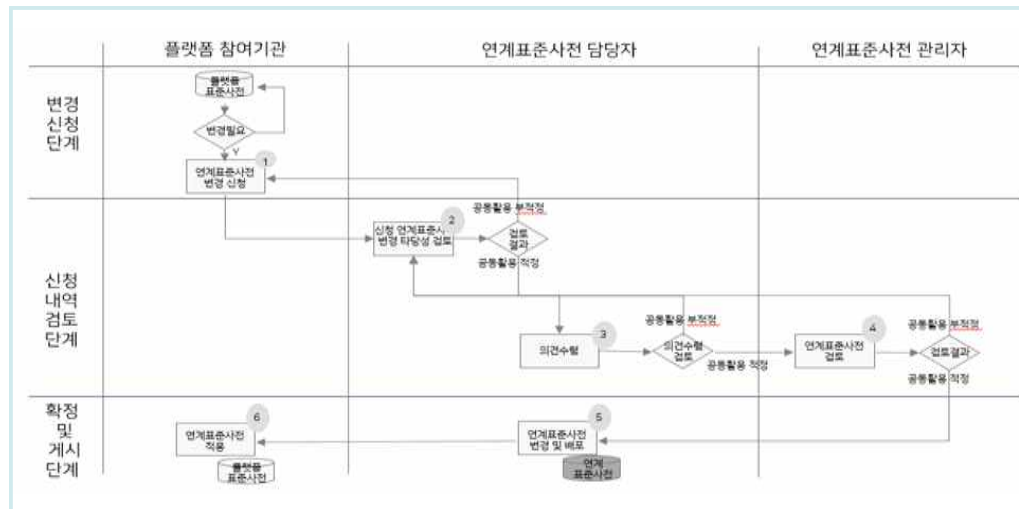


표 23 연계표준사전 변경 세부절차

구분	절차	절차 내용	수행주체
변경 신청 단계	1. 연계표준사전 변경 신청	<ul style="list-style-type: none"> 플랫폼 참여기관은 데이터베이스에 직접적인 영향이 없는 항목이 변경된 경우 해당 연계표준사전(용어, 단어, 도메인)에 대한 변경을 신청 단어정의, 도메인정의설명, 등록일 등 변경 시 신청 	플랫폼 참여기관
	2. 신청 연계표준사전 변경 타당성 검토	<ul style="list-style-type: none"> 연계표준사전 담당자는 연계표준사전 변경 신청내역을 검토하여 변경(안)을 정의 변경 기준인 단어정의, 도메인정의설명 등에 대하여 타당성 검토 	연계표준사전 담당자
신청내역 검토 단계	3. 의견수렴	<ul style="list-style-type: none"> 연계표준사전 담당자는 필요시, 외부전문가 및 표준화협의체, 플랫폼에 의견을 수렴하고 연계표준사전 관리자에게 표준(안)의 확정을 요청 	연계표준사전 담당자
	4. 연계표준사전 검토	<ul style="list-style-type: none"> 연계표준사전 관리자는 표준(안)을 검토하고, 그 결과를 반영하여 연계표준용어로 최종 확정 	연계표준사전 관리자
확정 및 게시 단계	5. 연계표준사전 변경 및 배포	<ul style="list-style-type: none"> 연계표준사전 담당자는 확정된 연계표준사전 변경내역을 반영하고, 전 플랫폼에서 반영할 수 있도록 배포 및 관리 	연계표준사전 담당자
	6. 연계표준사전 적용	<ul style="list-style-type: none"> 모든 플랫폼은 확정된 연계표준사전을 적용하고, 플랫폼 및 센터 관련 센터도 적용할 수 있도록 배포 및 관리 	플랫폼 및 센터 담당자

3.5 연계표준사전 폐기절차

연계표준사전(용어, 단어, 도메인)의 폐기는 기존에 사용하던 연계표준사전을 더 이상 사용하지 않는 것을 의미하며, 사용 중인 연계표준사전이 전 플랫폼의 데이터베이스에서 사용하지 않게 되는 경우와 연계표준사전 관리항목 중 데이터베이스에 직접적인 영향이 있는 항목이 변경되어 사용을 중지하고 변경 항목을 적용하는 경우에 해당한다.

관리항목의 변경이 발생한 용어 중 데이터베이스에 직접적인 영향이 있는 경우 “폐기” 후 “신규등록” 절차를 통해 재등록한다.

연계표준용어 폐기절차 및 세부절차는 다음과 같다.

그림 16 연계표준사전 폐기 절차

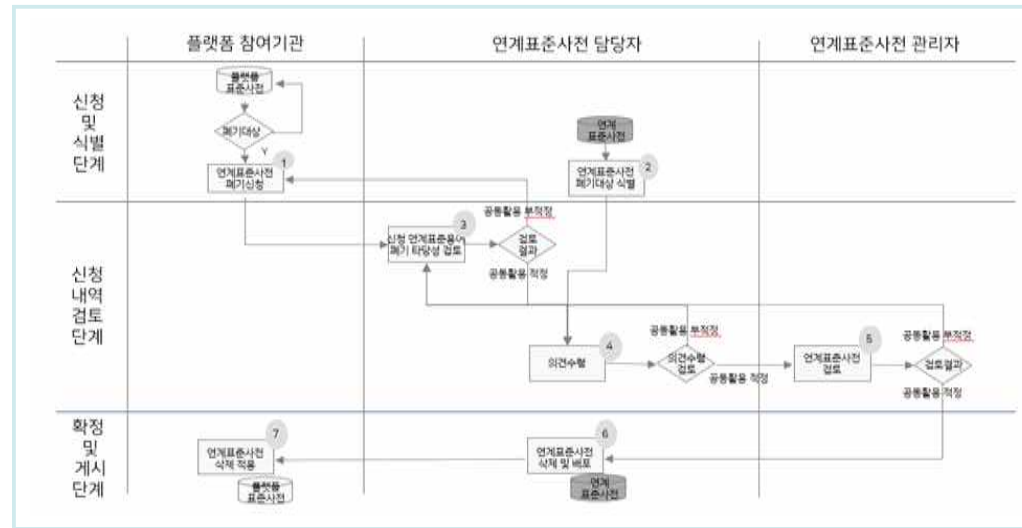


표 24 연계표준사전 폐기 세부절차

구분	절차	절차 내용	수행주체
신청 및 식별 단계	1. 연계표준사전 폐기 신청	<ul style="list-style-type: none"> 플랫폼 참여기관은 연계표준사전(용어, 단어, 도메인)의 관리항목 중 데이터베이스에 직접적인 영향이 있는 항목이 변경되는 경우 해당 연계표준사전에 대한 폐기를 신청 표준용어명, 용어영문약어명의 변경이 발생하는 경우 도메인한글명, 데이터타입, 데이터길이 등 신규생성 또는 변경 발생하는 경우 	플랫폼 참여기관
	2. 연계표준사전 폐기대상 발굴	<ul style="list-style-type: none"> 시스템 폐기 등으로 플랫폼에서 특정 연계표준사전(용어, 단어, 도메인)가 더 이상 사용되지 않는 경우 연계표준사전 담당자는 연계표준사전 폐기 대상을 식별 	연계표준사전 담당자
신청내역 검토 단계	3. 신청 연계표준사전 폐기 타당성 검토	<ul style="list-style-type: none"> 연계표준사전 담당자는 연계표준사전 폐기 신청내역을 검토하고 폐기(안)을 작성 연계표준용어의 표준용어명, 용어영문약어명 등 폐기 타당성 검토 	연계표준사전 담당자
	4. 의견수렴	<ul style="list-style-type: none"> 연계표준사전 담당자는 필요시, 외부전문가 및 표준화협의체, 플랫폼의 의견을 수렴하고 연계표준사전 관리자에게 표준(안)의 확정을 요청 	연계표준사전 담당자
	5. 연계표준사전 검토	<ul style="list-style-type: none"> 연계표준사전 관리자는 폐기 연계표준사전 표준(안)을 검토 한 후 폐기 대상으로 최종 확정 	표준용어사전 관리자

구분	절차	절차 내용	수행주체
확정 및 게시 단계	6. 연계표준사전 삭제 및 배포	<ul style="list-style-type: none"> 연계표준사전 담당자는 폐기된 연계표준사전을 전 플랫폼에서 적용할 수 있도록 배포 및 관리 	연계표준사전 담당자
	7. 연계표준사전 삭제 적용	<ul style="list-style-type: none"> 모든 플랫폼은 삭제된 연계표준사전을 적용하고, 플랫폼 및 센터 관련 센터도 적용할 수 있도록 배포 및 관리 	플랫폼 및 센터 담당자

01

개요



1.1 데이터 소유권 개요

- 데이터(Data)는 제4차 산업혁명 시대의 원유(Oil)이라고 하며, 데이터 기반 경제(Data Economy)의 효율성을 높이기 위하여 데이터에 대한 권리, 즉 데이터 소유권(Data Ownership)의 개념에 대한 논의가 대두
- 데이터 소유권(Ownership)은 데이터에 대한 권리를 개념적으로 지칭하는 것으로 현재 법적으로 명확한 정의나 개념이 제시되어 있는 용어는 아님
- 기존의 물산(부동산, 증권 등 재산)은 재산권에서 권리를 보호하여 왔으며, 정보, 저작물 등 지적 재산에 대해서는 저작권에서 세분화하여 권리를 보호하여왔음
- 2017년 유럽위원회(European Commission)의 'Building the European Data Economy' 프로젝트에서는 데이터 소유권을 제4차 산업혁명 시대의 새로운 권리로 보아 도입을 제안하였음¹⁾
- 국내에서도 2017년 후반부터 기존의 민법 및 저작권법에서 데이터 소유권의 인정과 보호가 불확실 한 것에 대해서 추가적인 입법이나 제도의 수립을 제안하여왔으며, 이에 수반되는 각종 활동이 전개되어 왔음
- 빅데이터 플랫폼과 데이터센터의 관점에서 데이터 소유권은 데이터의 생산·수집, 축적, 유통·거래의 전생애주기과정에 대해 이를 제도적으로 보호받고, 권리를 주장할 수 있는 권리설정의 개념으로 볼 수 있음
- 데이터의 생산·수집과정에서 데이터의 출처 및 이력(Provenance)에 의한 저작권의 성립여부와 빅데이터 플랫폼과 데이터센터의 역할 및 권리 설정의 필요성이 제기되고 있음
- 데이터의 유통·거래과정에서 빅데이터 플랫폼 및 데이터센터가 이용자들에게 데이터셋이나 분석데이터를 판매하고, 이용자가 이를 활용하거나 재가공하는 생태계가 조성되고 있는 것으로부터 이에 대한 권리주장과 법적 보호에 대한 필요성이 제기되고 있음
- 그러므로 빅데이터 플랫폼 및 데이터 센터의 관점에서 데이터의 생산·수집, 축적, 유통·거래의

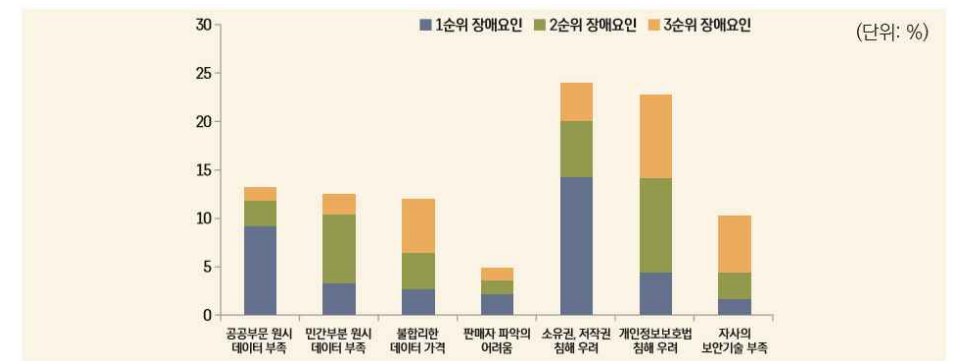
1) Francesco Bantre, Data Ownership in the Data Economy: A European Dilemma, EU Internet Law in the digital era, (edited volume based on the REDA 2017 conference), Springer, (2018) Forthcoming

전생애주기과정에 대한 플랫폼 및 데이터센터, 이용자 간의 권리를 현행 제도 상에서 정의하고 이를 실행하기 위한 이용약관의 개선이 데이터 소유권 및 이용약관 개선의 목적임

1.2 데이터 소유권 및 이용약관의 필요성

- 데이터의 생산·수집, 축적, 유통·거래의 전생애주기과정, 생태계 조성을 위해 '민·관 협력기반 데이터 플랫폼 발전전략' 등의 국가전략계획에 의해 빅데이터 플랫폼 및 데이터센터의 구축이 추진되고 있음
- 이러한 빅데이터 플랫폼 및 데이터센터 운영 등 생태계 조성에 있어 데이터의 공급자인 플랫폼 및 데이터 센터의 권리와 이용허락범위, 이용자의 권리와 의무 및 책임사항 등의 규정이 필요

그림 17 데이터 구매 경험기업의 데이터 구매 장애요인²⁾



- 또한, 데이터 구매한 경험이 있는 기업을 대상으로 장애요인을 조사한 사례를 볼 때, 데이터 소유권, 저작권이 데이터 유통·거래의 주요 장애요인으로 분석되고 있음
- 하지만, 현재 데이터 소유권이 민법이나 저작권법에서 인정되는 권리가 아니므로, 단기적인 데이터 유통·거래의 장애요인을 해소하기 위하여, 데이터 생태계의 공정거래 질서를 확립하고 법적 리스크를 사전예방하기 위한 데이터 소유권, 저작권의 기본체계를 보완하는 이용약관 등 기반이 필요
- 현재 빅데이터 플랫폼 및 데이터센터에서는 데이터의 생산·수집, 축적, 유통·거래의 전생애주기과정에 이에 참여하는 데이터의 공급자³⁾와 이용자⁴⁾에게 이용약관을 제시하여 이에 대한 동의와 의무 및 책임사항을 이행하도록 하고 있음
- 빅데이터 플랫폼 및 데이터센터는 개별적으로 이용약관을 개발하여 시행하고 있으며, 2020년에 개별 이용약관에 대한 점검과 개선을 수행하였으며, 그 결과 법률 검토 미비에 따른 법적 리스크에 노출된 사례가 있음

2) KISDI, ICT기반 신산업 발전을 위한 데이터 거래 활성화 방안 보고서, 2018

3) 데이터의 공급자는 데이터센터와 민간기업 등 데이터를 공급하는 객체

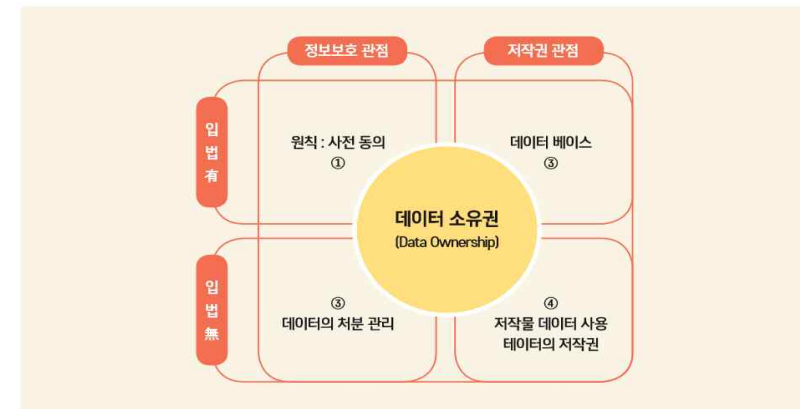
4) 데이터의 이용자는 공급자로부터 데이터를 입수·구매하여 직접이용하거나, 가공하여 재판매하는 객체

- 그러므로, 빅데이터 플랫폼 및 데이터센터 운영의 효율성 확보 차원에서 데이터 생태계의 공정 거래 질서확립, 법적 리스크 사전예방을 위해 이용약관을 사전진단하고 표준화 및 법적 검토를 수행하여 개선할 필요성이 있음
- 데이터 이용약관은 데이터 공급자인 빅데이터 플랫폼 및 데이터 센터와 데이터 이용자 간의 상호 동의를 통해 분쟁 및 이슈사항을 사전에 예방하고 상호 권리를 보호하는 것으로서 데이터 소유권 정립을 위한 기반으로 볼 수 있음

1.3 데이터 소유권의 정의와 현황

- 데이터 소유권(Data Ownership)은 데이터에 대한 가공처리에 관하여 법적 지위를 인정하는 개념으로 '4차 산업혁명 시대의 석유'로 언급되는 데이터에 대한 이해관계자의 권리를 보호하는 기반으로 볼 수 있음
- 소유권의 개념은 크게 민법에 의한 재산권과 저작권법에 의한 저작권으로 볼 수 있음
- 데이터 소유권의 개념에서 소유권은 민법 제211조와 제212조에 의한 '소유권의 객체인 물건' 또는 제98조의 '전기 기타 관리할 수 있는 자연력'으로 해석될 수 있으나, 데이터의 경우 지식정보자원이므로 현행 민법 및 재산권의 범주로 해석되기 어려움
- 데이터 소유권은 비록 민법의 소유권과 동일한 용어로 쓰이지만 같은 개념으로 보기 어려우며, 데이터는 무제한 복제·재생산·공유돼야 하기에 배타적·독점적 소유권 개념이 아니라 그 정보의 성질에 맞는 데이터 소유권 개념이 필요함
- 기존의 지식정보자원 보호를 관리하는 저작권의 경우, 저작권은 저작물을 창작한 자가 갖는 일련의 권리이므로 데이터 유통·거래의 대상인 데이터에 일반적으로 적용하기 어려움
- 예를 들어, 데이터의 경우, 저작권에 의해 보호될 수 있으면 대부분 창작성이 없거나, 데이터 원저작권자가 존재하는 경우, 저작권에 대한 보호가 어려움
 - 데이터베이스권과 같이 저작권법 범주 내의 권리도 데이터베이스 제작자의 권리로서 간접적으로 보호될 수 있는 것이며, 데이터 자체의 분석, 가공, 상품화에 대한 권리보호에는 한계가 있음
 - 또한, 데이터베이스권은 데이터베이스에 대한 독점적인 권리이므로, 데이터 유통·거래 과정에서 지속적으로 분석, 가공, 상품화되는 데이터의 특성으로 볼 때 법적 보호의 범위가 불확실하며 분쟁이나 이슈사항의 리스크가 존재
- 그러므로, 현재 데이터 소유권은 현행 법·제도 내에서 데이터 소유권을 정의하는 법률이 아직 미흡하므로, 소유권의 개념에서 데이터 소유권을 정의하기 어려운 실정임
- 현행 법·제도 내에서 데이터 소유권에 대한 관점을 나타낸 개념도는 아래와 같음

그림 18 데이터 소유권과 관련한 법·제도적 관점⁵⁾



- 현행 법·제도 내에서 정보보호의 관점과 저작권 관점, 입법 유무 등의 분석스코프를 통하여 데이터 소유권에 대한 관점을 4가지로 분류하였음
 - ①은 개인정보보호의 관점에서 개인정보활용동의에 의한 데이터의 수집과 축적에 대한 것으로 개인정보를 재산권적 요소와 인격권적 요소로 보아 이러한 정보가 포함된 데이터에 대해 데이터3법 개정, 가명정보처리나 익명화 등의 법·제도 개선이 이루어지고 있는 부분임
 - ②는 수집, 집적한 데이터셋(Data Set)의 사용, 수익창출, 처분 등의 권한에 대한 것으로 적법하게 데이터를 수집, 집적한 주체는 '점유자'로 볼 수 있으나, 이 주체를 이러한 데이터셋(Data Set)을 사용, 수익창출, 처분하는 등 행위를 할 수 있는지, 이것이 적절한 것인지에 대한 관점임
 - ③은 데이터베이스권에 대한 것으로 데이터베이스권은 저작권법에 의해 보호받아 데이터베이스를 구축한 자의 저작물로서 그 자체에 권리를 인정받으며 이에 대한 복제, 배포, 전송 등의 일련의 행위를 저작권법 제93조 제1항에서 정의하고 있어 일정부분 정리된 관점임
 - ④는 원저작권자가 있는 저작물에서 데이터를 추출하여 데이터셋으로 집적화 하거나, 데이터 간의 특정 패턴 또는 유사성을 분석하여 새로운 의미를 추출할 경우, 원저작권자의 동의가 필요여부나 새로운 저작물로서 인정을 받는지에 대한 관점임
- 대체적으로 데이터 3법이나 가명정보처리, 저작권법의 데이터베이스와 같이 기존의 데이터 관련 법·제도에 대해서 논의가 활발한 관점 ①, ③은 법·제도 개선방향성이 존재하나, 관점 ②, ④와 같이 아직 입법사항이 없거나, 데이터의 사용, 수익창출, 처분 등에 대한 핵심적인 사안에 대해서는 더 많은 연구와 사회적 합의가 필요함
- 특히, 데이터 소유권을 개념적으로 접근하였을 때, 관점 ②, ④와 같이 데이터센터, 데이터판매자 등 데이터를 수집, 직접한 주체가 단순히 개인정보활용동의나 별도의 동의서를 통해 이러한 행위가 민사 상 적법한 행위인지, 기존의 데이터를 활용하여 융복합데이터나 결합정보를 생성하였을 때 저작권을 인정받을 수 있는지 등이 주요 이슈로 발생하고 있음

5) 이종주, 데이터 소유권 동향, 월간SW중심사회 2019년 10월호, 2019

표 25 데이터 소유권 관련 국내외 법·제도 현황

구분	한국	EU	미국	일본
①개인정보공개	가명정보	GDPR규정	개별법	익명가공정보
②사용, 수익 처분	-	일부인정	일부인정	-
③데이터추출	-	EU저작권 규칙	판례존재	저작권법
④데이터베이스권	저작권법	EU지침	판례존재	저작권법

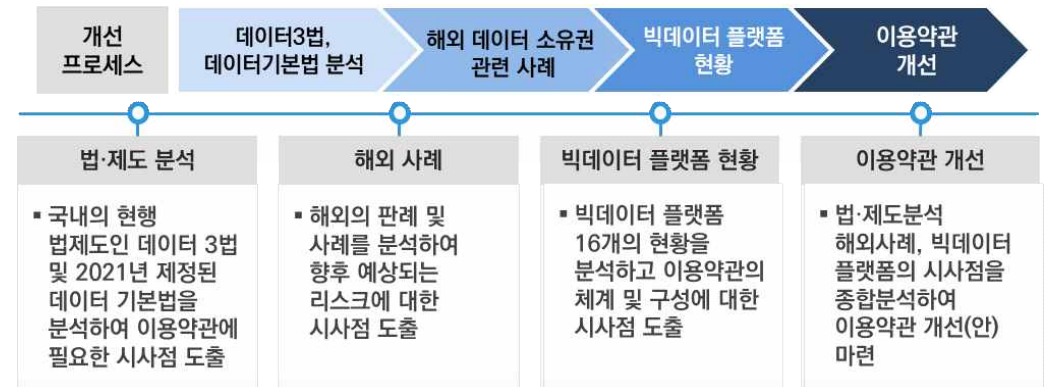
- 해외의 경우, 유럽연합(EU)의 경우, 과거 유사판례나 사례가 축적되어 일반개인정보보호법과 EU저작권 규칙, 각종 지침에서 데이터 소유권을 간접적으로 인정하거나 이에 대한 가이드를 안내하고 있음
- 미국과 일본의 경우, 판례로 데이터 소유권에 대한 내용을 다루고 있거나, 저작권법의 범주에서 데이터 소유권을 해석하고 있음
- 국내의 경우, 2020년 데이터 3법(개인정보보호법, 정보통신망법, 신용정보법)이 개정되며, 개인정보보호 및 가공과 관련하여 개인정보동의와 가명정보처리를 통해 활용이 가능해 졌으나, 해외 사례와 달리 아직 세부적인 지침이나 법·제도가 정비되지 않은 상태임
- 현재, 빅데이터 플랫폼 및 데이터 센터의 관점에서 데이터 소유권에 관련된 관점을 적용하였을 때 핵심키워드는 아래와 같이 정리할 수 있음
 - 데이터센터나 데이터판매자가 개인정보활용동의와 가명정보처리를 통해 데이터를 수집, 집적하여 플랫폼에서 판매할 시, 민사 상의 책임이 발생할 소지가 있는지
 - 데이터 센터나 데이터판매자가 빅데이터 플랫폼 내에서 데이터를 구매하여 융복합데이터나 결합정보를 생성하였을지, 이에 대한 수익창출에 대한 권리가 있는지
 - 빅데이터 플랫폼의 데이터 유통·거래 활성화를 위해 데이터 가치사슬(생성, 집적, 활용)의 행위자들이 적절한 보상이나 권리를 취득할 수 있는지
- 현실적으로 데이터에 대한 유통·거래가 활발하게 이루어지고 있으며, 이에 대해 재산적 가치가 존재함을 부정하지 않음
- 현재, 빅데이터 플랫폼 및 데이터센터에서 데이터의 유통·거래의 조건, 이용범위, 권리 등은 계약법에 따라 시장의 다양한 상업적 협의를 통해 정의되고 있으며, 데이터에 대한 법적 보호장치가 없어 오직 계약법이나 저작권(지식재산권)에 의해 부분적, 불완전하게 보호되고 있음
- 이는 데이터의 이용권의 개념이나 범위 내지는 그와 관련된 당사자간의 권리·의무를 규율하는 것이니 법률적으로 불명확성이 지속될 수 있기 때문에 장기적으로 악영향을 미칠 수 있음
 - 계약에 의한 권리·의무의 명시는 계약서에 의해 변동될 수 있으며, 모든 데이터 유통거래 과정이 동일한 기준 아래에서 보호를 받는 것이 아님
 - 그러므로, 향후 공정거래위원회의 표준계약서나 기준에 대한 법·제도가 충분히 뒷받침 되어야 시장질서 유지와 생태계 조성이 가능

- 데이터 계약은 당사자별로 상대적이고 유동적이기 때문에 법적 보호 수단으로 보기에는 미흡하며, 당사자 간의 협상 능력에 의해 데이터의 가치나 이용조건 등이 좌우되어 데이터 생태계 조성에 불확실성을 가중시킬 수 있음
- 그러므로 장기적으로는 데이터에 대한 안정적인 법적 보호가 있어야 데이터 생태계가 성장할 수 있으며 법적 보호 장치로서 데이터 소유권(Ownership) 개념의 도입이 필요하나, 현행 법·제도에서 데이터 소유권에 대한 명확한 정의와 권리설정이 미흡하므로, 빅데이터 플랫폼의 플랫폼·데이터판매자·데이터구매자 간의 상호동의체계와 데이터 3법의 개정사항과 데이터 산업진흥 및 이용촉진에 관한 기본법(약칭:데이터기본법) 등의 입법동향을 반영하여 이용약관 개선의 필요성이 있음

1.4 데이터 소유권 및 이용약관 개선방향성

- 데이터는 무제한 복제·재생산·공유돼야 하기에 배타적·독점적 소유권 개념이 아니라 그 정보의 성질에 맞는 데이터 소유권 개념으로 도입이 필요하며, EU도 역시 배타적이지 않은 데이터 특유의 소유권 개념을 전제로 논의 중임
- 데이터 소유권은 데이터의 자산 가치를 사회적으로 인정하는 것으로, 데이터 자산 가치를 사회적으로 수용하면서 궁극적으로 데이터의 거래나 유통 활성화, 나아가 데이터 가치의 극대화에 기여하는 것임

그림 45 데이터 소유권 및 이용약관 개선절차



- 데이터 3법 및 데이터기본법 분석에서는 개인정보보호법, 정보통신망법, 신용정보법의 개정사항과 데이터 기본법 제정에 의한 신규 법·제도를 분석하고 데이터 소유권 및 이용약관 개정에 필요한 시사점을 도출
 - 데이터 3법 개정에 따른 가명정보처리 가이드 준수여부와 데이터 유통·거래 현황에 반영이 중점적으로 필요
 - 데이터 기본법에 의한 표준계약서, 분쟁조정위원회 등 관련 사항 반영 필요

- 해외 사례에서는 유럽연합(EU)와 영국, 독일, 미국의 데이터 소유권과 관련 있는 판례를 분석하여 향후 데이터 소유권과 관련하여 예상되는 리스크를 식별하고 이에 대한 단기적인 시책을 시사점으로 도출
- 빅데이터 플랫폼 현황에서는 2020년 표준 이용약관을 기준으로 플랫폼 별 이용약관 개정현황과 진단을 수행하고, 이슈사항을 중심으로 이용약관 개정 필요 시시점을 도출
 - 플랫폼 회원제 운영여부, 플랫폼별 분쟁기구 설치 및 개입현황, 이용약관의 준거법 및 관할권에 이슈사항 분석이 필요
- 이용약관 개선에서는 데이터 3법 및 데이터기본법 등 법·제도, 해외사례, 빅데이터 플랫폼 현황의 시사점을 종합분석하여 이용약관 개선(안)을 마련하여 제시

02 해외 데이터 소유권 관련 사례 분석



2.1 해외 데이터 소유권 관련 사례 분석 개요

- 데이터 3법 개정과 데이터 기본법의 제정으로 데이터 유통·거래 활성화를 위한 법·제도적 기반이 마련되었으나, 데이터 소유권이 명확히 명시된 법이나 조문이 현재 제정되어 있지 않으므로 데이터 기본법의 '데이터 자산보호'와 같은 간접적인 형태로 보호받고 있음
- 기존의 데이터 자산에 대한 보호는 저작권법의 범주에서 지적재산권과 데이터베이스권의 형태로 보호받아 왔으나 데이터의 원저작권자가 권리행사를 하거나 데이터 자체가 아닌 데이터베이스 형성에 대한 권리를 보호 받아 데이터 소유권에 대한 적절한 판례가 부존재
 - 이는 데이터 소유권이나 데이터의 자산에 권리를 주장하기 위해서는 해당 데이터의 소유권자가 저작권자임을 증명해야되나, 이러한 판단기준이 불분명한 것에서 소가 성립하지 않거나 합의수준에서 종결되는 사례가 다수
- 데이터 소유권에 관련하여 이동진의 「데이터 소유권(Data Ownership, 개념과 그 실익」, 이상용의 「데이터 거래의 법적 기초」, 「빅데이터의 경쟁법적 함의에 관한 연구」, 최종모의 「빅데이터의 이용활성화를 위한 저작권법적 고찰 -데이터 마이닝 등을 중심으로-」 등의 연구가 지속되어 왔으나, 국내 법·제도 상 국내 판례를 제시하기 보다는 해외 사례를 분석하여 방향성을 도출
 - 국내의 법·제도 및 정책 상, 분쟁조정기구의 설치 등을 통해 불복 시, 민사소송을 통해 분쟁을 해결하거나 합의절차가 진행되며, 대한민국 법원 종합법률정보 등에 판례라 등록되지 않으므로 구체적인 사례 및 판례를 제시하기 어려운 문제점이 있음
 - 또한, 기존의 과학기술정보통신부 등에서 익명처리를 통해 일부 사례를 소개한 바 있으나 당해 이해관계자의 의사에 따라 비공개처리되는 상태이므로 절차적 적법성(Due Process) 있는 사례를 구조적으로 인용하기 어려움
- 그러므로 해외 데이터 소유권 관련 사례분석에서는 2021년 기준으로 해외에서 데이터 소유권 및 관련 동향으로 사법기관의 판단이 이루어진 사례를 소개하고, 빅데이터 플랫폼 및 데이터센터에서 분쟁사항에 대한 효율적인 대응을 위해 참고사례를 제공
 - 해외의 판례는 판단 이유나 결과에 대해서 시사점을 제공할 수 있으나, 국내법과 법률의

구현절차와 해석이 상이하므로 참고사례가 될 수 있으나 이를 인용한 법적 지침이 될 수는 없음

- 해외의 경우에도 데이터 소유권에 대한 법적 명시나 법·제도화가 진행 중인 상황으로 국내의 데이터 유통·거래 현황과 일치하는 사례가 발생하기 어려운 상황임
- 기존의 해외에 축적된 사례를 주요 동향과 데이터 포괄승계 및 이용, 저작물의 데이터 활용 관련 판례를 기반으로 빅데이터 플랫폼에서 발생할 수 있는 리스크를 분석
 - 주요 동향은 데이터 소유권에 대한 해외의 주요 판례와 시각을 분석
 - 데이터 포괄승계 및 이용은 데이터 판매자의 적법한 데이터 취득을 기준으로 빅데이터 플랫폼에 영향이 있는지 분석
 - 데이터 활용 관련 판례는 향후 빅데이터 활용 및 융복합데이터 개발이 원저작권자의 권리를 침해하는 것인지 분석

2.2 데이터 소유권 관련 주요 동향

- 국내와 같이 해외의 경우도 데이터 소유권에 대한 명시적인 법률이나 법·제도체계가 완성되어 있지 않은 실정
- 하지만, 데이터 산업에 대한 고려와 판례의 축적이 장기간에 걸쳐 이루어져 왔으므로, 국내보다 관련 사례 완성도와 관측이 용이
- 현재, EU GDPR(General Data Protection Regulation)과 같이 개인정보보호의 관점에서 데이터 소유권에 접근하는 경우가 있으며, 기존의 민법(재산권)이나 저작권법의 범주에서 데이터 소유권이 논의되고 있음

표 26 데이터 소유권에 대한 해외 각국의 주요 판례

구분	사건 개요	판단 이유
EU 및 유럽위원회 (Usesoft 사건)	<ul style="list-style-type: none"> 2005년 독일 중고소프트웨어 판매사이트에 오라클 영구라이센스를 중고소프트웨어로 판매 오라클 유럽지사에서 라이선스를 재판매한 것에 대해서 이용약관 위반 및 불법 취득으로 소송제기 2012년 유럽연합 법원에서는 영구라이선스 구매계약에 의해 소프트웨어를 취득하였으므로 구매자의 소유물은 본인 의사에 의해 가능하다 판결 	<ul style="list-style-type: none"> 사건의 대상이 된 오라클 영구라이선스는 인터넷에서 소프트웨어를 다운받고 인증키로 입력하는 방식 오라클 유럽지사의 주장은 영구라이선스라 하더라도 저작권자인 라이선스 인증키를 재판매한 것은 오라클을 침해한 것으로 인식 유럽연합 법원에서는 영구라이선스의 형태가 구매자에게 이용권리를 판매한 것이므로 이를 재판매하여도 공정이용으로 판결 또한, 이 라이선스를 구매, 또는 재구매한 구매자들도 공정이용에 해당

구분	사건 개요	판단 이유
독일 (연방소비자 연맹의견)	<ul style="list-style-type: none"> 데이터 소유권을 EU GDPR(일반정보보호규칙)에서 개인정보보호관련 규정으로 보완 시도 구체적인 입법과정에서 연합협의체에서 반대의견이 존재 독일의 경우, 타국의 판례를 분석하여 입법활동과 이에 대한 민간 의견 수렴 등의 절차가 수행 중 	<ul style="list-style-type: none"> 독일 연방소비자연맹은 데이터에 대한 권리 보장의 특별성, 기본 법·제도에 대한 보완이 선행필요, 정보주체 동의와 의사표시에 대한 방안과 개인정보 비식별화에 대한 실효성을 문제로 제기 데이터는 인터넷과 같이 정보가 공유될 수 있는 것이며, 이에 대한 소유권적 배타적 권리는 불필요하다 주장 저작물과 같이 저작권법에 범주에서 보호받거나 활용관점에서 권리보호가 필요하다고 보았음
영국 (Oxford v Moss 사건, Fairstar Heavy Transport NV v Adkins 사건)	<ul style="list-style-type: none"> 1968년 옥스퍼드대학교 토목공학과 학생인 Moss는 사전에 시험지를 확인하는 등 부정행위를 행함 이에 대학당국은 Moss를 절도죄로 처벌을 주장 1968년 당시, 법원은 사전에 시험지를 확인하여 기억하는 것을 절도로 볼 수 없으며 정보를 재산으로 볼 수 없다하여 기소를 기각 2013년 Adkins는 Fairstar Heavy Transport NV사의 CEO로서 일산상의 이유로 퇴사 Fairstar Heavy Transport NV사는 Adkins의 근무 중, 메일 등 업무기록은 회사 소유라 주장 해당 사건에 대해 영국 리버풀지방법원은 메일 등 업무기록은 재산의 한 형태로 볼 수 없으며, 회사의 보안문제라 판결 다만, 메일 등의 데이터에 대해서는 보호가 필요한 것으로 의견 첨부 	<ul style="list-style-type: none"> 영국의 경우, 고전적으로 저작물이라 하더라도 물리적 실체가 있어야하며, 정보를 기억하거나 구술하는 것은 절도로 보지 않았음 (Teff Act 1968) 2013년 사건의 경우, 법원에서 메일 등 업무기록(저장매체, 데이터, 파일 등을 포괄)은 회사 내 보안문제이며 절도로 보기 어렵다는 판례를 유지 이는 데이터의 복제라는 특성을 문서 등과 동일하게 본 것으로 전문위원 의견에서 법원이 법적 쟁점에 대한 접근이 부족한 것을 언급 다만, 데이터의 도난을 증명할 수 있는 기술(완전성, 진본성)이 있을 경우, 절도가 성립할 수 있으나 이에 대한 증거가 부족하여 기존의 법률과 판례를 근거로 논의할 수 밖에 없었던 한계가 있음
프랑스 (Blutouff 사건)	<ul style="list-style-type: none"> 2012년 Blutouff라는 가명으로 VPN을 통해 프랑스 OIV(중요기반시설) 인터넷에 접속하여 데이터(약82GB)를 무단으로 다운로드 및 공개 프랑스 형법 제311조에 따라 타인의 재산 절도 및 횡령으로 기소 데이터의 삭제가 없었고 데이터 	<ul style="list-style-type: none"> 사건의 핵심은 정부인트라넷에 접속한 것에 대한 것과 데이터의 사본복제가 절도 및 횡령인지에 대한 것 프랑스 Creteil 지방법원(1심)에서는 정부인트라넷의 결합이 있는 것과 데이터의 삭제, 이익추구가 없던 것에서 무죄 선고

구분	사건 개요	판단 이유
	<p>도난에 대한 증거가 없으므로 무죄 판결</p> <ul style="list-style-type: none"> 2015년 이후, 프랑스 형법 제323조의 3이 개정되어 데이터의 도난을 인정 	<ul style="list-style-type: none"> 2015년 이후, 인가되지 않은 데이터 다운로드에 대해서는 불법과 정부인트라넷에 접속을 대테러 행위로 규정

- 먼저, EU 및 유럽위원회의 Usesoft 사건은 독일의 소프트웨어 중고판매사이트에서 기존에 오라클 영구라이센스를 구매한 자가 사이트에서 영구라이센스 인증키를 재판매한 사례로 이를 구매한 사람의 권리가 적당한지에 대한 판례임
- 중고판매사이트를 통해 동일한 영구라이센스가 1차판매와 2차판매까지 이루어졌으며, 이러한 것이 오라클 유럽지사의 이익을 침해한다는 것으로 소송이 제기
- 해당 판례는 독일에서 소가 제기되어 유럽연합법원까지 상고된 케이스로 최종적으로 유럽연합법원에서는 영구라이센스 구매계약은 영구적으로 해당 라이선스 및 해당 버전의 소프트웨어를 권리를 양도하고 대가를 받는 것으로 물건과 같은 소유권이 인정된다고 보았음
- 그러므로, 이를 판매하고 재판매하는 것은 물건과 같은 소유권의 범주로 이해해야하며, 오라클 영구라이센스의 이용은 정당한 이용권리로 보았음
- 이는 계약에 의한 소유권의 특성과 효력에 대한 판례로 계약절차와 조건이 정당하였으므로 적법하다고 판단한 사례임
- 독일의 경우, 연방의 정책수립이나 의견수렴을 위한 연방협의체가 구성되어 있으며 연방협의체 중 하나인 연방소비자연맹에서 데이터 소유권에 대한 법·제도 방향성에 대해서 의견을 제시하였음
- 데이터는 인터넷과 같은 21세기의 자원이라 칭해지고 있으나, 이에 대한 배타적 권리(소유권)이 보장되는 것은 지식정보의 확산을 저해하는 것으로 인식하였으며, 데이터의 복제라는 특성으로 인해 고전적인 재산권적인 정의는 어렵다고 보았음
- 또한, 이미 EU GDPR 등 개인정보보호의 관점과 판례로 기존의 법·제도가 보완되고 있으므로 데이터만을 위한 별도의 법·제도 필요성에 대해서 의문성을 제기
- EU GDPR의 개인정보보호에 관한 규정에서 개인정보의 비식별화(가명정보)는 데이터가 조합될 경우 실효성을 잃으므로, 이것이 개인정보보호의 원칙과 부합하는 것인지에 대한 관점을 제시
- 마지막으로 마이데이터와 같은 자기정보결정권, 정보주체의 동의와 의사표시를 데이터 산업의 활성화 명목으로 강요하거나 의무화하는 것은 시민의 자유를 침해할 수도 있는 것이며 이와 관련된 논의와 체계가 수립필요를 주장
- 영국의 경우, 고전적으로 정보의 절도에 대하여 물리적 실체를 고려하는 시각이 있었으며, 2010년대에는 데이터 소유권이나 권리에 대한 논의가 없는 것은 아니나 아직 법·제도나

6) 연방소비자연맹은 연방협의체로 연방에서 인정한 시민단체 혹은 기관으로 정책수립 과정에 참여

판례상에서 이를 확인하기 어려움

- Adkins의 사례와 같이 무형의 지식정보에 대해서는 재산권을 적용하기 어렵다고 보고 있으며, 절도나 도난을 증명하기 위해서는 이에 대한 증거가 필요하다고 보는 입장
- 개념적으로는 데이터의 절도나 도난은 존재할 수 있으나, 절도나 도난당한 대상에 대해 본 소유자의 소유라는 것에 대한 증명이 필요
- 이러한 관점에서 데이터의 완전성, 진본성을 기술적으로 증명하는 블록체인 기술이나 암호화 등의 기술의 발전이 보완적으로 이루어지고 있음
- 다만, 영국의 경우에도 데이터 소유권에 대한 개념에 동의는 할 수 있으나 사회적 협의나 검토가 필요하다는 전문위원 의견이 판례에 존재
- 프랑스의 경우, 영국과 같이 데이터의 절도나 도난에 대한 증명이나 데이터가 복제되는 특성에 의해 데이터가 공개나 유포되어도 이를 어떠한 권리 침해로 볼 것인지에 대해서 이슈가 있었으나, 2015년 이후로 무단침입이나 대테러 등으로 형법을 보완
- 데이터의 삭제, 복제, 공개나 유포가 어떠한 권리를 침해하였는지에 대해 2012년 판례에서 검토가 이루어 졌으며, 2015년 이후에는 이를 타인의 권리침해나 형법적인 측면에서 접근한 사례가 있음
- 프랑스는 형법의 개정과 정부관점에서 데이터 소유권의 개념을 본 것으로 국내의 데이터 소유권에 대한 논의와는 거리가 있음
- 전반적으로 해외 사례에서는 고전적인 민법(재산권)에서 데이터 소유권을 인정하지 않으며, 저작권이나 계약의 적법성에 의해서 데이터에 대한 권리가 보호받고 있음
- 이는 기존의 국내에서 데이터 소유권에 관련된 연구와 맥락이 같은 것으로 현행 체계 내에서는 저작권과 계약 관련 법을 통해 기초적으로 데이터 소유권이 보호받을 수 있으며, 이러한 체계를 개선하기 위해서는 더 많은 판례와 논의가 필요
- 빅데이터 플랫폼의 관점에서는 해외 사례를 국내에 일반적으로 적용하기 어려우나, 계약에 의한 데이터 이용의 적법성 관점으로 볼 때, 데이터 판매자와 데이터 구매자 간의 계약이 공정한 조건으로 체결될 필요성이 있음

2.3 데이터 포괄승계 및 이용 판례

- 데이터 포괄승계 및 이용판례는 빅데이터 플랫폼에 기업 간의 결합 및 포괄승계(인수합병 등) 시, 데이터를 이용할 수 있는 권리가 해당 기업체에 있는지에 대한 내용⁷⁾
- 데이터 포괄승계 및 이용은 공정거래법 제3조의2 1항 1호에 깊게 관련되는 내용으로 시장독점적인 지위의 취득이나 기업 간의 결합 및 포괄승계 시, 보유한 데이터를 이용할 수 있는 권리가 정당한지에 대한 판례임

7) 이호영, 빅데이터의 경쟁법적 함의에 관한 연구, 2018

- 데이터 포괄승계 및 이용은 데이터 소유권과 관련이 적어보일 수 있으나, 빅데이터 플랫폼에 등록되는 데이터 생산취득에 적법성이 있는지, 이러한 데이터를 등록할 시 빅데이터 플랫폼에 리스크가 발생하는지 조사하기 위하여 판례를 소개

표 27 데이터 포괄승계 및 이용 관련 판례

구분	사건 개요	판단 이유
구글의 더블클릭인수 사건(2007)	<ul style="list-style-type: none"> 구글이 고객 데이터를 다량 보유한 더블클릭(DoubleClick)을 인수하면서 광고 시장에서 우위를 점하는 것에 대해 미국 연방거래위원회는 승인함 	<ul style="list-style-type: none"> 기업을 인수하는 것에 대해 우위를 점하는 것은 아니고, 개인정보보호를 위한 경우가 아닌 한 기업결합이 적법하다고 판단함 경쟁법 집행과 데이터 호의 관계를 본 격적으로 다룬 최초의 사례
톰슨의 로이터 인수 사건 (2007)	<ul style="list-style-type: none"> 투자자에게 재무 관련 정보를 제공하는 사업자인 톰슨(Thomson)이 동일한 서비스를 제공하는 경쟁자 로이터(Reuters)를 인수하기로 한 것에 대해 미국 연방법무부와 유럽 집행위원회는 결합 기업이 보유한 일정한 범위의 기초 데이터(Fundamentals Databases) 등을 제3자에게 매각하는 조건으로 기업결합 승인 	<ul style="list-style-type: none"> 재무 관련 정보를 취합하는 것은 시장에서 독점적 지위를 형성하지 않는다고 판단
리드엘세비어의 초이스포인트 인수 사건 (2008)	<ul style="list-style-type: none"> 수사기관이나 국가기관 등의 직무수행에 필요한 공적 기록 서비스와 정보분석 서비스를 제공하는 리드엘세비어(Reed Elsevier)가, 역시 미국 내 정부 기관을 포함한 고객들에게 다양한 경제적 위험관 리 서비스를 제공하는 초이스포인트(ChoicePoint)를 인수하기로 한 것에 대하여, 연방거래위원회는 미국 내 법집행기관에 대한 전자 공적 기록 서비스 시장에서 경쟁제한성이 인정된다고 판단하고 결합 기업에 대하여 초이스포인트의 전자공적 기록 서비스 사업 부문을 경쟁자인 톰슨 로이터스에 매각하도록 하는 내용의 동의명령을 내림 	<ul style="list-style-type: none"> 단 두 개뿐인 주요 사업자의 경쟁이 소멸될 가능성이 있고, 진입장벽이 높은 시장의 특성으로 인해 신규 경쟁자 발생이 어려워 두 회사 간의 기업결합 불허
닐센의 아비트론 인수 사건 (2012)	<ul style="list-style-type: none"> TV 및 교차 플랫폼 시청률 조사 서비스 등을 제공하는 사업을 영위하는 닐센(Nielson)이, 역시 라디오와 교차 플랫폼 시청률을 포함한 다양한 시청률 조사 서비스 등을 제공하는 사업을 영위하는 아비트론(Arbitron)을 	<ul style="list-style-type: none"> 관련 시장을 '미국 내 전국 연합 교차 플랫폼 시청률 조사 서비스 시장'으로 획정하면서 경쟁제한성을 인정하였고, 시장에서 기업결합은 서비스의 가격이 인상되고 혁신을 저해할 가능성이 높다고 판단함

구분	사건 개요	판단 이유
	인수하기로 한 것에 대하여, 연방거래위원회는 시장에서 경쟁제한성이 인정되므로 수집한 데이터를 외부에 매각하는 동의명령을 내림	
톰톰의 텔레아틀라스 인수 사건 (2007)	<ul style="list-style-type: none"> 이동식 내비게이션 장치와 내비게이션용 소프트웨어를 공급하는 톰톰(Tom Tom)이 유럽과 북미 지역에서 내비게이션에 사용되는 디지털 지도 데이터 베이스를 공급하는 텔레아틀라스(Tele Atlas)의 주식 전부를 공개매수하기로 한 것에 대하여 유럽 집행위원회는 심사를 진행한 결과 경쟁제한성이 인정되지 않는다고 판단하고 승인 	<ul style="list-style-type: none"> 텔레아틀라스의 데이터베이스 매출 감소로 인하여 발생하는 손실이 이동식 내비게이션 장치 시장에서 시장지배력을 강화함으로써 얻을 수 있는 이익보다 클 것 이므로, 톰톰이 봉쇄전략을 취할 유인이 인정되지 않고, 내비게이션 장치 시장을 볼 때 경쟁제한 효과가 나타나지 않을 것 이므로 경쟁제한 효과 발생하지 않을 것으로 판단
페이스북의 왓츠앱 인수 사건 (2014)	<ul style="list-style-type: none"> 사회관계망 서비스, 온라인 메신저 서비스 등을 제공하는 페이스북이 모바일 메신저 사업자인 왓츠앱(WhatsApp)을 인수한 것에 대하여 미국 연방거래위원회와 유럽 집행위원회 모두 경쟁제한성이 인정되지 않는다고 판단하여 인수 허가 결정 	<ul style="list-style-type: none"> 해당 시장은 이미 다수의 경쟁자가 존재하며, 기업 결합으로 인해 데이터를 활용하려면 양 기업의 프라이버시 정책을 변경해야 하는 것과 기술적 제약이 존재하므로 경쟁에서 부당한 우위를 차지하게 될 것으로 보지 않음
마이크로소프트의 링크드인 인수 사건 (2016)	<ul style="list-style-type: none"> 소프트웨어를 공급하고 온라인 광고사업도 영위하는 마이크로소프트(Microsoft)가 전문직 사회 관계망(Professional Social Network : PSN) 서비스, 온라인 구인 서비스 등을 제공하는 링크드인(LinkedIn)을 인수하기로 한 것에 대하여 유럽 집행위원회가 경쟁제한성을 인정하고, 마이크로소프트 소프트웨어의 링크드인을 자유롭게 삭제할 수 있는 것과 타 경쟁자가 해당 서비스를 지속적으로 운영하고 데이터에 자유롭게 접근하는 것을 조건으로 기업 인수를 승인함 	<ul style="list-style-type: none"> 마이크로소프트의 시장지배력이 링크드인의 시장까지 영향력을 끼쳐 경쟁자를 배제할 가능성이 높고, 소비자의 선택권 개인의 프라이버시권 등에 대해 침해 가능성이 높다고 판단함
페이스북의 개인 정보 수집 사건 (2019)	<ul style="list-style-type: none"> 독일 경쟁당국인 연방 카르텔청은 소셜네트워크 시장의 지배적 사업자인 페이스북이 개인정보를 착취하는 행위로 경쟁법을 위반하였다고 결정 	<ul style="list-style-type: none"> 페이스북이 실질적인 동의를 받지 않은 상황에서 제3자로부터 이용자들의 개인 정보를 수집하고 이를 이용자의 페이스북 계정에 연계·통합 시키는 부분은 개인정보보호법령에 위반되며 착취적 거래조건 강제행위로 경쟁법에 위반된다고 판단

- 데이터 포괄승계 및 이용에 대한 빅데이터 플랫폼 및 데이터센터에 미치는 영향은 3가지로 데이터의 빅데이터 플랫폼에 데이터셋을 등록하는 시점에서 작용
 - 적법한 데이터 자산의 생산·취득과 유통·거래
 - 개인정보보호의 관점에서 데이터 3법의 준수여부
 - 공정거래법 및 공정거래위원회의 소관 범위 확인
- 먼저, 데이터의 포괄승계 및 이용에 있어서 적법한 절차를 통해 데이터를 생산·취득하여 빅데이터 플랫폼에 등록, 데이터를 유통·거래하는 것은 관계 법률을 준수하여야 함
 - 대부분의 판례는 개인정보나 기조데이터를 기업 간의 인수·합병으로 인해 데이터 자산의 독점여부나 데이터를 이용하는데 개인정보보호 영역을 침해하는지에 대한 내용
 - 이 부분에 대해서는 데이터 소유권의 관점보다도 시장독점이나 개인정보보호 영역에서 판례의 방향성이 달라지며, 권리의 보호보다는 데이터 취득의 적법성에 가까움
- 두번째는 데이터의 포괄승계 및 이용에 있어서 개인정보보호 및 가명처리의 필요성과 이를 확인하는 절차가 이용약관의 요소로 포함되어야 함
 - 데이터의 생산·취득과 유통·거래는 데이터의 판매자 소관이나 빅데이터 플랫폼의 입장에서는 적법한 데이터를 유통·거래하도록 하여 법적 리스크를 해소해야함
 - 특히, 데이터 3법의 개정 및 가명정보 개념의 도입으로 특정목적에 한하여 개인정보보호를 침해하지 않는 방안이 있으며, 현재 국내에서 데이터의 결합이나 추출·활용에 대해서는 법·제도화되어 있지 않으나 개인정보보호위원회의 가명정보처리 가이드 준수하여 리스크를 최소화해야함
- 세 번째는 데이터의 생산·취득 적법성, 시장적 독점지위나 개인정보보호의 의무는 데이터를 등록하는 자에게 있으며, 기업 간의 인수·합병이나 데이터 자산 취득의 적법성은 공정거래법과 공정거래위원회의 소관으로 볼 수 있음
 - 데이터의 생산·취득 적법성은 공정거래법에 의해 민사소송이나 분쟁이 발생하였을 경우, 공정거래위원회에서 조정 및 판단할 문제임
 - 빅데이터 플랫폼의 입장에서는 데이터 3법 및 가명정보처리에 대한 의무는 데이터 판매자에게 있음을 방침으로 하고 이를 이용약관에서 명시할 필요성이 있음

2.4 저작물의 데이터 활용 관련 판례

- 기존의 데이터는 정보지식자원으로 간주되어 저작권법의 범주에서 권리를 보호 받는 경우가 많으나, 이는 데이터를 직접 생산한 자일 때 한정
- 데이터의 직접 생산하고 이를 저작권으로 등록하거나, 저작권자임이 확실할 때, 데이터 소유권의 맥락과 같이 데이터의 추출 및 활용에 대한 적법성이 판례로 남겨나 사례로 등록될 수 있으나, 이러한 사례가 대외적으로 소개되기 어려움

- 그러므로, 데이터에서 필요한 부분을 추출하여 재생산하거나 결합하는 경우의 판례를 참고하기 위해서는 기존의 저작물에 대한 판례를 살펴봐야 하며 어떠한 이유에서 공정이용으로 인정되었는지 조사할 필요성이 있음

표 28 저작물의 데이터 활용 관련 판례

구분	판시 내용	요약
Authors Guild, Inc. v. HathiTrust 사건	<ul style="list-style-type: none"> 원 저작물을 재발행하거나 재포장하는 것 보다는 새로운 목적 또는 추가적인 목적이거나 다른 성질 또는 새로운 표현을 수반하여 바꾸거나 의미 또는 메시지 등을 추가하여야 한다고 판시 	<ul style="list-style-type: none"> 변형적 이용일 경우 공정이용에 해당
Fox News Network, LLC v. TVEyes, Inc. 사건	<ul style="list-style-type: none"> Fox News Network, LLC 의 프로그램(저작물)을 피항소인(피고) TVEyes의 고객이 시청토록 하여 상업적 용도로 이용하여도 공정이용에 해당할 수 있다고 판시 	<ul style="list-style-type: none"> 상업적 용도로 이용하여도 공정이용에 해당
Campbell v. Acuff-Rose Music 5 사건	<ul style="list-style-type: none"> 공정이용 해당여부에 대하여 다른 이용과 동일하게 공정이용의 요소를 통하여 저작권법이 추구하는 목적에 따라 사안별로 판단함에 따라 이용의 목적이 상업적이거나 비영리적 등은 이용목적 및 성격의 한 요소이라고 판시하였을 뿐만 아니라, 공정이용 해당여부에 영향을 미치는 상업주의 등의 요소에 대한 중요도는 변형적 저작물에서는 감소한다고 판시 	<ul style="list-style-type: none"> 패러디 등의 변형적 저작물의 목적이 상업적이라고 할지라도 공정이용이 적용

- 특정 저작물에서 데이터나 일부를 추출하여 사용하거나, 이를 재가공하고 상업적으로 이용할 경우, 이러한 이용은 부당이용인지 아니면 공정이용인지에 대한 판례
 - 부당이용과 공정이용은 저작권자의 권리를 침해하여 부당한 사례로 손해배상의 책임이 있는지와 저작권자의 권리를 침해하지 않은 것으로 해석할 수 있음
 - 이는 공정이용이라도 상업적 목적으로 데이터나 일부를 추출하여 이익을 얻었을 경우, 이 행위에 대한 책임을 묻는 것이지, 이로 인해 얻은 이익을 독점적으로 보유한다는 의미는 아님
 - 즉, 공정이용이라도 원저작권에 대한 판권 및 원고료나 이익을 공유할 필요성이 있는 것이며, 이는 별도의 민사소송으로 다루어져야하는 문제임 (별도 소송 가능)
- 빅데이터 플랫폼 관점에서 저작물에서 데이터의 추출 및 이용은 데이터 결합이나 융복합데이터를 개발할 시, 이것이 부당이용인지에 대한 시사점으로 볼 수 있음
 - 저작물에서 데이터를 추출하여 이용할 시, 어느정도 변형이 있거나 저작물의 성격에 따라 공정이용으로

볼 수 있음

- 공정이용이 배타적인 이익독점이나 원저작권자에 대한 보상에서 자유로운 것은 아니나 부당이용이 아니므로 손해배상의 책임이나 불법성은 담보하기 어려움
- 2017년 지식사회를 위한 저작권법[Gesetz zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschafts(UrhWissG)]에서 데이터 마이닝에 대한 규정 신설과 해당법령 제60d에서 저작물의 데이터 추출 및 활용을 허용
- 그러므로 빅데이터 플랫폼에서 유통·거래된 데이터에서 데이터를 추출 또는 결합, 융복데이터로 개발하여 활용하였을 경우, 해외 판례 상 부당이용으로 보기 어려움
- 만약, 빅데이터 플랫폼에서 저작물의 데이터 추출 및 이용에 관한 이슈가 발생하였을 경우, 이를 플랫폼에서 이용약관 등으로 사전에 제약할 사항은 없으며, 원저작권자와 데이터 판매자, 데이터 추출 및 이용한 자 간의 분쟁은 이해관계자 간의 별도의 소송으로 다뤄져야할 문제로 볼 수 있음

2.5 해외 데이터 소유권 분석 시사점

- 해외 데이터 소유권에 대해 주요 동향 및 데이터 포괄승계 및 이용, 저작물의 데이터 활용 판례 및 사례를 분석하여 법·제도 분석 시사점으로 도출한 것을 종합적으로 분석한 결과는 아래와 같음

표 29 해외 데이터 소유권 관련 사례 분석

구분	시사점	종합
해외 데이터 소유권 주요 동향 분석	<ul style="list-style-type: none"> • EU 및 유럽위원회의 사례와 같이 무형자산, 데이터 등의 소유권은 계약의 형태에 의해 인정될 수 있음 	<ul style="list-style-type: none"> • 현행 법·제도 내에서 데이터 소유권 보호의 기본적인 사항은 데이터 판매자와 데이터 구매자 간의 공정한 계약 체결이 필요
	<ul style="list-style-type: none"> • 독일과 같이 데이터 소유권에 대한 비판적인 입장이 존재 	
	<ul style="list-style-type: none"> • 영국과 같이 데이터 소유권의 개념에 대해서 기존의 법·제도로 해석하는 사례가 존재 	
데이터 포괄적 승계 및 이용 분석	<ul style="list-style-type: none"> • 프랑스와 같이 데이터에 대한 이용목적에 의해 소관 법이 적용되는 사례가 존재 	<ul style="list-style-type: none"> • 빅데이터 플랫폼의 이용약관에 현행 법·제도에 대한 준수를 의무화 필요
	<ul style="list-style-type: none"> • 빅데이터 플랫폼에 적법한 데이터를 등록해야하며, 이에 대한 준수여부는 데이터를 등록하는 자의 책임 	
	<ul style="list-style-type: none"> • 개인정보가 포함된 데이터의 가명처리 등의 처리는 데이터를 등록하는 자가 수행 	
	<ul style="list-style-type: none"> • 공정거래법의 부당이익, 독점방지는 공정거래위원회의 소관이므로 이에 대한 준거법 포함 필요 	<ul style="list-style-type: none"> • 가명정보 처리 등 개인정보보호 관련은 개인정보보호위원회 및 데이터 3법, 가명정보처리 가이드 준수가 필요
		<ul style="list-style-type: none"> • 공정거래법은 데이터 시장과 관련된 내용을 포괄하므로 준거법으로 명시필요

저작물의 데이터 추출 관련 판례	<ul style="list-style-type: none"> • 저작물의 데이터 추출 및 활용에 대해서는 공정이용으로 보고 있음 	<ul style="list-style-type: none"> • 데이터 추출 및 활용은 공정이용의 범주로 볼 수 있으나, 데이터 판매자와 구매자 간의 이슈사항은 상호간의 민사소송이므로 빅데이터 플랫폼이 개입할 필요가 없음
	<ul style="list-style-type: none"> • 저작물의 데이터 추출 및 활용이 공정이용이라도 이에 대해 꼭 배타적 독점이 인정되는 것은 아니므로 이해관계자 간의 민사소송 제기 가능 	

- 소유권 관련 사례의 주요 시사점은 기존의 법·제도의 범주에서 계약법에 의한 권리보호로 볼 수 있으며, 해외 경우에도 데이터 소유권에 대한 정립이 진행중이므로 빅데이터 플랫폼에서는 사전에 해외 사례와 같은 리스크에 대해서 이용약관에서 방안을 마련할 필요성이 있음

03 빅데이터 플랫폼 이용약관 현황

3.1 빅데이터 플랫폼 이용약관 개요

- 데이터의 유통·거래 생태계 조성을 위해 빅데이터 플랫폼 및 데이터센터의 데이터 자산을 보호하고 활성화 기반이 되는 이용약관의 개선이 필요
- 빅데이터 플랫폼과 같은 서비스를 운영하기 위하여 플랫폼의 정책 및 방향성, 이용자의 권리·의무, 책임사항을 규정하여 법적 리스크를 사전에 예방하고 각 분야의 플랫폼 현실과 부합하는 이용약관 개선이 필요
- 2019년부터 빅데이터 플랫폼 및 데이터센터 구축이 수행되어 왔으나, 각 플랫폼 별로 이용약관을 작성하여 시간적, 비용적으로 비효율적인 측면이 있어 2020년부터는 표준 이용약관을 제정하여 플랫폼에 권장하였음
- 현재, 국내 법·제도상으로 데이터 자산에 대한 보호와 개인정보보호와 같은 책무 이행을 위해서는 데이터 3법과 데이터기본법과 계약법 및 계약체결의 기반이 되는 권리설정이 이용약관에 필요하며 이를 법률전문기관의 검토를 통해 주기적으로 현행화 및 고도화를 수행
- 빅데이터 플랫폼 이용약관 현황분석에서는 2020년 기준 표준 이용약관에 근거하여 분석 기준을 수립하여, 각 플랫폼에서 표준 이용약관의 준수여부와 개선방향성을 도출하여 이용약관 개선에 반영

3.2 빅데이터 플랫폼 이용약관 분석기준

- 빅데이터 플랫폼 이용약관의 분석 기준은 2020년 표준 이용약관의 준수여부와 플랫폼별 현황을 고려하여 변경내역이 있는지를 분석
 - 표준 이용약관 기준으로 플랫폼별 이용약관을 분석하고, 각 조문별로 표준이용약관의 내용이 포함되어 있거나, 일치할 경우 준수로 판정
 - 표준 이용약관에 해당하는 조문이나 내용이 이용약관에 포함되지 않거나, 표준 이용약관의 방향성과 다른 내용으로 구성되어 있을 경우 미준수로 판정

- 표준 이용약관의 내용을 인용하거나 유사하지만, 수정이 발생한 사례는 별도로 표시

3.3 빅데이터 플랫폼 이용약관 현황분석 결과

- 빅데이터 플랫폼 이용약관 현황분석은 2021년 7월 각 플랫폼에서 제출받은 이용약관과 동년 11월 기준으로 16개 빅데이터 플랫폼의 이용약관을 점검하여 분석수행
- 빅데이터 플랫폼 이용약관 현황분석은 2020년 표준 이용약관의 17개 조문을 기준으로 수행하고, 준수의 경우 '○'를, 미준수나 해당내용이 없는 경우는 공란처리, 특이사항이나 플랫폼에서 별도로 수정을 가한 것은 '△'로 표시하여 분석

표 30 빅데이터 플랫폼 표준 이용약관 준수현황 (1/2)

구분	금융	환경	문화	중소기업	교통	유통·소비	통신	헬스케어
제1조 목적	○	○	○	○	○	○	○	○
제2조 정의	○	○	○	○	○	○	○	△
제3조 약관 등의 명시·개정	○	○	○	○	○	○		△
제4조 서비스의 제공·변경	○	○	○	○	○	○		△
제5조 서비스의 중단	△	○	△	○	○	○	○	△
제6조 서비스 이용계약신	○	○	○	○	○	○		
제7조 서비스가입 신청·승낙	○	○	○	○	○	○	○	
제8조 서비스이용 계약종료	○	○	○	○	○	○	○	
제9조 회비 환불정책						○		
제10조 회원에 대한 통지	○	○	○		○	○	○	△
제11조 개인정보 제공 동의	○	○	○	○	○		○	○
제12조 유료 데이터상품의 구매	○	○	○	○	○	○	△	
제13조 데이터상품 분석 환경	○	○				○	△	
제14조 데이터상품 환불	○	○	○	○	○	○	△	
제15조 저작권의 귀속 등	○	○	○				○	○
제16조 분쟁해결								
제17조 재판권 및 준거법	○	○	○	○	○	○	○	○
약관시행·개정일	'19.12	'20.10	'21.06	'20.01	'20.03	'21.01	'19.12	'20.01

표 31 빅데이터 플랫폼 표준 이용약관 준수현황 (2/2)

구분	지역경제	산림	소방안전	스마트치안	해양수산	농식품	라이프로그	디지털산업
제1조 목적	○	○	○	○	○	○	○	○
제2조 정의	○	○	○	○	○	○	○	○
제3조 약관 등의 명시·개정	○	○	○	○	△	○	○	○
제4조 서비스의 제공·변경	○	○	○	○	○	○	○	○
제5조 서비스의 중단	○	○	○	○	○	○	○	○
제6조 서비스 이용약관·신청	○	○	○	○	○	○	○	○
제7조 서비스가입 신청·승낙	○	○	○	○		○	○	○
제8조 서비스이용 계약종료	○	○	○	○		○	○	○
제9조 회비 환불정책			○	○				○
제10조 회원에 대한 통지		○	○	○			○	○
제11조 개인정보 제공 동의	○	○	○	○	○	○	○	○
제12조 유료 데이터상품의 구매	○	○	○	○			○	○
제13조 데이터상품 분석환경	○		○	○			○	○
제14조 데이터상품 환불	○		○	○		○	○	○
제15조 저작권의 귀속 등	○			○	○	○	○	○
제16조 분쟁해결	○	△	△	○		△	○	○
제17조 재판권 및 준거법	○	○	△	○	○	○	○	○
약관시행일	'21.11	'19.12	'21.01	미명시	'20.06	미명시	'21.04	'21.02

- 빅데이터 플랫폼은 대체적으로 표준 이용약관을 준수하고 있으나, 플랫폼의 개별 현황을 반영하여 이용약관에 일부 사항을 반영하고 있는 것으로 보임
- 16개 플랫폼의 표준 이용약관 준수에 대한 주요동향은 3가지를 분석할 수 있음
 - 먼저, 대부분의 플랫폼이 회원제 운영 및 회비 정책에 대한 이용약관 내용이 누락되어 있는경우가 많으며 이것은 플랫폼에서 회원제를 미운영하기 때문으로 분석
 - 두 번째는 데이터 분석환경의 부분으로 일부 플랫폼의 경우 데이터 분석환경을 구축하여 데이터 안심구역과 같이 활용하고 있으나, 이러한 환경구축이 플랫폼의 정책에 따라 미구축된 경우가 있음

- 마지막으로 분쟁해결 등 플랫폼 내의 데이터 판매자와 데이터 구매자간의 분쟁사항 발생 시의 대응으로 대부분의 플랫폼이 분쟁해결에 플랫폼이 개입하지 않는 것을 명시하거나, 면책조항을 명시하고 있음
- 이는 각 플랫폼의 환경에 따라 회원제 및 데이터 분석환경 운영여부와 일반적으로 플랫폼이 분쟁사항에 개입하지 않고 법적 사항을 이해관계자 당사자 간에 해결하도록 하는 정책방향성이 존재하는 것으로 보임
- 또한, 기존의 2019년, 2020년 구축된 빅데이터 플랫폼과 2021년 구축된 빅데이터 플랫폼 간의 표준 이용약관 준수를 차이가 존재
- 기존에 구축된 10개 빅데이터 플랫폼의 경우, 2019년, 2020년에 이용약관을 수립하여 공시하고 있으며, 2021년에 표준 이용약관에 따라 이용약관 개정 미진행 상태
- 2021년 신규 구축된 빅데이터 플랫폼의 경우, 표준 이용약관을 그대로 준용하거나, 일부 수정을 통해 이용약관을 공시
- 16개 빅데이터 플랫폼 중, 일부 플랫폼에 특이사항이 있으며 이에 대한 내용은 아래와 같음
- 지역경제 빅데이터 플랫폼의 경우, 2021년 11월 이용약관 개정을 통해 표준 이용약관을 준수하고, 플랫폼의 현황에 맞춰 최적화가 가장 잘되어 있는 사례로 볼 수 있음
- 해양수산과 헬스케어 빅데이터 플랫폼의 경우, 표준 이용약관을 준용하지 않고 별도의 체계로 이용약관을 운용하고 있으며, 데이터의 유통·거래를 위한 내용이 아닌 별도의 이용약관을 개발하여 운용하고 있음
- 농식품 빅데이터 플랫폼의 경우, 2010년 시행일자 이용약관을 공시하고 있으며, 회원가입 시 표준이용약관을 사용하고 있는 것을 보아 약관시행일을 명확히 할 필요가 있음
- 표준 이용약관은 데이터 3법과 데이터 기본법과 같은 데이터 관련 법·제도의 준수와 빅데이터 플랫폼 및 데이터 센터, 데이터 판매자, 데이터 구매자의 권리를 보호하기 위하여 작성된 것으로 이에 대한 준수필요성과 플랫폼의 환경에 적합한 내용변경을 심도있게 검토가 필요

3.4 빅데이터 플랫폼 이용약관 시사점

- 16개 빅데이터 플랫폼의 이용약관에 대해 표준 이용약관의 준수여부와 동향, 특이사항에 대해 시사점을 도출하여 종합적으로 분석한 결과는 아래와 같음

표 32 빅데이터 플랫폼 이용약관 분석 시사점

구분	시사점	종합
빅데이터 플랫폼 이용약관 준수현황	<ul style="list-style-type: none"> 유료회원(회비회원) 등은 플랫폼의 특성에 의해 미운영인 상황이 존재 	<ul style="list-style-type: none"> 표준 이용약관의 조문을 필수와 선택으로 분류하여 플랫폼의 현황에 최적화 필요

	<ul style="list-style-type: none"> 데이터 분석환경의 경우, 일부 플랫폼이 운영 중이며 필수적인 사항으로 보기 어려움 (공공기관 특성) 	<ul style="list-style-type: none"> 데이터 대금결제 등의 환불조항에서 미분석, 미활용의 경우, 7일 이내라면 환불이 가능해야함 (공정거래법의 소관)
	<ul style="list-style-type: none"> 대부분의 플랫폼이 환불, 분쟁사항에 대해 미개입하거나 면책조항을 운영 	<ul style="list-style-type: none"> 분쟁조정 및 개입에 대한 정책이 재검토 필요함
빅데이터 플랫폼 이용약관 동향	<ul style="list-style-type: none"> 2019년, 2020년 구축된 빅데이터 플랫폼의 경우, 표준 이용약관을 준수하지 않는 경향 2021년 구축된 빅데이터 플랫폼은 표준 이용약관을 일부 개정하여 전면시행하고 있음 	<ul style="list-style-type: none"> 표준 이용약관에 대한 준수여부나 현행화 상태에 대한 진단 필요
빅데이터 플랫폼 이용약관 특이사항	<ul style="list-style-type: none"> 일부 플랫폼은 표준 이용약관을 준수하고 플랫폼의 현황에 적합하게 최적화를 수행 일부 플랫폼의 경우, 데이터 유통·거래체계와 거리가 있는 이용약관 운용 중 일부 플랫폼의 경우, 이용약관이 노후화 되거나 미현행화 상태임 	<ul style="list-style-type: none"> 빅데이터 플랫폼이 데이터 유통·거래체계와 거리가 있는 정책을 수립할 경우, 이에 대한 협의 필요

04 데이터 소유권 및 이용약관 개선



4.1 분석결과 시사점 종합분석

- 데이터 3법 개정과 데이터 기본법의 제정, 해외 데이터 소유권 관련 사례 및 빅데이터 플랫폼 16개 이용약관 현황을 분석한 결과, 각 부분의 시사점과 이를 개선필요사항으로 도출한 내용은 아래와 같음

표 33 분석결과 시사점 종합분석 및 개선 필요사항 도출

구분	시사점	개선필요사항
법·제도 시사점 분석 결과	<ul style="list-style-type: none"> 개인정보보호위원회의 가명정보처리 가이드 준수 명시 필요 표준 이용약관에서 가명정보처리 가이드 준수와 미준수 시 데이터의 적법성 등은 준거법에 의해 처리되는 것을 명시 필요 표준 이용약관에서 준거법으로 데이터 기본법, 개인정보보호법, 저작권법, 공공데이터 제공 및 활성화에 관한 법률, 공정거래법을 기본으로 하고 각 분야에 필요한 법을 추가 가능하도록 개선필요 공정거래위원회의 표준계약서 준수를 권장하도록 명시 필요 (준수권장을 통해 플랫폼 법적책임 해소) 기존의 빅데이터 플랫폼에서 데이터기구를 설치하여 개별 조정하는 것에서 데이터분쟁조정위원회의 조정신청, 데이터기본법 위반에 의한 손해배상 가능 내용을 명시 필요 	<ul style="list-style-type: none"> 개인정보보호위원회의 가명정보처리 가이드 준수와 플랫폼에 데이터 등록 시, 가명정보처리 의무 및 책임은 데이터 판매자에게 있음을 명시
해외 데이터 소유권 관련 사례 분석 결과	<ul style="list-style-type: none"> 현행 법·제도 내에서 데이터 소유권 보호의 기본적인 사항은 데이터 판매자와 데이터 구매자 간의 공정한 계약 체결이 필요 	<ul style="list-style-type: none"> 데이터 기본법 제17조 제4항 3, 제21조에 근거하여 데이터 유통·거래 시 표준계약서 사용권장 빅데이터 플랫폼 내에서 플랫폼 별로 분쟁조정기구 설치 및 개입을 데이터 기본법 제34조, 제35조에 근거하여 데이터 분쟁조정위원회 회부로 개정

과	<ul style="list-style-type: none"> 빅데이터 플랫폼의 이용약관에 현행 법·제도에 대한 준수를 의무화 필요 	<ul style="list-style-type: none"> 데이터 기본법 및 공정거래법 등 관련 법률에 근거하여 필수적으로 필요한 내용을 이용약관 명시
	<ul style="list-style-type: none"> 가명정보 처리 등 개인정보보호 관련은 개인정보보호위원회 및 데이터 3법, 가명정보처리 가이드 준수가 필요 	
	<ul style="list-style-type: none"> 공정거래법은 데이터 시장과 관련된 내용을 포괄하므로 준거법으로 명시필요 	
	<ul style="list-style-type: none"> 데이터 추출 및 활용은 공정이용의 범주로 볼 수 있으나, 데이터 판매자와 구매자 간의 이슈사항은 상호간의 민사소송이므로 빅데이터 플랫폼이 개입할 필요가 없음 	<ul style="list-style-type: none"> 유료회원제나 분석환경 등 빅데이터 플랫폼 별로 현행화가 필요한 부분을 선택사항으로 표준 이용약관에서 권장
빅데이터 플랫폼 현황분석 결과	<ul style="list-style-type: none"> 표준 이용약관의 조문을 필수와 선택으로 분류하여 플랫폼의 현황에 최적화 필요 	<ul style="list-style-type: none"> 데이터 3법 및 데이터기본법, 저작권법, 공공데이터 제공 및 활성화에 관한 법률, 공정거래법 등 준거법을 정비
	<ul style="list-style-type: none"> 데이터 대금결제 등의 환불조항에서 미분석, 미활용의 경우, 7일 이내라면 환불이 가능해야함 (공정거래법의 소관) 	
	<ul style="list-style-type: none"> 분쟁조정 및 개입에 대한 정책이 재검토 필요함 	
	<ul style="list-style-type: none"> 표준 이용약관에 대한 준수여부나 현행화 상태에 대한 진단 필요 빅데이터 플랫폼이 데이터 유통·거래체제와 거리가 있는 정책을 수립할 경우, 이에 대한 협의 필요 	<ul style="list-style-type: none"> 데이터 자산 및 빅데이터 플랫폼, 데이터 판매자, 데이터 구매자 권리보호를 위하여 표준 이용약관의 진당대상화 필요

- 주요 개선필요사항은 데이터 3법 개정 및 데이터 기본법 제정으로 인해 준거법 및 조문에 대한 현행화가 필요하며, 16개 빅데이터 플랫폼의 표준 이용약관 준수현황에 보완이 필요한 부분이 있으므로, 이에 대한 관리체계 개선이 필요

표 34 표준 이용약관 개선필요사항 정리

순번	구분	개선필요사항
1	가명정보처리 책무 명시	<ul style="list-style-type: none"> 개인정보보호위원회 가명정보처리 가이드 준수 필요 가명정보처리 가이드 준수책무는 데이터 판매자에 있음
2	표준계약서 사용권장 명시	<ul style="list-style-type: none"> 데이터 기본법 제17조 제4항 3, 제21조에 근거하여 데이터 유통·거래 시 표준계약서 사용권장 필요
3	분쟁조정위원회 조정 명시	<ul style="list-style-type: none"> 기존의 플랫폼별 분쟁조정기구 설치 및 개입을 데이터 기본법 제34조, 제35조에 근거하여 데이터 분쟁조정위원회 회부로 변경 필요

4	표준 이용약관 필수사항 지정	<ul style="list-style-type: none"> 데이터 기본법 및 공정거래법 등 관련 법률에 근거하여 필수적으로 이용약관 명시 필요
5	표준이용약관 선택사항 지정	<ul style="list-style-type: none"> 유료회원제나 분석환경 등 빅데이터 플랫폼 별로 현행화가 필요한 부분을 선택사항으로 표준 이용약관에서 권장 필요
6	관할권 및 준거법 정비	<ul style="list-style-type: none"> 데이터 3법 및 데이터기본법, 저작권법, 공공데이터 제공 및 활성화에 관한 법률, 공정거래법 등 준거법을 정비 필요
7	표준 이용약관 진당대상화	<ul style="list-style-type: none"> 데이터 자산 및 빅데이터 플랫폼, 데이터 판매자, 데이터 구매자 권리보호를 위하여 표준 이용약관의 진당대상화 필요

- 개선필요사항으로 표준 이용약관 조문 자체의 내용은 아니나, 빅데이터 플랫폼 중 일부 플랫폼은 표준 이용약관을 사용하지 않거나, 현행화가 없이 전문 그대로를 전사하는 사례가 존재
- 일부 플랫폼은 표준 이용약관의 제정이나 개선과 상관없이 이용약관을 개정하지 않고 그대로 사용하거나, 과거의 이용약관을 그대로 전용하는 등의 현황이 있음
 - 신규 플랫폼의 경우, 표준 이용약관을 그대로 전사하여 사용하거나, 일부 조문을 제외하는 수준에서 사용하고 있으나, 플랫폼 별 현황에 적합하게 일부 조문을 개정할 필요성이 있음
- 빅데이터 플랫폼 별로 플랫폼의 현황과 운영정책에 의해 이용약관의 현행화 필요성이 있으므로 이를 주기적으로 진단하고 개선사항을 점검할 진단이 필요

4.2 표준 이용약관 개선사항

- 기존의 표준이용약관은 3장 17조의 형태로 구성되어 있으며, 개선필요사항을 반영하여 조문 내용을 개선한 내용은 아래와 같음

표 35 표준 이용약관 개선사항

이용약관 개선(안)	기존 이용약관	필수유무	개선내용/비고
제1조 목적	제1조 목적	필수	
제2조 정의	제2조 정의	필수	<ul style="list-style-type: none"> 플랫폼 별로 변경가능 조건
제3조 약관 등의 명시·개정	제3조 약관 등의 명시·개정	필수	<ul style="list-style-type: none"> 근거 법률 추가보완
제4조 서비스의 제공 변경	제4조 서비스의 제공 변경	필수	<ul style="list-style-type: none"> 빅데이터 플랫폼 구축사항의 목적에 근거한 필수사항 (데이터 유통·거래 기능은 필수기능으로 볼 수 있음)
제5조 서비스의 중단	제5조 서비스의 중단	필수	<ul style="list-style-type: none"> 전자상거래 등에서의 소비자 보호에 관한 법률에 근거한 필수사항
제6조 서비스 이용약관신청	제6조 서비스 이용약관신청	필수	

이용약관 개선(안)	기존 이용약관	필수유무	개선내용/비고
제7조 서비스가입 신청승낙	제7조 서비스가입 신청승낙	필수	
제8조 서비스이용 계약종료	제8조 서비스이용 계약종료	필수	• 해당 조문 제5항은 선택사항
제9조 회비 환불정책	제9조 회비 환불정책	선택	• 유료회원제 미운영 시, 표준 이용약관 준수불필요
제10조 회원에 대한 통지	제10조 회원에 대한 통지	선택	•
제11조 개인정보 제공 동의	제11조 개인정보 제공 동의	필수	• 별도의 개인정보활용동의에 대한 약관 및 동의절차 존재 시 이용약관에 기포함으로 분류 • 데이터 3법 개정에 따른 소관법률 이전 반영
제12조 데이터의 가명정보처리 및 비식별성 확보	-	필수	• 가명정보처리 가이드 준용 및 비식별성 확보에 대한 조치 책무 명시
제13조 유료 데이터상품의 구매	제12조 유료 데이터상품의 구매	필수	
제14조 표준계약서 준용권장	-	필수	• 데이터 기본법 제17조, 제21조에 근거하여 표준계약서 준용
제15조 매매부적합 상품제한	-	필수	• 플랫폼의 운영정책에 의해 법적으로 사유가 있는 상품의 매매제한 조치
제16조 데이터상품 분석환경	제13조 데이터상품 분석환경	선택	• 플랫폼에 따라 분석환경을 운영하지 않을 수 있으므로 선택사항으로 분류
제17조 데이터 상품 환불	제14조 데이터상품 환불	필수	• 원칙적으로 환불불가이나, 7일 이내 미분석, 미활용 시 환불이 가능해야함 (공정거래법 저촉사항 회피)
제18조 저작권의 귀속 등	제15조 저작권의 귀속 등	선택	• 저작권 귀속은 운영정책에 속하므로 선택사항으로 분류
제19조 데이터분쟁조정위원회 조정신청	제16조 분쟁해결	필수	• 기존의 분쟁개입 및 피해보상처리기구 운영을 데이터 기본법 제34조, 제35조의 데이터 분쟁조정위원회 신청으로 변경
제20조 재판권 및 준거법	제17조 재판권 및 준거법	필수	• 개인정보보호법, 데이터 기본법, 공정거래법 준거법 보완

이용약관 개선(안)	기존 이용약관	필수유무	개선내용/비고
약관시행·개정일	약관시행·개정일	필수	• 주기적으로 진단 및 제·개정 확인필요

- 2020년의 표준 이용약관은 전자상거래법에 근거하여 작성된 이용약관으로 서비스에 관련한 내용은 전자소비자법 등 관련 법률이 존재하여 소비자 권익보호를 위해 필수적으로 필요한 사항임
- 또한, 별도로 개인정보활용동의 및 분석환경 이용 등에 대한 동의절차를 운영하는 플랫폼도 존재하나, 표준 이용약관은 빅데이터 플랫폼 운영정책 전반에 대한 기본사항을 정의하는 것이므로 이용약관에 명시 필요

4.3 표준 이용약관 개선(안)

제1장 총칙

제1조 [목적]

이 약관은 ○○회사가 운영하는 ○○빅데이터 플랫폼 (이하“플랫폼”이라한다)에서 제공하는 서비스를 이용함에 있어서 플랫폼과 이용자의 권리·의무 및 책임사항을 규정함을 목적으로 합니다.

제2조 [정의]⑧

이 약관에서 사용하는 용어의 정의는 다음과 같습니다.

- ① “플랫폼”이란 ○○ 회사가 데이터, UI, API 등을 (이하 “데이터 등”이라 함)을 이용자에게 제공하기 위하여, 데이터를 수집, 처리, 저장, 분석 및 시각화 등을 통하여 유료 및 무료회원들에게 제반을 혜택을 제공하는 시스템을 말하며, 아울러 플랫폼을 운영하는 사업자의 의미로도 사용합니다.
- ② “이용자”란 “플랫폼”에 접속하여 이 약관에 따라 “플랫폼”이 제공하는 서비스를 받는 회원 및 비회원을 말합니다.
- ③ “회원”이라 함은 “플랫폼”에 회원등록을 한 자로서, 계속적으로 “플랫폼”이 제공하는 서비스를 이용할 수 있는 자를 말합니다.
- ④ “비회원”이라 함은 회원에 가입하지 않고 “플랫폼”이 제공하는 서비스를 이용하는 자를 말합니다.
- ⑤ “아이디(ID)”라 함은 “회원”의 식별과 서비스이용을 위하여 “회원”이 정하고 “회사”가 승인하는 문자 또는 숫자의 조합을 말합니다.
- ⑥ “비밀번호(PASSWORD)”라 함은 “회원”이 부여받은 “아이디”와 일치되는 “회원”임을 확인하고 비밀보호를 위해 “회원” 자신이 정한 문자 또는 숫자의 조합을 말합니다.

8) 빅데이터 플랫폼의 현황에 적합하게 변경가능

제3조 [약관 등의 명시와 설명 및 개정]

- ① “플랫폼”은 이 약관의 내용과 상호 및 대표자 성명, 영업소 소재지 주소(소비자의 불만을 처리할 수 있는 곳의 주소를 포함), 전화번호·모사전송번호·전자우편주소, 사업자등록번호, 통신판매업 신고번호, 개인정보관리책임자등을 이용자가 쉽게 알 수 있도록 “플랫폼”의 초기 서비스화면(전면)에 게시합니다. 다만, 약관의 내용은 이용자가 연결화면을 통하여 볼 수 있도록 할 수 있습니다.
- ② “플랫폼”은 이용자가 약관에 동의하기에 앞서 약관에 정하여져 있는 내용 중 청약철회·환불조건 등과 같은 중요한 내용을 이용자가 이해할 수 있도록 별도의 연결화면 또는 팝업화면 등을 제공하여 이용자의 확인을 구하여야 합니다.
- ③ “플랫폼”은 「전자상거래 등에서의 소비자보호에 관한 법률」, 「약관의 규제에 관한 법률」, 「전자문서 및 전자거래기본법」, 「전자금융거래법」, 「전자서명법」, 「방문판매 등에 관한 법률」, 「소비자기본법」, 「개인정보보호법」, 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」, 「신용정보의 이용 및 보호에 관한 법률」, 「데이터 산업진흥 및 이용촉진에 관한 법률」, 「독점규제 및 공정거래에 관한 법률」⁹⁾ 등 관련 법을 위배하지 않는 범위에서 이 약관을 개정할 수 있습니다.
- ④ “플랫폼”이 약관을 개정할 경우에는 적용일자 및 개정사유를 명시하여 현행약관과 함께 플랫폼 초기화면에 그 적용일자 7일 이전부터 적용일자 전일까지 공지합니다. 다만, 이용자에게 불리하게 약관내용을 변경하는 경우에는 최소한 30일 이상의 사전 유예기간을 두고 공지합니다. 이 경우 “플랫폼”은 개정 전 내용과 개정 후 내용을 명확하게 비교하여 이용자가 알기 쉽도록 표시합니다.
- ⑤ “플랫폼”이 약관을 개정할 경우에는 그 개정약관은 그 적용일자 이후에 체결되는 계약에만 적용되고 그 이전에 이미 체결된 계약에 대해서는 개정 전의 약관조항이 그대로 적용됩니다. 다만 이미 계약을 체결한 이용자가 개정약관 조항의 적용을 받기를 원하는 뜻을 제3항에 의한 개정약관의 공지기간 내에 “플랫폼”에 송신하여 “플랫폼”의 동의를 받은 경우에는 개정약관 조항이 적용됩니다.
- ⑥ 이 약관에서 정하지 아니한 사항과 이 약관의 해석에 관하여는 전자상거래 등에서의 소비자보호에 관한 법률, 약관의 규제 등에 관한 법률, 공정거래위원회가 정하는 전자상거래 등에서의 소비자 보호지침 및 관계법령 또는 상관례에 따릅니다.

제4조 [서비스의 제공 및 변경]

- ① “플랫폼”은 다음과 같은 업무를 수행합니다.
 1. 데이터 또는 분석환경에 대한 정보 제공 및 무료, 유료 구매 계약의 체결
 2. 데이터 거래 및 활용을 위한 플랫폼 운영 ¹⁰⁾
 3. 데이터 상품 분석 환경 제공 서비스 : 유료회원에게는 구매한 데이터를 분석하기 위한

9) 데이터 3법 및 데이터 기본법, 공정거래법 등 유관 법률 보완

10) 제4조 제2항 1, 2호는 필수사항

분석 환경을 제공하며, 사용시에 별도의 추가 비용 발생

4. 빅데이터, AI 관련 교육 및 컨퍼런스 기회 제공 ¹¹⁾
5. 기타 “플랫폼”이 정하는 업무
- ② “플랫폼”의 기술적 사양의 변경 등의 경우에는 장차 체결되는 계약에 의해 제공할 데이터 및 분석 환경 상품 정보와 관련 된 내용을 변경할 수 있습니다. 이 경우에는 변경된 내용 및 제공일자를 명시하여 현재의 데이터 및 분석 환경 상품 내용을 게시한 곳에 즉시 공지합니다.

제5조 [서비스의 중단] ¹²⁾

- ① “플랫폼”은 컴퓨터 등 정보통신설비의 보수점검·교체 및 고장, 통신의 두절 등의 사유가 발생한 경우에는 서비스의 제공을 일시적으로 중단할 수 있습니다.
- ② “플랫폼”은 제1항의 사유로 서비스의 제공이 일시적으로 중단됨으로 인하여 이용자 또는 제3자가 입은 손해에 대하여 배상합니다. 단, “플랫폼”이 고의 또는 과실이 없음을 입증하는 경우에는 그러하지 아니합니다.

제2장 서비스 이용약관

제6조 [서비스 이용약관 및 가입신청]

- ① 플랫폼 서비스를 이용하고자 하는 자는 “플랫폼”이 정한 가입 양식에 따라 회원정보를 기입한 후 이 약관에 동의한다는 의사표시를 함으로서 회원가입을 신청합니다. 가입신청 시 필요한 정보는 사실대로 기재하여야 하며 정해진 인증절차를 준수해야 합니다.
- ② 회원가입신청서에는 다음 사항을 기재해야 합니다. 1호 내지 3호의 사항은 필수사항이며, 그 외의 사항은 선택사항입니다.
 1. “회원”의 성명과 주민등록번호 또는 인터넷상 개인식별번호
 2. “아이디”와 “비밀번호”
 3. 전자우편주소
 4. 이용하려는 “데이터 또는 분석환경 서비스”의 종류
 5. 기타 “플랫폼”이 필요하다고 인정하는 사항
- ③ 유료회원은 “플랫폼”에서 정한 연회비 또는 월회비 납입을 통해 이용신청이 완료 됩니다.

제7조 [서비스가입 신청의 승낙]

- ① “플랫폼”은 “이용자”가 약관의 내용에 대해서 동의를 하고, 서비스 가입신청에 대하여 승낙함을 원칙으로 합니다. 단, 무료 회원은 이메일 인증완료 시 서비스 가입신청에 대한 승낙이 완료 됩니다.
- ② “플랫폼”은 다음 각 호에 해당 하는 신청에 대하여는 승낙을 하지 않을 수 있습니다.

11) 제4조 제2항 3, 4, 5호는 선택사항

12) 「전자상거래법 등에서 소비자 보호에 관한 법률」에 근거하여, 플랫폼의 과실여부 및 배상책임 필수사항

1. 실명이 아니거나 타인의 명의를 이용한 경우
 2. 허위의 정보를 기재하거나, 플랫폼에서 제시하는 내용을 기재하지 않은 경우
 3. 가입신청자의 귀책사유로 인하여 승인이 불가능한 경우
 4. “플랫폼”으로부터 자격정지 조치 등을 받은 회원이 그 조치 기간 중에 이용계약을 임의 해지하고 재이용신청을 하는 경우
 5. 기타 본 약관에 위배되거나 위법 또는 부당한 이용신청임이 확인 된 경우 및 플랫폼의 합리적 판단에 의하여 필요하다고 인정되는 경우
- ③ 제 1항에 따른 신청에 있어서 “플랫폼”은 전문기관을 통한 실명확인 및 본인인증, 계좌검증 및 사업자등록 확인 등을 요청할 수 있습니다.
- ④ 서비스 이용계약의 성립시기는 제 1항에 따른 “플랫폼”의 승낙이 완료되고 해당 내용을 회원에게 발송한 시점으로 합니다.

제8조 [서비스이용 계약의 종료]

- ① 회원은 “플랫폼”에 언제든지 탈퇴를 요청할 수 있으며 “플랫폼”은 즉시 회원탈퇴를 처리합니다.
- ② 회원이 다음 각 호의 사유에 해당하는 경우, “플랫폼”은 회원자격을 제한 및 정지시킬 수 있습니다.
 1. 가입 신청 시에 허위 내용을 등록한 경우
 2. “플랫폼”을 이용하여 구입한 데이터 상품 등의 대금, 기타 “플랫폼” 이용에 관련하여 회원이 부담하는 채무를 기일에 지급하지 않는 경우
 3. 다른 사람의 “플랫폼” 이용을 방해하거나 그 정보를 도용하는 등 전자상거래 질서를 위협하는 경우
 4. “플랫폼”을 이용하여 법령 또는 이 약관이 금지하거나 공서양속에 반하는 행위를 하는 경우
- ③ “플랫폼”이 회원 자격을 제한·정지 시킨 후, 동일한 행위가 2회 이상 반복되거나 30일 이내에 그 사유가 시정되지 아니하는 경우 “플랫폼”은 회원자격을 상실시킬 수 있습니다.
- ④ “플랫폼”이 회원자격을 상실시키는 경우에는 회원등록을 말소합니다.
이 경우 회원에게 이를 통지하고, 회원등록 말소 전에 최소한 30일 이상의 기간을 정하여 소명할 기회를 부여합니다.
- ⑤ 유료회원의 이용기간 만료¹³⁾
 1. 유료개인회원 : 이용기간 만료 시, 이용 기간을 연장하지 않은 경우 유료회원의 계정은 무료회원으로 변경 됩니다. 구매 하신 데이터는 사용기한(1년) 내 사용이 가능하며, 분석 환경 사용은 중지됩니다.
 2. 유료법인회원 : 이용기간 만료 시, 이용 기간을 연장하지 않은 경우 회원에게 부여 된 관리자 및 판매자의 권한이 무료회원과 동일한 권한으로 변경됩니다. 이에 따라 데이터

13) 제8조 제5항, 제6항, 제7항은 빅데이터 플랫폼 운영정책에 의해 선택사항

판매, 분석 환경 이용 등의 권한이 제한됩니다.

- ⑥ 이용기간 만료 후 갱신
 1. 유료개인회원: 이용기간이 만료된 회원이 유료회원으로 갱신을 요청하여 회비를 결제하면 결제 확인 후 유료회원으로 갱신됩니다.
 2. 유료법인회원: 이용기간이 만료된 회원이 관리자 계정으로 이용기간이 만료된 때로부터 1년 내에 회비를 결제하면 결제 확인 후 이용 권한이 회복됩니다.
- ⑦ 휴면 계정
 1. 회원 이용기간 만료 후 1년 이상 플랫폼에 로그인 하지 않을 경우에는 회원의 계정은 휴면 계정으로 전환 되며, 휴면 계정의 개인정보는 다른 계정과 별도로 관리됩니다.
 2. 판매자로 활동하던 회원의 계정이 휴면계정으로 전환 될 경우, 구매자가 없는 데이터 상품은 즉시 삭제됩니다.

제9조 [회비 환불 정책]¹⁴⁾

- ① “플랫폼”은 회비를 결제한 회원에 대하여, 다음과 같이 환불 조치할 수 있습니다.
 1. “플랫폼”의 귀책사유로 인한 환불의 경우: 회원이 서비스 이용으로부터 얻은 이익에 해당하는 금액을 아래와 같이 공제하고 환불할 수 있습니다.
가. 데이터 상품 분석 환경 제공 서비스: 공제액 = 기 이용일수 * 일일 이용대금 * 단, 유료회원으로 가입시 제공한 분석 환경을 위한 크레딧은 소멸함
 2. 회원의 귀책사유로 인한 환불의 경우: 가목에 따라 회원이 서비스 이용으로부터 얻은 이익에 해당하는 금액을 공제한 잔여 금액에서 위약금(10%) 및 지불수단별 결제대행수수료를 제외한 금액을 환불할 수 있습니다.
- ② 다음에 해당하는 경우 회비의 환불이 불가합니다.
 1. “플랫폼”이 발행한 쿠폰을 통해 유료회원의 자격을 얻은 경우
 2. 기타 회원이 직접 결제하지 않은 경우
- ③ “플랫폼”은 회원이 신용카드 또는 전자화폐 등의 결제수단으로 회비를 지불한 경우에는 지체없이 당해 결제수단을 제공한 사업자로 하여금 이용요금의 청구를 정지 또는 취소하도록 요청합니다. 다만, 제1항 및 제2항의 금액공제가 필요한 경우에는 그러하지 아니할 수 있습니다.
- ④ 기결제된 결제금액 또는 결제취소가 불가능한 결제금액의 환불 신청 시는 해당 요금의 정상납입 확인 후 환불조치가 가능하며, 결제취소기간 내 환불 의사를 밝힌 경우에는 결제취소처리를 우선적으로 고려하여 처리할 수 있습니다.
- ⑤ 회비를 환불 받을 금융기관 계좌의 예금주 명의로 “플랫폼”에 등록된 회원 ID의 명의가 동일한 경우에만 제1항 및 제2항에 따른 금액의 입금이 가능합니다. 단 해외금융기관의 계좌인 경우에는 환불이 불가능할 수 있습니다.¹⁵⁾
- ⑥ 개인정보 도용 및 결제정보 도용 또는 부정결제 등으로 인한 경우에는 회비의 환불이

14) 회비 환불 정책은 빅데이터 플랫폼이 유료회원제 미운영 시 선택사항

15) 제9조 제5항은 빅데이터 플랫폼의 운영정책에 의해 선택사항

불가하며, 환불받을 금액에 관하여 정당한 권한이 있음을 수사기관 또는 법원의 재판 등을 통한 입증이 필요합니다.

- ⑦ “플랫폼”은 환불요청을 접수한 날로부터 10 영업일 이내에 제1항 및 제2항에 따른 금액을 반환하는 것을 원칙으로 합니다. 다만, 후불제 결제수단(신용카드, 휴대전화) 등을 이용해서 결제하신 경우, 결제내역의 입금사실을 확인하는데 필요한 기간만큼 환불처리기간이 추가로 소요될 수 있습니다.
- ⑧ 회비의 납부여부의 확인이 불가능한 경우 환불이 불가능할 수 있으므로, 결제와 관련된 납부영수증 등의 증빙자료 보관에 유념하시기 바랍니다.
- ⑨ 회원탈퇴 및 환불요청 시 회원님의 오기(誤記)등의 귀책사유로 인해 발생하는 불이익에 대해서 “플랫폼”은 책임을 지지 않습니다.

제10조 [회원제에 대한 통지]

- ① “플랫폼”이 회원에 대한 통지를 하는 경우, 회원이 “플랫폼”과 미리 약정하여 지정한 전자우편 주소로 할 수 있습니다.
- ② “플랫폼”은 불특정다수 회원에 대한 통지의 경우 1주일이상 “플랫폼” 게시판에 게시함으로써 개별 통지에 갈음할 수 있습니다. 다만, 회원 본인의 거래와 관련하여 중대한 영향을 미치는 사항에 대하여는 개별통지 합니다.

제11조 [개인정보 제공 동의 등]

- ① “플랫폼”이 제3자에게 구매자 개인정보를 제공할 필요가 있는 경우, 개인정보를 제공받는 자, 개인정보를 제공받는 자의 개인정보 이용목적, 제공하는 개인정보의 항목, 개인정보를 제공받는 자의 개인정보 보유 및 이용기간을 구매자에게 알리고 동의를 받아야 합니다.
- ② “플랫폼”이 제3자에게 구매자의 개인정보를 취급할 수 있도록 업무를 위탁하는 경우에는 개인정보 취급위탁을 받는 자, 개인정보 취급위탁을 하는 업무의 내용을 구매자에게 알리고 동의를 받아야 합니다.
- ③ 다만, 서비스제공에 관한 계약이행을 위해 필요하고 구매자의 편의증진과 관련된 경우에는 「개인정보보호법」¹⁶⁾에서 정하고 있는 방법으로 개인정보 취급방침을 통해 알림으로써 고지절차와 동의절차를 거치지 않아도 됩니다.

제3장 데이터상품 이용약관

제12조 [데이터의 가명정보처리 및 비식별성 확보]

- ① 회원은 플랫폼에서 거래되는 데이터 상품에 특정 개인의 개인정보가 포함되어 있는 경우, 이를 가명정보처리 및 비식별화 조치한 다음에 거래하여야 합니다.
- ② 회원은 플랫폼에서 거래되는 데이터 상품에 대하여 다음과 같은 보호조치를 취하여야

합니다.

1. 관리적 보호조치

- 가. 상품의 비식별화 조치 관리를 위한 담당자 지정
- 나. 비식별화 조치된 상품의 대장관리
- 다. 비식별화 조치된 상품의 재식별 시 대응계획 수립
- 라. 기타 필요한 관리적 보호 조치

2. 기술적 보호조치

- 가. 비식별화 조치된 상품에 대한 접근권한 관리 및 접근통제
- 나. 비식별화된 데이터에 대한 접속기록 관리
- 다. 악성코드 방지 등을 위한 보안프로그램 설치 및 운영
- 라. 기타 필요한 기술적 보호 조치

- ③ 회원은 「개인정보보호법」과 개인정보보호위원회의 고시, 지침 등에 근거하여 가명정보처리 및 비식별화 조치를 수행해야 합니다.

제13조 [유료 데이터상품의 구매 등]

- ① 회원이 유료데이터 상품을 이용하기 위해서는 해당 유료 데이터 상품의 대금을 결제해야 하며, 이와 관련하여 회원이 입력한 정보 및 그 정보와 관련하여 발생하는 책임과 불이익은 전적으로 회원이 부담합니다.
- ② 회원이 유료 데이터상품을 주문한 후 일정 기간 내에 대금을 결제하지 않을 경우 “플랫폼”은 당해 주문을 회원의 동의 없이 취소할 수 있습니다.
- ③ 회원은 상품 주문 정보를 마이페이지를 통해 확인 할 수 있으며, “플랫폼”은 데이터 상품의 주문의 취소 방법 및 절차를 안내합니다.
- ④ “플랫폼”은 회원이 대금 결제 시 사용한 결제수단에 대해 정당한 사용권한을 가지고 있는지의 여부를 확인 할 수 있으며, 이에 대한 확인이 완료 될 때까지 거래 진행을 중지하거나, 확인이 불가능한 거래를 취소할 수 있습니다.
- ⑤ 회원이 실제로 결제하는 금액은 판매회원이 정한 공급원가(실구매액)에 플랫폼 수수료 및 전자결제서비스 수수료가 포함된 금액이며 구매회원에게 발행되는 구매증빙서(세금계산서, 신용카드매출전표, 현금영수증 등)은 실구매액으로 발행됩니다.
- ⑥ “플랫폼”에서 구매한 데이터상품에 대한 대금지급방법은 다음 각 호의 방법중 가용한 방법으로 할 수 있습니다.
 1. 폰뱅킹, 인터넷뱅킹, 메일 뱅킹 등의 각종 계좌이체
 2. 선불카드, 직불카드, 신용카드 등의 각종 카드 결제
 3. 온라인무통장입금
 4. 전자화폐에 의한 결제
 5. 수령 시 대금지급
 6. 마일리지 등 “플랫폼”이 지급한 포인트에 의한 결제
 7. “플랫폼”과 계약을 맺었거나 “플랫폼”이 인정한 상품권에 의한 결제

16) 데이터 3법 개정으로 인해 정보통신망법에서 개인정보보호법으로 소관 법률 이전

8. 기타 전자적 지급 방법에 의한 대금 지급 등

- ⑦ “플랫폼”은 구매한 데이터 상품을 이용할 수 있는 분석환경을 지원합니다. 분석환경을 이용하기 위해서는 별도의 이용신청을 하여야 하며, 분석환경 이용시 추가적인 이용수수료가 발생할 수 있습니다.

제14조 [표준계약서 준용권장]

- ① 회원은 플랫폼에서 데이터 상품을 거래할 때, 「데이터 산업진흥 및 이용촉진에 관한 법률」제21조 표준계약서에 근거하여 과학기술정보통신부 고시의 표준계약서를 준용할 수 있습니다.
- ② 회원은 플랫폼에서 데이터 상품을 거래할 때, 「데이터 산업진흥 및 이용촉진에 관한 법률」 제12조 데이터 자산의 보호에 근거하여 권리를 보호 받을 수 있습니다.

제15조 [매매부적합 상품제한]

- ① 회원은 매매부적합 상품을 판매하여서는 안되며, 매매부적합 상품은 매매불가상품과 매매제한상품으로 구분합니다.
1. 매매불가상품은 개인정보, 지식재산권 등 권리침해상품 등 관련법령상 판매 또는 유통이 불가능한 상품을 말합니다.
 2. 매매제한상품은 상품의 판매방식, 판매장소 또는 판매 대상자 등에 대한 법적 제한이 있는 상품, 소비자에게 피해가 발생할 염려가 있는 상품, 기타 사회통념상 매매에 제한이 있거나 운영기관의 정책에 의하여 매매가 제한되는 상품, 또는 상품에 음란물 등에 대한 법적 제한이 있는 상품, 기타 사회통념상 매매에 부적합하거나 운영기관의 정책에 의하여 매매에 부적합한 상품을 말합니다.
- ② “플랫폼”은 매매부적합 상품이 발견된 경우, 사전 통보 없이 당해 상품의 판매를 중지시킬 수 있으며, 당해 상품이 이미 판매된 경우 그 거래를 취소할 수 있습니다.

제16조 [데이터상품 및 분석환경의 이용 등]¹⁷⁾

- ① “플랫폼”은 판매회원이 승인 및 취소에 대한 내역을 구매회원에게 통보합니다.
- ② “플랫폼”에서 구매한 데이터는 판매회원이 명시한 이용범위 내에서 활용이 가능합니다. 이를 위반시 발생하는 손해에 대해서는 이용회원이 책임집니다. “플랫폼”에서 거래되는 데이터상품의 경우에 아래 항목을 공통 이용범위로 합니다.
1. 구매한 데이터 상품을 재판매하는 행위 금지
 2. 구매한 데이터 상품에 대한 개인정보 식별 시도 금지
 3. 구매한 데이터 상품을 활용·가공 및 타 데이터와 결합하여 새로운 데이터 상품을 만드는 행위 금지. 단, 구매한 데이터 상품의 판매회원과 협의된 경우는 제외함
- ③ “플랫폼”에서 제공하는 분석환경은 이용신청 범위 안에서 활용 가능합니다. 단, 분석환경에서 사용되는 데이터의 개인정보보호 등 조치에 대한 책임은 이용자에게

17) 빅데이터 플랫폼에서 분석환경 미운영 시, 선택사항

있습니다.

- ④ 구매 데이터 상품의 전송과 관련하여 판매회원과 이용회원 사이에 발생한 분쟁은 당사자들 간의 해결을 원칙으로 합니다.

제17조 [데이터상품 구매의 취소 및 환불]¹⁸⁾

- ① “플랫폼”을 통해 거래되는 상품은 디지털화된 상품의 특성상 원칙적으로 청약 철회, 반품, 취소, 환불이 제약됩니다. 다만 아래의 경우에는 예외로 합니다.
1. 판매회원의 구매 승인이 필요한 데이터 상품에 대하여 판매회원이 7 일 내 승인을 하지 아니하여 결제가 자동 취소되는 경우
 2. 구매회원이 상품 구매 후 상품 다운로드를 하지 않거나 분석환경에서 이용하지 않은 상태에서 구매 완료일로부터 7 일 내 환불을 신청한 경우

제18조 [저작권의 귀속 및 이용제한]

- ① “플랫폼”이 작성한 저작물에 대한 저작권 기타 지적재산권은 “플랫폼”에 귀속합니다.
- ② 이용자는 “플랫폼”을 이용함으로써 얻은 정보 중 “플랫폼”에게 지적재산권이 귀속된 정보를 사전 승낙 없이 복제, 송신, 출판, 배포, 방송 기타 방법에 의하여 영리목적으로 이용하거나 제3자에게 이용하게 하여서는 안됩니다.

제19조 [데이터분쟁조정위원회 조정신청]

- ① 회원 간의 데이터 상품 및 용역 등에 관한 정보 또는 유통·거래에 관하여 분쟁이 발생한 경우, 「데이터 산업진흥 및 이용촉진에 관한 법률」제34조에 근거하여 데이터분쟁조정위원회에 조정신청을 할 수 있습니다.
- ② “플랫폼”은 회원 간의 분쟁에 개입하지 않으며 그 분쟁의 결과로 인한 모든 책임은 회원이 부담하여야 합니다.
- ③ “플랫폼”은 회원 간의 분쟁과 관련하여 “플랫폼”이 제 3자에게 손해를 배상하거나 기타 비용을 지출한 경우 “플랫폼”은 귀책사유 있는 자에게 구상권을 행사 할 수 있습니다.
- ④ “플랫폼”과 회원 간에 발생한 분쟁과 관련하여 이용자의 피해구제신청이 있는 경우에는 공정거래위원회 또는 데이터분쟁조정위원회의 조정에 따를 수 있습니다.

제20조 [재판권 및 준거법]

- ① “플랫폼”및 이용자 간에 발생한 전자상거래 분쟁에 관한 소송은 제소 당시의 이용자의 주소에 의하고, 주소가 없는 경우에는 거소를 관할하는 지방법원의 전속관할로 합니다. 다만, 제소 당시 이용자의 주소 또는 거소가 분명하지 않거나 외국 거주자의 경우에는 민사소송법상의 관할법원에 제기합니다.
- ② “플랫폼”및 회원 간의 분쟁에 관한 소송 등은 공공데이터의 제공거부 및 제공중단 관련한 분쟁은 「공공데이터의 제공 및 이용 활성화에 관한 법률」 및 「데이터기반행정 활성화에

18) 「독점규제 및 공정거래에 관한 법률」에 근거하여 특정사유에 한해 환불가능 필요

관한 법률], 개인정보와 관련한 분쟁은 「개인정보보호법」, 저작권에 관련한 분쟁은 「저작권법」, 공정거래 및 시장행위에 관련한 분쟁은 「독점규제 및 공정거래 촉진에 관한 법률」, 데이터 자산보호에 관련한 분쟁은 「데이터 산업진흥 및 이용촉진에 관한 법률」을 준거법으로 합니다.

- ③ “플랫폼” 및 회원 간의 이용계약 및 회원 상호간의 분쟁에 대해 “플랫폼”을 당사자로 하는 소송이 제기될 경우에는 “플랫폼”의 본점 소재지를 관할하는 법원을 합의관할법원으로 정하기로 합니다.

빅데이터플랫폼 및 센터 데이터 품질관리 가이드

발 행 2021년 12월

발 행 인 문용식

기획/작성 한국지능정보사회진흥원 빅데이터플랫폼팀
이용진 단장, 손기문 팀장, 유시형 수석, 김은영 연구원, ㈜씨에이에스

문 의 ysh@nia.or.kr [한국지능정보사회진흥원 빅데이터플랫폼팀 유시형 수석]

주 소 [41068] 대구광역시 동구 첨단로 53

홈페이지 www.nia.or.kr

- * 본 가이드의 내용은 한국지능정보사회진흥원의 공식견해와 다를 수 있습니다.
- * 본 가이드 내용에 대한 무단 전제를 금하여, 가공 및 인용할 때에는 반드시 출처를 명기하여 주시기 바랍니다.
- * 본 책자는 정보통신기금으로 제작되었습니다.(비매품)