

National Archives Standard

I 행정정보 데이터세트 기록관리 기준 - 관리기준표 작성 및 이관규격

Record Keeping Criteria for Dataset
- Composition of Dataset Management Reference Table &
Exchange of Dataset -

Version 1.0

- 제정자 : 행정안전부 국가기록원장
- 제정일 : 2020년 10월 16일(국가기록원 고시 제2020-43호)
- 심 의 : 국가기록관리위원회, 기록·정보정책전문위원회
- 원안 작성 :
 - 왕호성(국가기록원 기록연구사)
- 검 토 :
 - 백영미(한국철도시설공단) • 이희원(한국원자력연구원)
 - 한상진(육군기록관리정보단) • 김명옥(국가기록원)
 - 김재훈(국가기록원) • 이주광(국가기록원)
 - 이지영(국가기록원) • 김수영(국가기록원)
 - 김정애(국가기록원) • 조영주(국가기록원)
- 자 문 :
 - 양동민(전북대학교) • 설문원(부산대학교)
 - 임진희(서울특별시청)
- 관 리 :
 - 국가기록원 정책기획과

(1) 이 표준의 열람은 홈페이지를 이용하시고, 의견 또는 질문은 아래 전화로 연락 주십시오.

- 표준열람 : 국가기록원(<http://www.archives.go.kr>)→기록관리업무→기록관리 표준→표준화 현황
- 행정안전부 국가기록원 기록정책부 디지털기록혁신과(042-481-1762)
정책기획과(042-481-6231)

(2) 이 표준은 「저작권법」 제24조의 2(공공저작물의 자유이용)에 따라 저작권자인 국가기록원의 허락 없이 자유롭게 이용할 수 있다. 다만, 저작물을 이용하는 자는 그 출처를 명시하여야 하며, 영리를 목적으로 하는 무단 복제 및 배포는 금지합니다.

Copyright© National Archives of Korea(2020). All Rights Reserved.



목 차

머리말	ii
개요	iii
1 적용범위	1
2 적용근거	1
2.1 법적 근거	1
2.2 인용표준	1
2.3 다른 표준과의 연계	2
3 용어정의	2
4 행정정보 데이터세트에 대한 일반사항	3
5 행정정보시스템의 선정	6
6 행정정보 데이터세트 관리기준표	7
6.1 관리기준표의 구성	8
6.2 관리기준표의 작성	12
7 행정정보 데이터세트의 이관규격	13
부속서 A (참고) 관리기준표 작성 예시	15
부속서 B (참고) SIARD 규격	21
참고문헌	49

머리말

이 표준은 디지털 환경에서 공공기관이 행정정보시스템을 통해 운영하고 있는 행정정보 데이터세트의 기본적인 기록관리 기준을 제시하기 위해 제정되었다.

이 표준은 2020년에 신설된 「공공기록물 관리에 관한 법률」 시행령 제25조 제6항, 제34조의3조에 따라 행정정보시스템의 관리대상 선정, 관리기준표의 작성, 이관 방법 등에 관한 사항을 정리하였다.

이 표준은 다음과 같이 구성하였다. 제1절부터 제3절에서는 표준의 적용범위와 적용근거 및 용어를 정의하였다. 제4절에서는 행정정보 데이터세트에 대한 일반사항을 기술하였으며, 제5절에서는 행정정보시스템의 선정, 제6절에서는 행정정보 데이터세트 관리기준표의 구성항목과 작성방법을 기술하였다. 제7절에서는 행정정보 데이터세트를 관할 영구기록물관리기관으로 이관할 경우 참고할 수 있는 이관규격을 소개하였다. 부속서에서는 행정정보 데이터세트 관리기준표의 예시를 소개하고 이관이 필요한 경우 활용할 수 있는 참고규격인 스위스 연방기록원의 SIARD규격을 수록하였다.

이 표준은 기록·정보정책전문위원회 및 국가기록관리위원회 심의를 거쳐 제정하였으며 국가기록원이 유지·관리한다. 이 표준은 관련 법령의 개정, 관계 기관 및 이해 당사자의 요청 등 개정 사유가 발생할 경우 그 필요성 및 타당성 검토 후 개정안을 마련하고 의견수렴 및 심의 절차를 거쳐 개정을 추진한다.

개요

전자정부의 추진에 따라 종이문서는 전자문서로 전환되었으며 행정업무의 상당 부분이 행정정보시스템을 통해 이루어지면서 데이터세트 기록이 양산되고 있다. 데이터세트를 생산·운용하는 행정정보시스템 역시 급증하여 2020년 현재 범정부 EA포털(Government-wide Enterprise Architecture)에 등록된 행정정보시스템 수는 16,000여개를 넘어선 상태이다.

2010년에 이미 「공공기록물 관리에 관한 법률」을 개정하여 행정정보 데이터세트가 법령상 기록관리 대상에 포함되었지만, 그동안 문서유형 중심으로 설계된 기존의 기록관리체계는 행정정보 데이터세트의 관리에 제대로 기능하지 못하였다.

행정정보 데이터세트의 기록관리 방안을 설계하기 위하여 다양한 이해관계자들의 의견과 분야 전문가들의 자문을 받아 수년간의 연구개발 및 시범사업을 지속적으로 수행하였다. 다음은 그간의 행정정보 데이터세트의 기록관리 방안 설계를 위하여 수행한 관련 업무의 추진 경과이다.

- 2016 행정정보시스템 현황 조사 및 기록관리 방안 초안 마련
- 2017 행정정보 데이터세트 기록관리 체계 설계 연구 및 시범 적용
- 2018 행정정보 데이터세트 기록관리 체계(안) 마련 및 기관 설명회(5회)
- 2019 행정정보 데이터세트 기록관리 체계 구축 시범사업(1차, 14개 시스템)
- 2019 데이터세트 유형 전자기록의 장기보존기술 연구(SIARD 검증)
- 2020 행정정보 데이터세트 기록관리 체계 구축 시범사업(2차, 50개 시스템)
- '15~'20 행정정보 데이터세트 기록관리 표준화작업반 운영(11회)

그간의 관련 연구와 시범적용 등 검증을 거쳐 2020년 3월 「공공기록물 관리에 관한 법률」 시행령에 행정정보 데이터세트 관리와 관련된 조항이 신설되었고 이를 근거로 이 표준이 만들어졌다.

중앙기록물관리기관은 관련 연구와 시범사업의 실행을 통해 획득되는 모범사례 및 최신 기술 정보를 축적하여 단계적으로 법률과 표준에 반영할 예정이다. 따라서 이 표준은 기능하지 않고 검증되지 않은 절차와 방안을 배제하

고 행정정보 데이터세트 관리의 기본적 요건만을 제시하였다. 표준의 내용은 관리대상의 선정, 관리기준표의 작성, 이관규격 등을 정리하였으며 좀 더 세부적인 절차와 방안은 매년 발간하는 지침에서 다루도록 한다.

행정정보 데이터세트 기록관리 기준

- 관리기준표 작성 및 이관규격

1 적용범위

이 표준은 행정정보시스템을 운영하는 공공기관이 「공공기록물 관리에 관한 법률」 시행령 제2조11호, 제25조제6항, 제34조의3에 근거하여 행정정보 데이터세트를 기록물로 관리하기 위한 기본적인 절차와 방법을 규정한다. 이 표준은 행정정보 데이터세트와 관련한 업무를 수행하는 담당자들이 이해하고 따라야 한다.

이 표준은 행정정보 데이터세트 관리방안의 고도화를 위하여 표준이 제시한 방안 이외의 기술과 절차를 적용하는 것을 제약하지 않는다. 별도의 관리방안이 필요한 공공기관은 관할 영구기록물관리기관과 협의하여 이를 적용할 수 있다. 다만, 그 적용 결과는 중앙기록물관리기관과 공유하여 법령과 표준 개정예 참고할 수 있도록 하여야 한다.

2 적용근거

2.1 법적 근거

이 표준의 구체적인 법적 근거는 다음과 같다.

- 「공공기록물 관리에 관한 법률」 시행령 제2조11호(정의)
- 「공공기록물 관리에 관한 법률」 시행령 제25조제6항(기록관리기준표)
- 「공공기록물 관리에 관한 법률」 시행령 제34조의3(행정정보 데이터세트의 관리)
- 「행정 효율과 협업 촉진에 관한 규정」 제3조제12호(정의)

2.2 인용표준

다음의 표준은 이 표준의 적용을 위해 인용된 표준이다. 발행연도가 표기된 인용표준은 인용된 판만을 적용한다. 발행연도가 표기되지 않은 인용표준은 최신판(모든 개정내용을 포함)을 적용한다.

- KS X ISO 15489-1: 2007 문헌정보-기록관리-제1부:일반사항
- KS X ISO 16175-3 : 2010 문헌정보-전자사무환경에서 기록에 대한 원리 및 기능요건-제3부:업무시스템의 기록관리지침 및 기능요건

2.3 다른 표준과의 연계

다른 기록관리 표준 중에서 이 표준의 적용을 위해 필요하거나 직접적으로 연관이 있는 표준은 없다.

3 용어정의

이 표준의 목적을 위하여 다음의 용어와 정의를 적용한다.

3.1 단위기능

행정정보 데이터세트 관리를 위한 기본단위이며 행정정보 데이터세트 관리 기준표의 작성단위

3.2 데이터

정보시스템에서 생산, 수집, 가공, 저장, 검색, 제공, 송신, 수신 등을 위해 조합된 문자, 숫자, 도형, 이미지 및 그 밖의 개체

[「공공기록물 관리에 관한 법률」 시행령 제2조의 내용 개작]

3.3 데이터세트

정보시스템에서 생산, 수집, 가공, 저장, 검색, 제공, 송신, 수신 등을 위해 조합된 문자, 숫자, 도형, 이미지 및 그 밖의 데이터 집합

[「공공기록물 관리에 관한 법률」 시행령 제2조의 11호의 내용 개작]

3.4 행정정보시스템

행정기관이 행정정보를 생산·수집·가공·저장·검색·제공·송신·수신하고 활용할 수 있도록 하드웨어·소프트웨어·데이터베이스 등을 통합한 시스템

[「행정 효율과 협업 촉진에 관한 규정」 제3조제12호]

3.5 행정정보 데이터세트

공공기관의 행정정보시스템에서 생산, 수집, 가공, 저장, 검색, 제공, 송신, 수신 등을 위해 조합된 문자, 숫자, 도형, 이미지 및 그 밖의 데이터 집합

[「공공기록물 관리에 관한 법률」 시행령 제2조제11호의 내용 개작]

3.6 행정정보 데이터세트 관리기준표

행정정보 데이터세트의 기록관리를 위해 필요한 정보를 작성할 수 있도록 중앙기록물관리기관이 정한 서식

4 행정정보 데이터세트에 대한 일반사항

행정정보 데이터세트는 공공기관의 행정정보시스템을 통해 생산, 운용되는 모든 형태의 정보자료를 말한다. 대부분의 공공기관은 관계형 데이터베이스를 사용하여 행정정보 데이터세트를 생산·운용하고 있지만, 최근 빅데이터 처리를 위해 비관계형 데이터베이스의 사용이 증가하는 추세이다. 행정정보 데이터세트는 이러한 데이터베이스의 종류와 관계없이 공공기관의 업무 데이터 및 이와 연계된 파일 등 모든 형태의 정보자료를 포함한다.

행정정보시스템은 일반적으로 애플리케이션, 정형 데이터인 데이터베이스 및 이들과 연계된 비정형 데이터인 첨부파일 등으로 구성된다. 행정정보 데이터세트는 행정정보시스템에 기술적으로 종속되며 업무 목적에 따라 다양하고 고유한 기능적 특성을 가진다.

공공기관에서 가장 많이 활용되고 있는 관계형 데이터베이스를 기준으로 행정정보 데이터세트는 다음과 같은 5계층의 구조를 가진다.

· 데이터 - 필드 - 테이블 - 데이터베이스 - 행정정보시스템

비고 1 데이터베이스는 관련 데이터의 조직화된 축적체이다. 테이블은 각기 연관된 데이터 요소로 구성된 하나 내지 그 이상의 관련 데이터베이스 필드들의 집합이며 필드는 데이터베이스 내의 정보 범주를 재현하는 일련의 하나 혹은 그 이상의 관련 데이터 요소를 의미한다.

그림 1은 ‘KS × ISO 16175-3 : 2010 문헌정보-전자사무환경에서 기록에 대한 원리 및 기능요건-제3부 : 업무시스템의 기록관리지침 및 기능요건’에서 제시하고 있는 데이터베이스 내의 구성요소를 나타낸다.



그림 1 - 데이터베이스 구성 요소

비고 2 데이터베이스의 구성에 대한 보다 자세한 사항은 ‘KS X ISO 16175-3 : 2010 문헌정보-전자사무환경에서 기록에 대한 원리 및 기능요건 - 제3부: 업무시스템의 기록관리지침 및 기능요건’을 참조한다.

행정정보 데이터세트는 업무기능에 따라 개별 데이터 또는 데이터의 조합이 기록이 될 수도 있고, 테이블 또는 데이터세트 전체를 대상으로 기록에 대한 처분행위를 수행할 수도 있다. 이는 행정정보 데이터세트가 행정정보시스템의 계층이나 대상에 관계없이 기록관리 단위가 될 수 있음을 의미한다.

특정 업무기능의 보존기간이 종료되어도 이 기능과 연계된 데이터세트와 다른 현용 업무기능이 연계되어 있다면 이 데이터세트를 폐기해서는 안된다.

단위기능은 행정정보 데이터세트의 기록관리 단위를 정할 때 기준이 되며, 실제 행정정보시스템에 구현된 여러 데이터세트 계층에 제약 없이 부여하여 처분행위를 실행할 수 있다.

그림 2는 단위기능에 연계될 수 행정정보시스템의 다양한 계층과 단위를 도식화한 개념도이다.

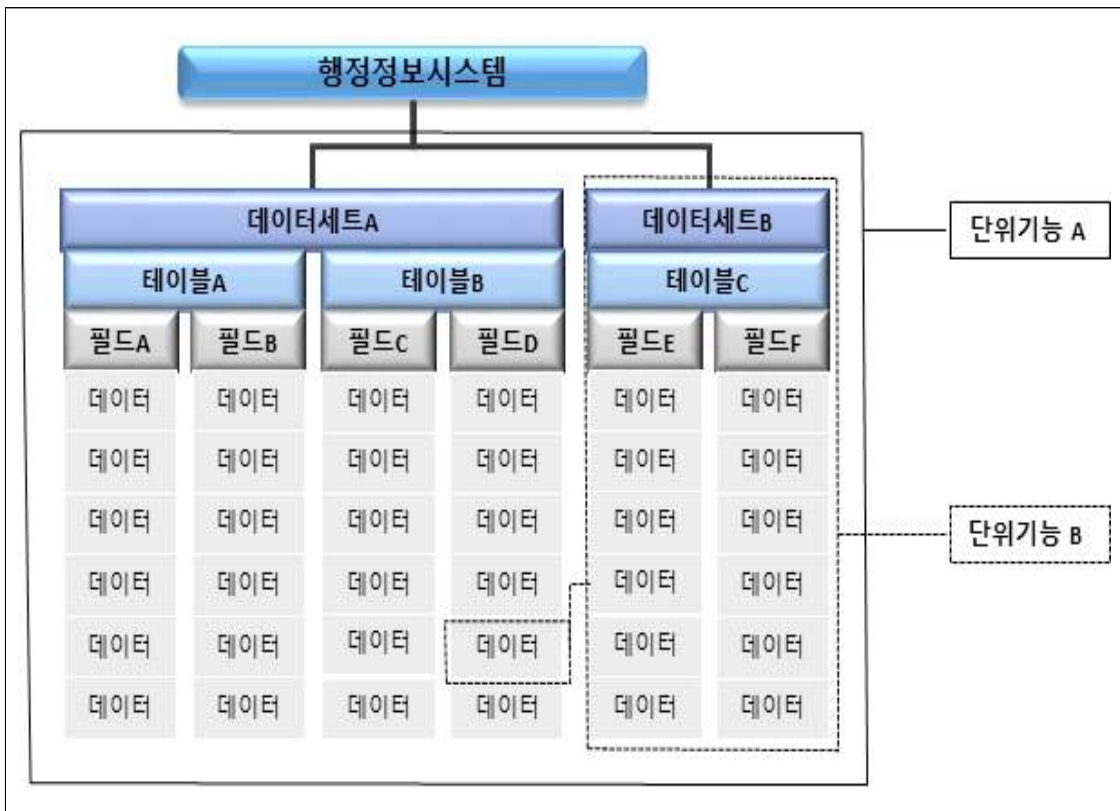


그림 2 - 단위기능과 행정정보시스템의 연계

시스템을 기본단위로 하여 전체 행정정보 데이터세트를 관리하는 경우에는 하나의 단위기능으로 관리가 가능하다. 위 그림의 검은색 실선은 단위기능 A가 연계된 것을 그림으로 표현한 것이다.

만약 검은색 점선과 같이 데이터세트 하위의 특정 데이터 단위를 관리해야

하는 경우에는 단위기능 B를 추가로 만들어 2개의 단위기능으로 확장하는 것이 가능하다. 이처럼 추가적인 세부 단위로 데이터세트 관리가 필요한 경우에는 필요한 만큼 단위기능을 추가하여 사용할 수 있다.

5 행정정보시스템의 선정

「공공기록물 관리에 관한 법률」 시행령 제34조의3에 따라 공공기관은 행정정보시스템을 기본단위로 하여 행정정보 데이터세트를 관리해야 한다. 이를 위하여 공공기관은 기관이 보유한 모든 행정정보시스템을 대상으로 행정정보 데이터세트의 존재 유무를 확인하여 관리대상을 선정하여야 한다.

이때 행정정보 데이터세트의 존재 유무를 확인하고 관리대상을 선정하는 작업은 행정정보시스템을 운용하는 처리과와 기관의 기록관이 협의하여 공동으로 수행한다.

만약, 여러 기관이 공동으로 시스템을 운영하는 경우에는 시스템과 데이터세트의 운용 주체가 각각 다를 수 있으므로, 이 경우에는 관련 기관이 공동으로 선정 작업을 수행할 수 있다.

비고 여러 기관이 공동으로 활용하는 시스템의 경우 시스템 관리권한을 가진 기관이 데이터 보유권한을 가진 기관의 의견수렴 등을 통하여 기록으로 관리되어야 하는 행정정보시스템을 선정할 수 있다.

공공기관은 아래와 같은 선정 기준에 따라 행정정보시스템을 4가지 유형으로 구분하여 관리대상을 선정한다.

표 1 - 행정정보시스템 선정 기준

유형	시스템 구분	내용	비고
A	고유 업무용 시스템	개별기관의 고유 행정업무를 시스템으로 구현하고 그 내용을 데이터세트로 관리	행정정보 데이터세트 관리방안 적용
B	공통 업무용 시스템	여러 기관의 공통 행정업무를 시스템으로 구현하고 그 내용을 데이터세트로 관리	
C	기록관리 적용 시스템	이미 기록관리 절차가 적용된 시스템(업무관리시스템, 전자문서시스템, RMS, AMS, 웹기록물 등)	별도 관리방안 적용
D	단순지원 시스템/보안·NW 등 기술시스템	업무 내역이 DB로 저장되지 않은 단순 중계 시스템/ICT 운영 등을 위한 기술적 시스템	관리방안 적용 안함

단일기관에서 고유 업무를 위하여 구축된 행정정보시스템은 A 유형으로 구분하고, 여러 기관에서 공통 업무를 위하여 구축된 행정정보시스템은 B 유형으로 구분한다. 이 두 가지 유형의 행정정보시스템은 관리대상으로 선정하여 행정정보 데이터세트 기록관리 방안을 적용하여야 한다.

C 유형으로 구분하는 전자문서, 웹기록물 등은 행정정보 데이터세트와 같이 전자기록물이지만 이미 기록관리 절차를 적용하여 관리하고 있으므로 행정정보시스템 선정 대상에서 제외한다.

행정정보 데이터세트가 생산되지 않고 기능만을 제공하는 시스템은 D 유형으로 구분하고, 선정 대상에서 제외한다.

공공기관은 기관의 행정정보시스템에 대해 관리대상 선정과정을 거친 후 그 결과를 관할 영구기록물관리기관에 통보하여야 한다.

6 행정정보 데이터세트 관리기준표

행정정보 데이터세트 관리기준표는 행정정보 데이터세트의 기록관리를 위해 필요한 정보를 작성할 수 있도록 중앙기록물관리기관이 정한 서식이다. 관리기준표는 개별 시스템의 특성을 고려하여 합리적인 기록관리 체계를 마련하기 위한 기본적인 정보자료이다.

공공기관의 장은 시스템을 기본단위로 하여 행정정보 데이터세트 관리기준표를 작성하여야 하며, 관리기준표를 통해 획득되는 정보를 바탕으로 시스템 하위의 세부단위로 관리기준표를 작성할 수 있다.

행정정보 데이터세트는 개별 행정정보시스템에 따라 데이터의 분류, 구조 등이 각각 달라 하나의 기록관리 단위를 확정하여 관리하기 어려운 특징이 있다. 이는 단위과제라는 하나의 관리단위에 기록물을 집합적으로 묶어 동일한 보존기간으로 관리하는 기록관리기준표와는 차이가 있다.

비고 이 표준에서는 정부기능분류체계(BRM : Business Reference Model)에 따라 단위과제를 기록의 관리계층으로 정해 관리기준을 제시한 기록관리기준표와는 다른 개념으로 행정정보 데이터세트 관리기준표를 사용한다.

6.1 관리기준표의 구성

행정정보 데이터세트 관리기준표는 관리정보, 법규정보, 업무정보, 기록관리정보, 시스템정보, 데이터정보의 6개 영역으로 구분되며 모두 36개의 항목으로 구성된다.

관리정보는 행정정보시스템을 운영·관리하는 기관, 부서 등을 기입하는 항목이다. 법규정보는 행정정보시스템의 구축 근거 또는 해당 시스템으로 수행하는 행정업무의 수행 근거를 기입하는 항목이다. 시스템정보는 행정정보시스템의 명칭, 구축년도, 정형데이터가 저장되는 DBMS, 개발 산출물 내역 등을 기입하는 항목이다. 데이터정보는 대표 데이터, 파일 형식의 비정형데이터, 암호화와 연계시스템의 유무 등을 기입하는 항목이다. 업무정보는 행정정보시스템이 제공하는 기능을 통해 수행하는 업무명과 내용 등을 기입하는 항목이다.

목이다. 하나의 시스템을 기본단위로 관리하더라도 시스템을 구성하는 업무 단위는 여러 개로 구성되는 것이 일반적이다. 이 경우 업무정보 역시 복수로 구성될 수 있다.

기록관리정보는 앞의 5개 영역을 종합·분석하여 기록관리에 필요한 정보를 기입하는 항목으로 관리 기준단위인 단위기능명, 데이터의 소유권한과 시스템의 관리권한, 접근권한, 정보공개 구분, 보존기간, 처분 정보를 기입한다.

행정정보 데이터세트 관리기준표의 기본 서식은 다음과 같다.

표 2 - 행정정보 데이터세트 관리기준표

(필수:M / 선택:O)

영역	항목	내용	구분
1. 관리정보	기관명	행정정보시스템(이하 시스템) 관리기관 명을 기입한다. 예시) 인사혁신처	M
	부서명	시스템 운영부서 명을 기입한다. 예시) 정보화담당관	M
	담당자명	시스템 운영담당자 명을 기입한다. 예시) 홍길동	O
2. 법규정보	법령/규정	시스템과 관련된 법령과 규정을 기입한다. 예시) 국가공무원법 제19조2(인사관리의 전자화)	M
3. 시스템정보	시스템명	시스템의 공식명칭(통용명칭)을 기입한다. 예시) 전자인사관리시스템(e-사람)/기록관리교육훈련시스템	M
	시스템 개요	시스템을 이해할 수 있는 개요정보를 기입한다. 예시) 중앙행정기관의 인사행정을 처리하는 공용 표준시스템	O
	구축년도	시스템의 구축년도를 기입한다. 예시) 2012년	M
	개발 산출물	시스템 개발 최종산출물을 기입하고 해당 파일을 첨부한다.	M

		예시) ERD, 테이블명세서, 기능명세서, 사용자 지침서 별첨		
	EA 정보	범정부 EA포털에 등록된 시스템 정보를 PDF 저장하여 첨부한다. 예시) EA정보_전자인사관리시스템.PDF	O	
	DBMS 정보	DBMS 정보를 기입한다. 예시) Oracle Standard 11g, 오라클 10	M	
		업무 DB명	데이터세트와 관련된 직접적인 업무DB명을 기입한다. 예시)e-사람의 복무DB(모듈), 급여DB(모듈), 인사카드DB(모듈), 인사평가DB	M
4. 데이터 정보	대표데이터		DBMS가 관리하는 대표적인 데이터세트나 테이블명을 기입한다. 예시) 인사/급여/복무/평가 정보 데이터 그룹(인사모듈/급여모듈 등)	M
	비정형데이터 (첨부파일의 종류)	종류	비정형 데이터의 종류를 기입한다.(DB 이외 첨부되는 파일) 예시) 사진파일, 증명서 사본, 건축도면 등	M
		포맷	비정형 데이터의 포맷정보를 기입한다. 예시) PDF, hwp, xls, dwg 등	M
		구동 SW	비정형 데이터 파일을 구동하는 SW를 기입한다. 예시) 오즈뷰어, 엑셀	O
		저장 방법	비정형 데이터 파일의 저장방법을 기입한다. 예시) 별도 스토리지 저장, BLOB, CLOB 등	O
	암호화데이터 사용유무		암호화된 데이터를 사용하는 경우 기입한다. 예시) 주민번호는 암호화, 첨부파일 경로 암호화 등	O
	연계시스템 유무		연계시스템이 있는 경우 기입한다.	O
	연계내역 첨부		연계시스템의 정보가 담긴 파일을 첨부한다.	O
5. 업무 정보	업무명	시스템이 제공하는 주요 업무명을 기입한다.(여러 개일 경우 5번 영역을 반복하여 작성) 예시) 복무, 급여, 인사관리 등 여러 개일 수 있다.	M	
	업무내용	업무의 상세내용을 기입한다. 예시) 공무원의 복무상황 및 근무상황	M	

	업무목적	예시) 공무원 복무 관련 지원	M
	대상	해당 업무의 서비스 대상을 기입한다. 예시) 전체 공무원	O
6. 기록관리정보	단위기능명	데이터세트 기록관리 대상이 되는 단위기능명을 기입한다. 예시) e사람의 경우, 복무-급여-인사 등 여러 개 /기록관리교육훈련시스템의 경우 기록관리 교육 1개로 지정 ※ 시스템에 단위기능이 여러 개 있을 경우 기준표는 1개 작성 가능(필요시 단위기능 개수별로 작성도 가능)	M
	업무활용 목적	기록관리 단위기능의 업무활용 목적을 기입한다. 예시) 공무원의 복무관리 지원	M
	주제어	단위기능과 관련된 주제어를 기입한다. 예시) 복무, 근무상황, 연가, 초과근무 등	M
	데이터의 소유권한	시스템을 활용하여 데이터를 생산하고 처리하는 기관명을 기입한다. 예시) 각 사용부처(행안부, 인사처 등 42개 중앙행정기관)	O
	시스템의 관리권한	시스템과 DBMS를 운영·관리하는 기관명을 기입한다. 예시) 인사혁신처	O
	접근권한	단위기능 데이터에 접근할 수 있는 권한을 가진 자를 기입한다. 예시) 부서담당자, 기록관 담당자, 시스템 관리자	O
	정보공개 구분	단위기능 데이터 정보의 공개(부분공개)/비공개 사항을 기입한다. 예시) 개인정보는 비공개며 그 외 정보는 공개	M
	보존기간	단위기능별로 보존기간을 기입한다.(BRM, 기록관리기준표 등 보존기간 참조) 예시) 3년	M
	보존기간 책정사유	보존기간을 책정한 법률적 사유나 근거를 기재한다. 예시) 기관공통업무 보존기간표에 따라 3년 책정	M
	적용 범위	단위기능에 해당하는 테이블들이나 영역을 명시한다. 소형시스템의 경우 전체를 적용하나, 대형시스템의 경우 해당기능의 조회-출력용 쿼	M

		리 등을 확인하여 단위기능에 해당하는 범위를 효율적으로 지정할 수 있다. 예시) 단위기능 전체 or 급여관련 테이블만 지정	
	적용범위 관련 정보	적용범위가 일부 테이블일 경우 해당하는 DB 테이블 정보를 기입한다.(적용범위가 전체일 경우 테이블명을 구분할 필요가 없다. 예시) TN_PAVGroupusw(급여테이블)...등. 단순 시스템의 경우 전체로 표기	○
	처분의 제약 발생사항 (처분지연의 사유)	대상 데이터가 다른 기능과 연계되어 사용되는 기간이나, 다른 업무에 참고하는 기능에서 사용되는 기간과 사유 등을 입력한다.(기록관리 처분이 보류되어야 하는 기간으로 시스템적 기술검토 후 작성) 예시) 복무시간 DB는 급여업무에서 초과근무수당 계산에 활용되어 7시간 보류 필요(급여는 보존기간 10년)/통계모듈에서 참조되어 5년간 보류 필요 등	○
	처분 방법	보존기간 도래 시 처분방법을 기입한다. 예시) 처분의 제약 기간 7년과 보존기간 3년이 모두 경과되면 평가심의 후 평가심의 결과에 따라 해당 기간의 데이터를 처분한다.	○

행정정보 데이터세트 또는 행정정보시스템의 특성과 기관 업무의 고유성 등을 반영하기 위하여 추가 정보 등을 기입할 필요가 있는 기관에서는 관찰 영구기록물관리기관과 협의하여 관리기준표의 항목을 변경할 수 있다.

6.2 관리기준표의 작성

행정정보 데이터세트 관리기준표는 행정정보시스템을 운용하는 처리과와 기관의 기록관이 협의하여 공동으로 작성하여야 한다.

비고 1 여러 기관이 공동으로 활용하는 시스템의 경우 시스템 관리권한을 가진 기관이 관리기준표를 작성하되, 작성 과정에서 데이터 보유권한을 가진 기관의 의견을 수렴할 수 있다.

행정정보 데이터세트 관리기준표는 행정정보시스템을 기본단위로 하여 작성

하여야 한다. 다만, 행정정보 데이터세트를 세부적인 단위기능으로 관리해야 하는 경우에는 세부 단위기능의 수만큼 관리기준표를 작성하여야 한다.

비고 2 ‘4 행정정보 데이터세트에 대한 일반사항’의 행정정보 데이터세트 계층구조를 참고하여 관리기준표를 작성할 수 있다.

관리기준표의 모든 항목은 행정정보시스템 개발 산출물과 실제 시스템 구성 내용을 확인하여 정확하게 기입하여야 한다. 실제 행정정보시스템을 대상으로 작성한 행정정보 데이터세트 관리기준표의 예시는 부속서 A를 참조한다.

7 행정정보 데이터세트의 이관규격

행정정보 데이터세트의 이관은 기록관리를 위하여 행정정보 데이터세트의 전체 또는 일부를 기록물관리기관으로 이관하거나 기록관리 권한을 이전하는 것을 말한다.

행정정보 데이터세트를 영구기록물관리기관 등 외부기관으로 이관하는 경우에는 관할 영구기록물관리기관이 지정한 이관방식을 사용하여야 한다. 행정정보 데이터세트의 이관에 활용되는 도구는 특정 기술에 대한 의존성이 낮고 비독점적이며 공개된 방식이 권장된다.

행정정보시스템을 운용하는 처리과는 행정정보 데이터세트의 이관이 정상적으로 수행될 수 있도록 기술을 지원하여야 하며, 기관의 기록관은 이를 관리·감독하여야 한다.

공공기관이 행정정보 데이터세트를 영구기록물관리기관 등으로 이관하기 위하여 이관도구를 개발하는 경우에는 스위스 연방기록원의 SIARD(Software Independent Archiving of Relational Databases) 형식규격(부속서 B)을 참고할 것을 권고한다.

비고 1 2019년 국가기록원은 SIARD 형식규격에 따라 이관도구를 고도화하는 연구개발 및 시범사업을 수행하여 규격을 검증하였다. 이에 따라

이관이 필요한 공공기관은 이 규격을 참고할 것을 권고한다. 다만, 향후 기술 및 국제표준의 변화에 따라 참고규격은 변경될 수 있다.

비고 2 부속서 B는 SIARD 2.1(표준번호 eCH-0165)을 한글 번역하여 제공하며 예시가 담긴 부록은 원문을 참조한다.

SIARD 형식규격은 행정정보 데이터세트 이관도구의 개발을 위해 참고용으로 제공되는 것이며 반드시 따라야하는 의무 규격은 아니다. 기관의 사정에 따라 별도의 이관방안이 필요한 경우에는 관할 영구기록물관리기관과 협의하여 사용할 수 있다.

부속서 A (참고)

행정정보 데이터세트 관리기준표 작성 예시

A.1 단위기능이 1개인 경우

영역	항목		내용	구분	
1.관리정보	기관명		국가기록원	M	
	부서명		기록관리교육센터	M	
	담당자명		○○○	O	
2.법규정보	법령/규정		공무원인재개발법 제4조, 공무원인재개발법 시행령 제18조	M	
3.시스템 정보	시스템명		기록관리교육훈련시스템	M	
	시스템 개요		기록물관리 교육과 관련하여, 수강신청, 수강이력 및 교육생 수료관리 등 기록관리교육훈련 운영 전반 온라인 관리	O	
	구축년도		2009년	M	
	개발 산출물		운영성과측정_업무수행영향도_근거자료(기록관리교육훈련시스템).xlsx, EDU-02-기록관리교육_Ver1.01.erwin, 기록관리교육훈련시스템_FP.xlsx	M	
	EA 정보		EA정보_기록관리교육훈련시스템.pdf (20190819)	O	
	DBMS 정보	DBMS 정보	DBMS 정보	Oracle 10g	M
			업무 DB명	기록관리 교육	M
4.데이터 정보	대표 데이터		교육운영자료, 교육강사,	M	
	비정형 데이터 (첨부파 일의	종류	교육자료	M	
		포맷	동영상, 한글, 파워포인트, 이미지 파일 (.pdf, .hwp, .avi, .jpeg, .ppt)	M	

	종류)	구동 SW	윈도우 미디어 플레이어, 그림판, 한글, PDF 리더, MS 파워포인트	O
		저장 방법	파일저장스토리지	O
	암호화데이터 사용유무	파일 저장시 파일경로를 MD5 알고리즘을 통한 저장 사용	O	
	연계시스템 유무	없음	O	
	연계내역 첨부	없음	O	
5.업무정보	업무명	기록관리교육	M	
	업무내용	기록관리교육훈련 운영과 관련하여 교육생의 수강신청, 수강이력, 수료관리, 사이버컨텐츠와 강사명단 관리	M	
	업무목적	기록관리교육훈련	M	
	대상	기록관리 교육생, 강사	O	
6.기록관리 정보	단위기능명	기록관리교육	M	
	업무활용 목적	기록물 관련 교육정보를 등록하고 내역을 관리	M	
	주제어	강사관리, 교육운영, 교육참고자료	M	
	데이터 보유권한	국가기록원	O	
	시스템 관리권한	국가기록원	O	
	접근권한	기록관리교육센터 관리자, 교육강사, 전 공무원	O	
	정보공개 구분	공개	M	
	보존기간	10년	M	
	보존기간 책정사유	2019년 발행 기록물 관리지침(공통매뉴얼) 내 기관공통업무 보존기간표에 따라 인사관리-교육훈련 내 국내훈련 관련하여 보존기간 10년 적용	M	
적용 범위	전체	M		

	적용범위 관련 정보		T B _ P R O F E S S O R S , T B _ C O N T E N T _ M E T A , T B _ C O N T E N T _ R E S O U R C E S , T B _ C O N T E N T S , T B _ C O U R S E _ L E C T U R E S , T B _ C O U R S E _ S C H E D U L E S , T B _ C O U R S E _ U S E R S	O
	처분 제약 발생 사항 (처분 지연의 이유)		향후 통계정보 등 활용을 위해 보존 관리(보존관리기간 10년)	O
	처분 방법		처분 제약에 따른 기간 동안 보존관 리 후 다음연도 1월1일을 기산일로 하여 보존기간이 지난 후 평가심의를 거쳐 폐기 확정된 경우 처분 관련 정 보의 DB 데이터 및 전자파일을 폐기 처리	O
부가정보	주요 서비스 내용	서비스 내용	○○○○	O
		화면 캡처	화면 캡처 파일 별첨으로 제공	
		표현 쿼리	SELECT E.* FROM (SELECT ROWNUM AS rnum, C.* FROM WHERE ROWNUM<= 10) E WHERE E.rnum >= 1	
	이관대상 기관명		○○○	
	이관대상 부서명		○○○	
	이관대상 담당자명		○○○	

A.2 단위기능이 2개인 경우

영역	항목		내용	구분
1.관리정보	기관명		○○○○	M
	부서명		정보화○○○	M
	담당자명		김○○/정○○	O
2.법규정보	법령/규정		국가공무원법, 공무원 인사기록	M
3.시스템 정보	시스템명		전자인사관리시스템(e-사람)	M
	시스템 개요		60개 중앙행정기관 인사관리	O
	구축년도		2012년	M
	개발 산출물		ERD, 기능명세서, 매뉴얼	M
	EA 정보		EA정보_전자인사관리시스템.pdf	O
	DBMS 정보	DBMS 정보	Oracle Standard 11g	M
업무 DB명		통합인사DB	M	
4.데이터 정보	대표 데이터		인사, 급여, 복무, 평정 정보데이터	M
	비정형 데이터 (첨부파일 의 종류)	종류	개인 증명사진, 증명서 사본, 인사정보 연계파일	M
		포맷	JPG, PDF, excel, HWP, zip	M
		구동 SW	그림판, PDF Viewer, MS office, 한컴 오피스	O
		저장 방법	파일저장스토리지	O
	암호화데이터 사용유무		주민번호, 첨부파일 경로, 사용자 테이블 내 패스워드	O
	연계시스템 유무		디브레인	O
	연계내역 첨부		예산회계 정보.xml	O
5.업무 정보(1)	업무명		복무	M
	업무내용		복무상황, 근무상황	M
	업무목적		복무관련 지원	M
	대상		전체 공무원	O
5.업무	업무명		급여	M

정보(2)	업무내용	급여관리, 비공무원관리	M
	업무목적	공무원, 비공무원 급여 관련 지원	M
	대상	전체 공무원	O
6.기록관리 정보 (1)	단위기능명	복무	M
	업무활용 목적	근무형태, 근무상황, 초과근무 등	M
	주제어	근무형태, 근무상황, 초과근무 등	M
	데이터 보유권한	각 사용부처	O
	시스템 관리권한	○○○○○	O
	접근권한	본인, 부서담당자, 기록관담당자, 시스템 관리자	O
	정보공개 구분	개인정보 비공개(다만, 개인정보가 제외된 공무 출장 정보는 공개 가능)	M
	보존기간	5년	M
	보존기간 책정사유	복무 제도 정책 개선참고	M
	적용 범위	단위기능 전체	M
	적용범위 관련 정보	TN_BSR(복무평가테이블), TN_BSS(조직테이블)...	O
	처분 제약 발생 사항 (처분 지연의 이유)	향후 통계정보 등 활용을 위해 보존 관리	O
처분 방법	보존기간 종료시 평가심의 후 처분	O	
6.기록관리 정보 (2)	단위기능명	급여	M
	업무활용 목적	공무원, 비공무원 급여 관련 정보 확인	M
	주제어	급여관리, 연말정산, 기여금, 건강보험	M
	데이터 보유권한	각 사용부처	O
	시스템 관리권한	○○○○○	O
	접근권한	본인, 부서담당자, 기록관담당자, 시스템 관리자	O

	정보공개 구분	비공개	M	
	보존기간	30년	M	
	보존기간 책정사유	2019년 발행 기록물 관리지침	M	
	적용 범위	단위기능 전체	M	
	적용범위 관련 정보	TN_STMSD(급여테이블), TN_USERG (사용자테이블)	O	
	처분 제약 발생 사항 (처분 지연의 이유)	향후 통계정보 등 활용을 위해 자체 보존	O	
	처분 방법	처분 제약에 따른 기간 동안 보존관 리 후 다음연도 1월1일을 기산일로 하여 보존기간이 지난 후 평가심의 를 거쳐 폐기 확정된 경우 처분 관련 정보의 DB 데이터 및 전자파일을 폐기 처리	O	
부가정보	주요 서비스 내용	서비스 내용	○○○○	O
		화면 캡처	인사관리카드 화면캡처 자료 첨부	
		표현 쿼리	SELECT E.* FROM (SELECT INSANUM AS rnum, C.* FROM WHERE INSANUM<= 10) E WHERE E.rnum >= 1	
	이관대상 기관명	○○○		
	이관대상 부서명	○○○		
	이관대상 담당자명	○○○		

부속서 B (참고)

SIARD

- Software Independent Archiving of Relational Databases-

B.1 개요

이 부속서는 SIARD 2.1 규격의 표준 명세를 담고 있다. SIARD는 스위스 연방기록원이 특정 소프트웨어에 종속 없이 관계형 데이터베이스들을 장기보존하기 위해 개발한 파일형식으로 유럽연합의 전자정부 표준(eCH-0165)이다.

SIARD 규격형식은 ISO 표준인 유니코드·XML·SQL:2008과 인터넷 표준인 URI, 산업표준인 ZIP을 채택하여 개발되었다. 다음은 사용된 표준 목록이다.

eCH-0150	eCH-0150 Change und Release Management von eCH-Standards http://www.ech.ch/
RFC 1738	URL specification – in particular the “file:” URL/URI https://www.ietf.org/rfc/rfc1738.txt
RFC 1951	Specification of the “deflate” algorithm https://www.ietf.org/rfc/rfc1951.txt
SQL:2008	ISO/IEC 9075(1-4,9-11,13,14):2008: Information technology -- Database languages – SQL http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=53681
Unicode	Unicode 6.1.0 Unicode, Inc. http://www.unicode.org/versions/Unicode6.1.0/ (corresponds to ISO/IEC 10646:2012: Information technology -- Universal Coded Character Set (UCS) http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=56921)
XML	Extensible Markup Language (XML), 1.1 (Second Edition) W3C Recommendation 16 August 2006, edited in place 29 September 2006 http://www.w3.org/TR/2006/REC-xml11-20060816/ (corresponds to ISO/IEC 19503:2005: Information technology -- XML Metadata Interchange (XMI), http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=32622)
ZIP	ZIP File Format Specification, Version 6.3.3 September 1, 2012 PKWARE Inc. http://www.pkware.com/documents/casestudies/APPNOTE.TXT

SIARD 형식을 국제 표준으로 채택하고자 하는 목적은 관계형 DB가 폭넓게 사용되고 있음에도 불구하고 ‘덤프’ 방식으로 인해 특정 업체에 종속되는

현상을 탈피하여 데이터베이스 간의 보다 쉬운 교환과 접근 그리고 장기보존을 보장하기 위함이다.

부속서의 참고 규격은 2020년 현재 유럽연합이 채택하고 있는 최신 규격이며 현재 버전은 2.1이다. 세부 내용은 필요에 따라 eCH의 전문가 그룹에 의해 개정될 수 있으므로 최신 규격은 eCH 홈페이지를 참조한다.

국가기록원은 행정정보 데이터세트를 영구기록물관리기관으로 이관할 경우 SIARD 규격을 참고하여 표준화된 방식으로 이관·보존·활용할 것을 권장한다.

B.2 문서의 구조

B.2.1 장의 구조

이 명세서의 각 장은 동일한 패턴에 따라 구성되어 있다.

ID	요구 조건의 기술	의무/선택사항(M/O)
요구 조건의 ID 기재	요구 조건의 문구 기재	의무/선택사항 여부를 명기

종종 요구사항에 관하여 다음과 같이 특정된 권고, 주석, 예시를 통한 추가 설명이 있다.

ID	요구 조건의 기술	M/O
A_3.1-1	요구 조건의 문구 예시 예시 문구 노트 노트 문구 권고 권고 문구(이탤릭체)	M

B.2.2 요구조건에 대한 ID

ID를 통해 명확하게 요구조건을 구분한다.

ID
G_2.2-1

ID는 다음과 같은 패턴에 의해 구성된다:

G_ Letter + _ 주요 장(chapter)
 G_ = 주요 요구조건 / 원칙
 T_ = 테이블 데이터의 요구조건
 M_ = 메타 데이터의 요구조건
 P_ = 패키지 구조의 요구조건

2.1-1 숫자는 각 장의 번호로 시작하며(각 장은 같은 주제의 요구조건 함께 묶는다), 대시 다음의 숫자는 연속적이기에 각 장의 모든 요구조건을 나타낸다.

B.2.3 필수적 요구조건과 선택적 요구조건의 구분

각 요구조건은 필수적이거나 선택적이다. 이는 약어에 의해 표시된다:

약어	의미
M	필수적 요구조건 이는 유효한 SIARD 파일을 얻기 위해 반드시 충족해야 하는 요구조건이다.
O	선택적 요구조건 이 요구조건은 충족해야 한다. 이는 조작을 간단하게 해주며, 우수한 실행을 이루어준다.

B.2.4 폴더, 파일, 폴더 구조의 표기법

다음의 기호와 변수는 폴더와 파일 등의 표기에 사용한다.

기호	의미
/	폴더
header/	"header"란 이름을 가진 폴더
xy.txt	파일("txt" 파일 확장자를 가진)
dir/	샘플 폴더(적색 표기)
abc.pdf	샘플 파일(적색 표기)
...	설명과 관계없는 파일 또는 폴더의 자리표시자(placeholder)
[]	문자열("string"), 정수("interger") 등의 표현 또는 기본 타입의 자리표시자
<xx>	원하는 문자열 스트링(string of characters)에 대한 자리표시자

B.3 일반 요구조건 / 원칙

B.3.1 표준의 사용

장기간 DB 콘텐츠를 해석할 수 있도록 유지하는 것을 보장하기 위해, SIARD 포맷은 기본적으로 두 개의 ISO 표준, XML과 SQL:2008에 기반을 둔다.

ID	요구 조건의 기술	M/O
G_3.1-1	모든 DB 콘텐츠는 XML 스키마 1.0 ¹⁾ 의 스키마 정의를 따르는, XML 1.0 포맷 ²⁾ 파일의 집합에 보관된다. 스키마 정의와 SQL 코드는 각 케이스마다 ISO/IEC 9075에 근거한 SQL:2008을 따른다. 분리된 이진 파일과 텍스트 파일로 저장되나, BLOB와 CLOB (이진 대형 객체와 문자 대형 객체) 데이터만 예외로 둔다.	M

B.3.2 문서로서의 데이터베이스

개별 테이블 데이터 간의 관계를 보존하기 위해서, 하나의 RDB는 단일의 아카이브된 문서로 다뤄진다.

ID	요구 조건의 기술	M/O
G_3.2-1	하나의 RDB는 단일의 SIARD 파일로 아카이브된다. 더 큰 의미에서, 이 파일은 DB에 관계된 저장된 대형 객체(LOB)를 참조할 수 있다.	M

B.3.3 문자셋과 문자

ID	요구 조건의 기술	M/O
G_3.3-1	모든 데이터는 ISO 10646에 대응하는 유니코드 문자셋으로 저장된다.	M
G_3.3-2	다른 문자셋을 지원하는 DB에서 추출할 때, 해당하는 유니코드 문자셋으로의 매핑이 이루어진다. 이러한 이유로, DB 제품의 national 문자열 타입(NCHAR, NCHAR VARYING, NCLOB)은 일반적으로 반드시 non-national 타입(CHAR, VARCHAR 또는 CLOB)으로 변환되어야 한다.	M
G_3.3-3	SIARD 포맷의 XML 파일에서, XML 문법 상 특별한 의미를 가지는 모든 문자는 모든 필드에서 xs:string 타입의 단위 참조로 대체된다. 덧붙여, 유니코드 제어 문자 0-31 과 127-159는, XML 파일의 유효성을 보장하기 위해, 사선("\")을 사용하여 코드화된다.	M

1) <https://www.w3.org/TR/xmlschema-1/>, <https://www.w3.org/TR/xmlschema-2/>, <https://www.w3.org/TR/xmlschema-ref/>

2) <https://www.w3.org/TR/REC-xml/>

ID	요구 조건의 기술	M/O																						
G_3.3-4	이스케이프 문자 “\” 뿐만 아니라, 유니코드로 표현될 수 없는 문자들(코드 0-8, 14-31, 127-159)과 다수의 공백 문자는 XML에서 \u00<xx>로서 예외처리된다. 따옴표, 부등호 그리고 앰퍼샌드는 XML에서 엔티티 참조를 의미한다.	M																						
	<table border="1"> <thead> <tr> <th>원본 문자</th> <th>SIARD 포맷 문자</th> </tr> </thead> <tbody> <tr> <td>0 ~ 8</td> <td>\u0000 ~ \u0008</td> </tr> <tr> <td>14 ~ 31</td> <td>\u000E ~ \u001F</td> </tr> <tr> <td>32</td> <td>\u0020, 다수의 공백(for multiple spaces)</td> </tr> <tr> <td>"</td> <td>&quot;</td> </tr> <tr> <td>&</td> <td>&amp;</td> </tr> <tr> <td>'</td> <td>&apos;</td> </tr> <tr> <td><</td> <td>&lt;</td> </tr> <tr> <td>></td> <td>&gt;</td> </tr> <tr> <td>\</td> <td>\u005c</td> </tr> <tr> <td>127 ~ 159</td> <td>\u007F ~ \u009F</td> </tr> </tbody> </table>	원본 문자	SIARD 포맷 문자	0 ~ 8	\u0000 ~ \u0008	14 ~ 31	\u000E ~ \u001F	32	\u0020, 다수의 공백(for multiple spaces)	"	"	&	&	'	'	<	<	>	>	\	\u005c	127 ~ 159	\u007F ~ \u009F	
원본 문자	SIARD 포맷 문자																							
0 ~ 8	\u0000 ~ \u0008																							
14 ~ 31	\u000E ~ \u001F																							
32	\u0020, 다수의 공백(for multiple spaces)																							
"	"																							
&	&																							
'	'																							
<	<																							
>	>																							
\	\u005c																							
127 ~ 159	\u007F ~ \u009F																							

B.3.4 파일 URI 스키마

저장된 LOB를 외부적으로 참조하기 위해 사용되는, 파일 URI 스키마는 RFC 1738³⁾에 의해 정의된다.

ID	요구 조건의 기술	M/O
G_3.4-1	모든 외부 참조 파일은, RFC 1738에 명시된 파일 URI를 이용하여, 명시된다.	M
G_3.4-2	파일 URI는 URL로 인코딩 된 ASCII 문자열로서 SIARD 아카이브에 저장된다.	M
G_3.4-3	파일 URI가 ZIP파일 내의 개별 엔트리를 지정할 있는, 파일시스템에 근거하고 있다면, 외부적으로 저장된 LOB는 ZIP파일로 통합될 수 있다.	O

3) http://en.wikipedia.org/wiki/File_URI_scheme, <http://tools.ietf.org/html/rfc1738>

B.3.5 식별자와 일반(regular) 식별자

SQL:2008에는 대소문자를 구분하지 않지만 SIARD 아카이브에는 대문자로 저장되는 공백과 특수문자를 제외한 일반 식별자⁴⁾와, 대소문자를 구분하며 특수문자 또는 SQL 키워드 '='(equal) 을 포함하는 한정(delimited) 식별자가 있다. 이들은 표현식에서 큰 따옴표로 묶여야 한다. SIARD 아카이브에서는 따옴표 없이 저장된다.

SQL 표준에서 특수문자 또는 키워드를 구성하는 것의 정의가 설정된다. 글자의 대문자는 유니코드 표준에 의해 정의된다.

메타데이터에서, 일반 식별자는 대문자로 저장되고 다른 모든 식별자들이 따옴표 없이 저장된다. SQL:2008 표준은 특정 식별자가 일반 식별자에서 허용되지 않거나 SQL 키워드와 동일한 문자를 포함하는 즉시, 이를 한정 식별자로 간주할 것을 규정한다.

ID	요구 조건의 기술	M/O
G_3.5-1	모든 식별자는 유니코드 문자셋으로 저장된다.	M
G_3.5-2	일반 식별자는 따옴표가 없고, 대문자이다.	M
G_3.5-3	한정 식별자는 따옴표 없이 저장된다.	M

B.4 포맷구조를 위한 요건

B.4.1 SIARD 아카이브 파일의 구성

SIARD 아카이브 파일은 ZIP 아카이브로 구현된다.

ID	요구 조건의 기술	M/O
G_4.1-1	SIARD 아카이브 파일은 PkWare사에서 발행한 6.3.2 버전 ⁵⁾ 의 명세를 따른, 단일의 ZIP 아카이브에 저장된다.에 기술된 "deflate" 알고리즘을 사용하여 압축되어져야 한다. 권고 <i>deflate 알고리즘을 사용할 것을 권장한다.</i>	M
G_4.1-2	SIARD 파일은 압축되지 않거나, RFC 1951 ⁶⁾	M
G_4.1-3	SIARD 파일은 암호로 보호되거나 암호화되지 않는다.	M
G_4.1-4	ZIP32 및 ZIP64 ⁷⁾ 모두 ZIP 아카이브에 허용된다.	M
G_4.1-5	ZIP 아카이브는 ".siard" 파일 확장자를 갖는다.	M

4) SQL:2008의 식별자는 최대 128자로, 문자(A-Z) 또는 언더스코어(_)로 시작해야 하며, 이후에 문자(A-Z), 숫자(0-9) 또는 언더스코어(_)로 구성된다.

5) ZIP 파일은 본래 Phil Katz에 의해 정의되었으며, 현재 사실상 표준으로 광범위하게 사용된다. PkWare에서 발행된 현재 버전 6.3.2의 명세는 <https://support.pkware.com/display/PKZIP/Application+Note+Archives> 에서 찾을 수 있다.

6) <https://www.ietf.org/rfc/rfc1951.txt>

7) ※(NAK)ZIP32 and ZIP64 variants

B.4.2 SIARD 아카이브 파일의 구조

SIARD 포맷으로 아카이브된 RDB는 두 가지 구성요소로 구성된다: 아카이브된 DB의 구조를 기술하는 메타데이터와, 테이블의 내용을 나타내는 테이블 데이터. 또한 메타데이터는 다양한 테이블 데이터를 아카이브에서 찾을 수 있는 위치를 나타냅니다.

ID	요구 조건의 기술	M/O
P_4.2-1	<p>테이블 데이터는 content/ 폴더에, 메타데이터는 header/ 폴더에 위치한다. 이외의 폴더나 파일은 허용되지 않는다.</p> <p>예시 SIARD 파일의 구조 (도식화)</p> <pre data-bbox="411 674 1066 763"> ech-0165_oe.siard content/ header/ </pre>	M
P_4.2-2	<p>content/ 폴더에는 개별 테이블 폴더들이 위치한, 하나 이상의 스키마 폴더가 있다. 이외의 폴더나 파일은 허용되지 않는다.</p> <p>예시 SIARD 파일의 구조 (도식화)</p> <pre data-bbox="411 931 954 1200"> ech-0165_oe.siard content/ schema0/ table0/ table1/ table2/ ... schema1/ table0/ ... </pre> <p>권고 스키마와 테이블 폴더 이름을 표준화하고, 예를 들어 실제 이름보다는 schema0/ 과 table0/ 을 사용할 것을 권장한다. (P_4.2-5의 제한사항 참조)</p>	M
P_4.2-3	<p>개별 테이블 폴더는 XML 파일과 XSD 파일을 가지고 있으며, 그 이름(폴더 지정 및 두 파일 이름)은 동일해야 한다. BLOB 및 CLOB 폴더와 하위 콘텐츠를 제외하고, 다른 폴더나 파일은 허용되지 않는다.</p> <p>예시 SIARD 파일의 구조 (도식화)</p> <pre data-bbox="411 1529 1034 1854"> ech-0165_oe.siard content/ ... schema1/ ... table6/ table6.xml table6.xsd table7/ table7.xml table7.xsd lob1⁸⁾/ record0.xml record1.xml </pre> <p>권고 lob 폴더와 lob 파일 이름을 표준화하고, 예를 들어 실제 이름 대신 lob1/ 과 record0.bin, record0.txt 또는 record0.xml 을 사용하고, 이 lob 파일의 MIME 타입과 관련된 파일 확장명을 사용할 것</p>	M

	<i>을 권장한다.</i>	
P_4.2-4	SIARD 포맷(예: PRONOM)의 인식을 용이하게 하기 위해, .SIARD 포맷의 버전을 명시하는 빈 폴더 /header/siardversion/2.1/ 가 반드시 존재해야 한다.	M
P_4.2-5	<p>metadata.xml 과 metadata.xsd 파일은 반드시 header/ 폴더에 있어야 한다. 스타일 시트와 같은 추가 파일은 허용된다.</p> <p>예시 SIARD 파일의 구조 (도식화)</p> <pre> ech-0165_oe.siard content/ ... header/ metadata.xml metadata.xsd siardversion/ 2.1/ ... </pre>	M
P_4.2-6	<p>SIARD (ZIP64) 파일 내의 요소를 참조하는 모든 파일과 폴더 이름은 다음과 같이 구조화 되어야 한다: 이름은 반드시 문자 [a-z 또는 A-Z]로 시작하여야 하며, 이후 다음과 같은 문자만 포함할 수 있다.</p> <ul style="list-style-type: none"> ● a-z ● A-Z ● 0-9 ● - ● . (확장자와 이름을 분리할 때에만 사용할 수 있다) <p>권고 가능한 Windows에서 지나치게 긴 경로 길이의 문제를 피하기 위해, 파일 및 폴더 이름의 길이가 20자를 넘지 않아야 한다.</p>	

...

8) 예시에서, 2열(column 2)에는 lob1에 저장되는 추가 lob 파일들이 들어있다.

B.4.3 메타데이터와 테이블 데이터와의 대응

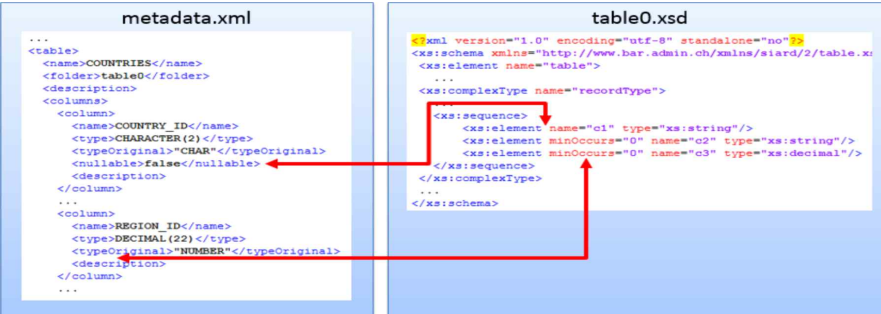
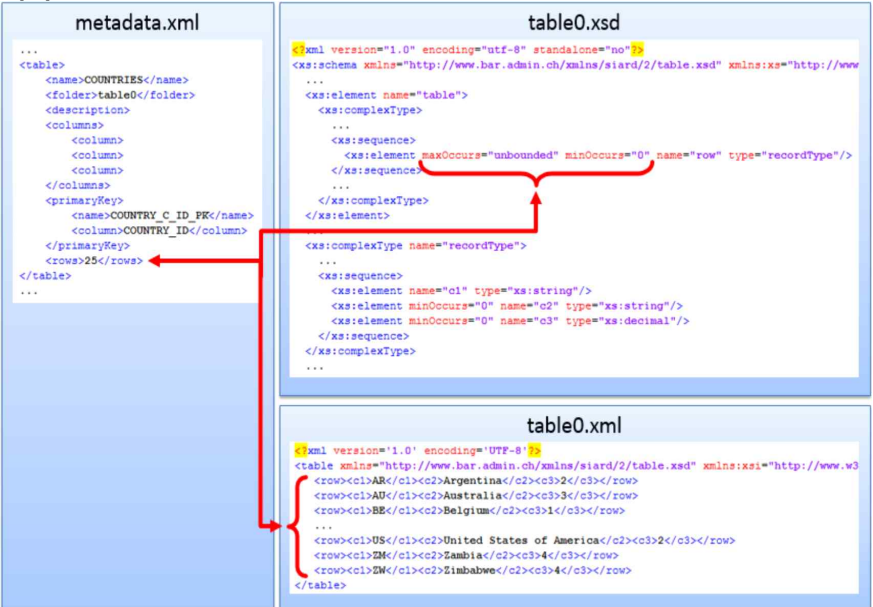
ID	요구 조건의 기술	M/O
P_4.3-1	<p>metadata.xml 에서 지정된 구조는 content/ 폴더의 구조와 동일해야 한다.</p> <p>예시</p> <div style="display: flex; justify-content: space-around;"> <div style="border: 1px solid #ccc; padding: 5px; width: 45%;"> <p style="text-align: center; margin: 0;">metadata.xml</p> <pre style="font-family: monospace; font-size: 0.9em; margin: 0;"> ... <dbname>OE Sample Database enhanced</dbname> ... <schemas> <schema> <name>HR</name> <folder>schema0</folder> <tables> <table> <name>COUNTRIES</name> <folder>table0</folder> ... </table> <table> <name>DEPARTMENTS</name> <folder>table1</folder> ... </table> <table> <name>EMPLOYEES</name> <folder>table2</folder> ... </table> <table> <name>JOB_HISTORY</name> <folder>table3</folder> </pre> </div> <div style="border: 1px solid #ccc; padding: 5px; width: 45%;"> <p style="text-align: center; margin: 0;">folder</p> </div> </div>	M
P_4.3-2	<p>Metadata.xml 에서 지정된 테이블의 열(columns)의 개수는 해당 table[number].xsd 파일의 열 수와 동일해야 한다.</p> <p>예시</p> <div style="display: flex; justify-content: space-around;"> <div style="border: 1px solid #ccc; padding: 5px; width: 45%;"> <p style="text-align: center; margin: 0;">metadata.xml</p> <pre style="font-family: monospace; font-size: 0.9em; margin: 0;"> ... <table> <name>COUNTRIES</name> <folder>table0</folder> <description> <columns> <column> <column> <column> </columns> <primaryKey> <rows>25</rows> </table> ... </pre> </div> <div style="border: 1px solid #ccc; padding: 5px; width: 45%;"> <p style="text-align: center; margin: 0;">table0.xsd</p> <pre style="font-family: monospace; font-size: 0.9em; margin: 0;"> <?xml version="1.0" encoding="utf-8" standalone="no"?> <xs:schema xmlns="http://www.bar.admin.ch/xmlns/siard/2/table.x <xs:element name="table"> ... <xs:complexType name="recordType"> ... <xs:sequence> <xs:element name="c1" type="xs:string"/> <xs:element minOccurs="0" name="c2" type="xs:string"/> <xs:element minOccurs="0" name="c3" type="xs:decimal"/> </xs:sequence> </xs:complexType> ... </xs:schema> </pre> </div> </div>	

<p>P_4.3-3</p>	<p>Metadata.xml 의 열 정의에 대한 데이터 타입 정보는 해당 table[number].xsd 파일의 데이터 타입 정의와 동일해야 한다.</p> <p>SQL:2008 내장 데이터 타입⁹⁾은 다음 표에 따라 table[번호].xsd 스키마 파일에서 XML 타입으로 변환된다.</p> <table border="1" data-bbox="363 481 1259 1527"> <thead> <tr> <th>SQL:2008</th> <th>XML</th> </tr> </thead> <tbody> <tr><td>BIGINT</td><td>xs:interger</td></tr> <tr><td>BINARY LARGE OBJECT(...), BLOB(...)</td><td>blobType¹⁰⁾</td></tr> <tr><td>BINARY VARYING(...), VARBINARY(...)</td><td>xs:hexBinary / clobType¹⁴⁾</td></tr> <tr><td>BINARY(...)</td><td>xs:hexBinary / blobType</td></tr> <tr><td>BOOLEAN</td><td>xs:boolean</td></tr> <tr><td>CHARACTER LARGE OBJECT(...), CLOB(...)</td><td>clobType¹⁴⁾</td></tr> <tr><td>CHARACTER VARYING(...), CHAR VARYING(...), VARCHAR(...)</td><td>xs:string / clobType¹⁴⁾</td></tr> <tr><td>CHARACTER(...), CHAR(...)</td><td>xs:string / clobType¹⁴⁾</td></tr> <tr><td>DATE</td><td>dateType</td></tr> <tr><td>DECIMAL(...), DEC(...)</td><td>xs:decimal</td></tr> <tr><td>DOUBLE PRECISION</td><td>xs:double</td></tr> <tr><td>FLOAT(p)</td><td>xs:double</td></tr> <tr><td>INTEGER, INT</td><td>xs:integer</td></tr> <tr><td>INTERVAL <start> [TO <end>]</td><td>xs:duration</td></tr> <tr><td>NATIONAL CHARACTER LARGE OBJECT(...), NCHAR LARGE OBJECT(...), NCLOB(...)</td><td>clobType¹⁴⁾</td></tr> <tr><td>NATIONAL CHARACTER VARYING(...), NATIONAL CHAR VARYING(...), NCHAR VARYING(...)</td><td>xs:string / clobType¹⁴⁾</td></tr> <tr><td>NATIONAL CHARACTER(...), NCHAR(...), NATIONAL CHAR(...),</td><td>xs:string / clobType¹⁴⁾</td></tr> <tr><td>NUMERIC(...)</td><td>xs:decimal</td></tr> <tr><td>REAL</td><td>xs:float</td></tr> <tr><td>SMALLINT</td><td>xs:integer</td></tr> <tr><td>TIME(...)</td><td>timeType</td></tr> <tr><td>TIME WITH TIME ZONE(...)</td><td>timeType</td></tr> <tr><td>TIMESTAMP(...)</td><td>dateTimeType</td></tr> <tr><td>TIMESTAMP WITH TIME ZONE(...)</td><td>dateTimeType</td></tr> <tr><td>XML</td><td>clobType¹⁴⁾</td></tr> </tbody> </table>	SQL:2008	XML	BIGINT	xs:interger	BINARY LARGE OBJECT(...), BLOB(...)	blobType ¹⁰⁾	BINARY VARYING(...), VARBINARY(...)	xs:hexBinary / clobType ¹⁴⁾	BINARY(...)	xs:hexBinary / blobType	BOOLEAN	xs:boolean	CHARACTER LARGE OBJECT(...), CLOB(...)	clobType ¹⁴⁾	CHARACTER VARYING(...), CHAR VARYING(...), VARCHAR(...)	xs:string / clobType ¹⁴⁾	CHARACTER(...), CHAR(...)	xs:string / clobType ¹⁴⁾	DATE	dateType	DECIMAL(...), DEC(...)	xs:decimal	DOUBLE PRECISION	xs:double	FLOAT(p)	xs:double	INTEGER, INT	xs:integer	INTERVAL <start> [TO <end>]	xs:duration	NATIONAL CHARACTER LARGE OBJECT(...), NCHAR LARGE OBJECT(...), NCLOB(...)	clobType ¹⁴⁾	NATIONAL CHARACTER VARYING(...), NATIONAL CHAR VARYING(...), NCHAR VARYING(...)	xs:string / clobType ¹⁴⁾	NATIONAL CHARACTER(...), NCHAR(...), NATIONAL CHAR(...),	xs:string / clobType ¹⁴⁾	NUMERIC(...)	xs:decimal	REAL	xs:float	SMALLINT	xs:integer	TIME(...)	timeType	TIME WITH TIME ZONE(...)	timeType	TIMESTAMP(...)	dateTimeType	TIMESTAMP WITH TIME ZONE(...)	dateTimeType	XML	clobType ¹⁴⁾	<p>M</p>
SQL:2008	XML																																																					
BIGINT	xs:interger																																																					
BINARY LARGE OBJECT(...), BLOB(...)	blobType ¹⁰⁾																																																					
BINARY VARYING(...), VARBINARY(...)	xs:hexBinary / clobType ¹⁴⁾																																																					
BINARY(...)	xs:hexBinary / blobType																																																					
BOOLEAN	xs:boolean																																																					
CHARACTER LARGE OBJECT(...), CLOB(...)	clobType ¹⁴⁾																																																					
CHARACTER VARYING(...), CHAR VARYING(...), VARCHAR(...)	xs:string / clobType ¹⁴⁾																																																					
CHARACTER(...), CHAR(...)	xs:string / clobType ¹⁴⁾																																																					
DATE	dateType																																																					
DECIMAL(...), DEC(...)	xs:decimal																																																					
DOUBLE PRECISION	xs:double																																																					
FLOAT(p)	xs:double																																																					
INTEGER, INT	xs:integer																																																					
INTERVAL <start> [TO <end>]	xs:duration																																																					
NATIONAL CHARACTER LARGE OBJECT(...), NCHAR LARGE OBJECT(...), NCLOB(...)	clobType ¹⁴⁾																																																					
NATIONAL CHARACTER VARYING(...), NATIONAL CHAR VARYING(...), NCHAR VARYING(...)	xs:string / clobType ¹⁴⁾																																																					
NATIONAL CHARACTER(...), NCHAR(...), NATIONAL CHAR(...),	xs:string / clobType ¹⁴⁾																																																					
NUMERIC(...)	xs:decimal																																																					
REAL	xs:float																																																					
SMALLINT	xs:integer																																																					
TIME(...)	timeType																																																					
TIME WITH TIME ZONE(...)	timeType																																																					
TIMESTAMP(...)	dateTimeType																																																					
TIMESTAMP WITH TIME ZONE(...)	dateTimeType																																																					
XML	clobType ¹⁴⁾																																																					

9) BIT 및 BIT VARYING은 이전 SQL 정의에서 비롯된 데이터 타입이며, SQL:2008의 BOOLEAN 및 BINARY 로 대체되었다. BIT(1)은 BOOLEAN으로, BIT(n)은 BINARY((n+7/8))으로 변환된다. ※(NAK)오타유의: BINARY((n+7/8))

10) XML 데이터 타입인 blobType 및 clobType 의 경우, G_3.1-1 참조.

<p>P_4.3-3</p>	<p>예시</p> <div style="display: flex; justify-content: space-around;"> <div style="border: 1px solid #add8e6; padding: 5px; width: 45%;"> <p style="text-align: center; margin: 0;">metadata.xml</p> <pre style="margin: 0;"> ... <table> <name>COUNTRIES</name> <folder>table0</folder> <description> <columns> <column> <name>COUNTRY_ID</name> <type>CHARACTER(2)</type> <typeOriginal>CHAR</typeOriginal> <nullable>false</nullable> <description> </column> ... <column> <name>REGION_ID</name> <type>DECIMAL(22)</type> <typeOriginal>NUMBER</typeOriginal> <description> </column> ... </pre> </div> <div style="border: 1px solid #add8e6; padding: 5px; width: 45%;"> <p style="text-align: center; margin: 0;">table0.xsd</p> <pre style="margin: 0;"> <?xml version="1.0" encoding="utf-8" standalone="no" ?> <xs:schema xmlns="http://www.bar.admin.ch/xmns/siard/2/table.x" ... <xs:element name="table"> ... <xs:complexType name="recordType"> <xs:sequence> <xs:element name="c1" type="xs:string"/> <xs:element minOccurs="0" name="c2" type="xs:string"/> <xs:element minOccurs="0" name="c3" type="xs:decimal"/> </xs:sequence> </xs:complexType> ... </xs:schema> </pre> </div> </div> <p>모든 날짜와 시간 타입은 UTC "시간대(time zone)"로 지정됩니다. 그것들은 "Z"로 끝나는 것이 권장된다. dateType은 UTC의 0001부터 9999사이의 연도에 대한 xs:date의 제한사항이다. (SQL:2008 제한사항). timeType은 UTC "시간대"에 대한 xs:time의 제한사항이다. dateTimeType은 UTC "시간대"의 0001부터 9999사이의 연도에 대한 xs:dateTime의 제한사항이다. (SQL:2008 제한사항). clobType은 xs:string의 확장이다. CLOB 값이 직접 저장되는 인라인 값에는 추가 속성이 필요하지 않다. CLOB 값이 별도의 파일에 저장되면 file 및 length 속성은 필수사항, digestType과 digest 속성은 선택사항이다. 파일 속성의 값은 (URL 인코딩된) 파일 URI(가능한 가장 인접한 lobFolder와 관련된)이며, 여기에 CLOB이 저장된다. length 속성의 값은 length in(UTF-8) 이며, optional messageDigest 속성은 동등한 optional digestType 속성에 따라 무결성 정보를 포함한다. BlobType은 xs:hexBinary의 확장이다. BLOB 값이 직접 저장되는 인라인 값에는 추가 속성이 필요하지 않다. BLOB 값이 직접 저장되는 인라인 값에는 추가 속성이 필요하지 않다. BLOB 값이 별도의 파일에 저장되면 file 및 length 속성은 필수사항, digestType과 digest 속성은 선택사항이다. 파일 속성의 값은 (URL 인코딩된) 파일 URI(가능한 가장 인접한 lobFolder와 관련된)이며, 여기에 BLOB이 저장된다. length 속성의 값은 length in bytes 이며, optional messageDigest 속성은 동등한 optional digestType 속성에 따라 무결성 정보를 포함한다.</p>	
<p>P_4.3-4</p>	<p>명명된 DISTINCT 데이터 타입은 기본 타입을 나타내는데 사용되는, table[번호].xsd 스키마 파일에서 XML 데이터 타입으로 변환된다.</p>	<p>M</p>
<p>P_4.3-5</p>	<p>배열은 table[번호].xsd 스키마 파일에서 배열의 베이스 타입에 대응하는 XML 데이터 타입으로 변환되는, 일련의 구조화된 XML요소인 <a1>, <a2>, ... 으로 변환된다.</p> <p>예시 appendix D.3c.의 table0.xsd 예시</p>	<p>M</p>
<p>P_4.3-6</p>	<p>명명된 사용자 정의 데이터 타입(UDT)는 table[number].xsd 스키마 파일에서 각 속성 타입에 대응하는 XML 데이터 타입으로 변환되는, 일련의 구조화된 XML 요소 <u1>, <u2>, ... 로 변환된다.</p> <p>예시 appendix D.3c.의 table0.xsd 예시</p>	<p>M</p>

<p>P_4.3-7</p>	<p>Metadata.xml의 열 정의에 대한 nullable 정보는 대응하는 table[number].xsd 파일의 정보와 동일해야 한다.</p> <p>예시</p>  <p>노트 SQL:2008의 "<nullable>true</nullable>"는 XML에서 "minOccurs="0""이 된다. "<nullable>false</nullable>"는 XML에서 "minOccurs="1""에 대응한다. 하지만, "minOccurs="1""이 기본값이기 때문에, 종종 생략된다. "<nullable>"의 표시가 없다는 것은 "<nullable>true</nullable>"를 의미한다.</p>	<p>M</p>
<p>P_4.3-8</p>	<p>metadata.xml의 열 순서는 대응하는 table[number].xsd의 열 순서와 동일해야 한다.</p>	<p>M</p>
<p>P_4.3-9</p>	<p>metadata.xml의 테이블 정의에 있는 필드 시퀀스는 대응하는 table[number].xsd의 필드 시퀀스와 동일해야 한다.</p>	<p>M</p>
<p>P_4.3-10</p>	<p>metadata.xml의 테이블에 있는 행 수는 대응하는 table[number].xml의 행 수와 동일해야 한다.</p> <p>metadata.xml의 테이블에 있는 행 수는 대응하는 table[number].xsd에 지정된 영역에 맞아야 한다.</p> <p>예시</p>  <p>권고 table[number].xsd에서 범위는 0에서 무한대를 사용할 것을 권장한다. (maxOccurs="unbounded" minOccurs="0"). 이는 table[number].xsd에 대해 table[number].xml의 유효성을 확인할 때 문제를 피할 수 있다.</p>	<p>M</p>

B.5 메타데이터 요구조건

SIARD 아카이브의 메타데이터는 아카이브된 DB의 구조를 저장하고, 아카이브에서 다른 테이블 데이터가 있는 위치를 나타낸다

모든 메타데이터는 header/폴더 안의 단일 metadata.xml로 통합(collated)된다. 이 파일은 계층적 구조를 갖는다.

metadata.xsd는 metadata.xml 파일에 대한 스키마 정의이다. 이것 또한 header/폴더에 저장된다.

ID	요구 조건의 기술	M/O
M_5.0-1	metadata.xml 파일은 스키마 정의 metadata.xsd를 준수해야 한다.	M

개별 레벨의 내용은 아래에 정의되어 있다.

B.5.1 DB 레벨 메타데이터

metadata.xml 파일은 DB 레벨에서 다음의 전역 정보를 담고있다:

ID	요구 조건의 기술	M/O
M_5.1-1	DB 레벨의 metadata.xsd에서 필수로 지정된 모든 메타데이터는 다음에 따라 작성되어야 한다.	M

다음의 DB 메타데이터는 metadata.xml 파일에 저장된다:

식별자	뜻	M/O
version	SIARD 포맷 버전	M
dbname	Short DB 식별자	M
description	전체적인 DB의 의미(meaning)와 내용(content)에 대한 기술	O
archiver	DB로부터 테이블 데이터의 보존을 수행한 사람의 이름	O
achiverContact	DB로부터 테이블 데이터의 아카이빙을 수행한 사람의 연락처 상세 (전화, 이메일)	O
dataOwner	DB 내의 데이터 소유자; 아카이빙 당시, 데이터 사용권한을 부여할 권리가 있고 데이터 보호 가이드라인과 같은 법적 의무를 준수할 책임이 있는 기관이나 인물	M
dataOrigin-Timespan	DB에 있는 데이터의 시작기간; 텍스트 형식으로 근사 표시	M
lobFolder	LOB(Large Object)의 외부 저장 위치를 나타내는 상대 URL에 대한 기본 URI를 가리키는 "file:" URI. 누락된 경우, ZIP 파일 내의 root 폴더의 기본 값이 사용된다. 열 메타데이터 내의 상대 lobFolder URI는 이 값과 연계된다. 노트 "file:" URI가 ZIP 파일이 폴더로 취급되는 확장 파일 시스템을 나타내는 경우, 상대 URI "..."은 SIARD 아카이브가 있는 외부 폴더를 참조합니다. 이러한 파일 시스템 확장이 지원되지 않으면, 절대 "file:" URI가 LOB파일의 외부 저장 위치를 지정하는데 이용되어야 한다. 모든 열의 lobfolder 엔트리와 모든 LOB 파일 속성은 상대 URL로 선택하는 것이 강력하게 권장된다. 따라서 SIARD 파일이나 이 파일의 정보 패키지의 재배포시, 이 전역 URI만 새 위치를 가리키도록 변경해야 한다.	O

producerApplication	SIARD 파일을 다운로드한 어플리케이션의 이름과 버전.	O
archivalDate	테이블 데이터가 보관된 ¹¹⁾ 날짜	M
messageDigest	<p>digestType (MD5, SHA-1, 또는 SHA-256) 및 대응하는 digest의 구성. Digest는 이진 버퍼를 16진수 문자열, 또는 SHA-1 이나 SHA-256인 경우에는 base64 문자열로서 나타낸다. 16진수나 base64 문자열이 사용되는 지의 여부는 digest와 문자열의 길이에 따라 달라진다.</p> <p>digest는 content/ 폴더를 통해 계산된다. (서로 다른 알고리즘에 기반한) 하나 이상의 digest 가 저장될 수 있다. Message digest가 저장되지 않으면, SIARD 파일 외부에 message digest 같은 것을 저장하여 무결성을 보장해야 한다.¹²⁾</p> <p>예시 appendix D.2.의 metadata.xml 예시</p> <p>권고 message digest 옵션을 사용하는 경우, 다음을 구현해야 한다: content와 header 디렉토리는 분리된 (비어있는) content/ 및 header/ 엔트리로, ZIP 파일 내에 저장된다. 기본 데이터의 무결성을 보장하기 위해서, header 디렉토리의 엔트리는 content/ 엔트리의 모든 기본 데이터 뒤에, 남은 모든 메타데이터 엔트리 앞에 삽입되어야 한다. 아래에 언급된 message digest는 offset 0부터</p>	O
clientMachine	보존을 수행한 (클라이언트) 컴퓨터의 DNS 이름	O
databaseProduct	테이블 데이터의 보존이 수행된 DB 제품 및 버전	O
connection	테이블 데이터를 보존하는데 사용된 연결 문자열(connection string)	O
databaseUser	DB에서 테이블 데이터를 아카이빙하는 SIARD 도구(tool) 사용자의 DB 유저 ID	O
schemas	DB의 스키마 목록	M
users	DB 유저 목록	M
roles	DB 역할 목록	O
privileges	유저 및 역할 권한 목록	O

11) (NAK)원문: Date on which the table data were archived

12) SIARD 파일을 위조하는 공격자에 의해 위조될 수 있기 때문에, SIARD 파일 내에 저장된 message digest 코드는 그 자체로 무결성을 보장하지 않는다. Message digest 코드를 외부에 저장해야만 이를 피할 수 있다. 그러나 다운로드시, 내부적으로 메타데이터 독립적인 message digest 코드를 생성함으로써 외부 저장소를 지원할 수 있다.

B.5.2 스키마 레벨 메타데이터

스키마 메타데이터는 DB의 전역 정보와 같이 metadata.xml 파일에 보관된다.¹³⁾

ID	요구 조건의 기술	M/O
M_5.2-1	스키마 레벨에서 metadata.xsd 안에서 필수로 지정된 모든 메타데이터는 다음에 따라 작성되어야 한다.	M

다음의 속성 메타데이터는 metadata.xml 파일에 저장된다.

식별자	뜻	M/O
name	DB의 스키마 이름	M
folder	SIARD 아카이브 내 content/ 하위의 스키마 폴더 이름	M
types	스키마의 (명명된) 고급 또는 구조화된 타입 목록	O
description	스키마의 의미(meaning)와 내용(content)에 대한 기술	O
tables	스키마의 테이블 목록	O
views	스키마에 저장된 뷰 목록	O
routines	스키마의 루틴(이전의 "저장된 프로시저") 목록	O

B.5.3 타입 레벨 메타데이터

ID	요구 조건의 기술	M/O
M_5.3-1	metadata.xml 안에서 보관될 수 있는 ¹⁴⁾ 스키마의 타입 메타데이터	O

고급 또는 구조화된 데이터 타입이 보관되는¹⁵⁾ 경우, 다음의 타입 메타데이터가 metadata.xml 파일에 저장된다.

식별자	뜻	M/O
name	스키마의 타입이름	M
category	고급 또는 구조화된 타입의 카테고리 ("distinct" 또는 "udt").	M
underSchema	타입이 super 타입을 기반으로 할 때, super 타입의 스키마 이름	O
underType	타입이 super 타입을 기반으로 할 때, super 타입의 이름	O
instantiable	타입이 인스턴스화 될 수 있다면 True, 아니면 False	M
final	이 타입에 하위 타입을 생성할 수 없다면 True, 아니면 False	M
base	카테고리가 "distinct"인 경우, 기 정의된 기본 타입의 이름	O
attributes	카테고리가 "udt"인 경우, 속성 목록	O
description	데이터 타입의 의미(meaning)와 내용(content)에 대한 기술	O

13) (NAK)원문: ~ are archived in ~

14) (NAK)원문: ~ can be archived in ~

15) (NAK)원문: ~ is archived:

B.5.4 속성 레벨 메타데이터

ID	요구 조건의 기술	M/O
M_5.4-1	metadata.xml 에서 필수로 지정된, 속성 레벨의 모든 메타데이터는 다음에 따라 작성되어야 한다.	M

다음의 속성 메타데이터는 metadata.xml 파일에 저장된다:

식별자	뜻	M/O
name	속성의 이름	M
type	SQL:2008 내장 속성 데이터 타입	O
typeOriginal	내장 타입의 원본 열 타입(original column type)	O
	노트 SQL과 호환된다고 기술된 다양한 DB 프로그램들이 매우 다른 데이터 타입을 허용하기 때문에, SQL:2008 타입 뿐만 아니라, 원본 타입도 여기에 나열된다. 독점 타입을 SQL:2008 타입으로 변환하는 것은 SIARD 포맷을 지원하는 각 DB 프로그램에서 정의되고 문서화되어야 한다.	
nullable	속성의 nullable 요소 권고 Nullable 요소의 사용은 권장하지 않는다.	O
typeSchema	고급 또는 구조화된 데이터 타입의 스키마	O
typeName	고급 또는 구조화된 데이터 타입의 이름	O
defaultValue	속성의 기본값	O
description	루틴의 의미와 기능에 대한 기술	O
cardinality	속성이 배열(array)인 경우, 요소의 (최대)수	O

B.5.5 테이블 레벨 메타데이터

DB 및 스키마 메타데이터에 대한 전역 정보와 마찬가지로, 테이블 레벨 메타데이터는 metadata.xml 파일에 보관된다¹⁶⁾.

ID	요구 조건의 기술	M/O
M_5.5-1	metadata.xsd 에서 필수로 지정된 모든 테이블 레벨 메타데이터는 다음에 따라 작성되어야 한다.	M

다음의 테이블 메타데이터는 metadata.xml 파일에 저장된다:

식별자	뜻	M/O
name	스키마의 테이블 이름	M
folder	스키마 폴더 내의 테이블 폴더 이름	M
description	테이블의 의미와 콘텐츠에 대한 기술	O
columns	테이블에 있는 열의 목록	M
primaryKey	테이블의 기본키	O
foreignKeys	테이블에 있는 외래키의 목록	O
candidateKeys	테이블에 있는 후보키의 목록	O
checkConstraints	테이블에 있는 제약조건의 목록	O
triggers	테이블에 있는 트리거의 목록	O
rows	데이터세트의 수	M

16) (NAK)원문: ~ are archived in ~

B.5.6 열(column) 레벨 메타데이터

DB, 스키마 메타데이터, 그리고 테이블 레벨 메타데이터에 대한 전역정보와 마찬가지로, 열 레벨 메타데이터는 metadata.xml 파일에 보관된다¹⁷⁾. 열 메타데이터는 테이블과 뷰의 열을 기술한다.

ID	요구 조건의 기술	M/O
M_5.6-1	metadata.xsd 에서 필수로 지정한 모든 열 레벨 메타데이터는 적절하게 작성되어야 한다.	M

다음의 열 메타데이터는 metadata.xml 파일에 저장된다:

식별자	뜻	M/O
name	테이블 또는 뷰의 열 이름 열 이름은 주어진 테이블 내에서 모호하지 않아야 한다.	M
lobFolder	외부 파일 시스템 내의 상대 또는 절대 "file:"URI로서 작성된 LOB 폴더의 이름. 대형 객체(large objects)의 내부 및 외부 저장에 사용된다. 예시 Appendix D.2.의 metadata.xml 예시 노트 이 엔트리는 열이 LOB 열(예: BLOB, CLOB, 또는 XML 타입)인 경우에만 의미가 있다. 누락된 경우, ""과 같다고 가정한다. 즉, DB 레벨에서 lobFolder와 동일한 폴더를 의미한다. 그렇지 않으면, 이 값은 이 LOB 열의 파일이 저장될 폴더를 가리키는 "file:" - 되도록 상대 - URI 이어야 한다. 이 값이 상대 URI인 경우, DB 레벨의 전역 lobFolder 엔트리에 연계되어 있다고 가정한다. 이 열에 있는 셀의 상대 파일 속성은 이 폴더에 연계된 것으로 해석한다.	O
type	SQL:2008의 기본 제공 데이터 타입 노트 이 열의 데이터 타입이 내장 데이터 타입이라면, 이 필드를 반드시 사용하여야 한다. 그렇지 않으면 typeName 필드가 타입 목록에서 정의된 타입을 반드시 참조하여야 한다.	M
typeOriginal	원본 열 타입(original column type) 노트 SQL을 준수하는(SQLcompliant) 다양한 DB 프로그램들이 매우 다른 데이터 타입을 허용하기 때문에, SQL:2008 타입 뿐만이 아니라 원본 타입이 여기에 나열된다. 독자적(proprietary) 타입을 SQL:2008 타입으로 변환하는 것은 SIARD 포맷을 지원하는 각 DB 프로그램에 대응하는 어플리케이션에서 정의되고, 문서화되어야 한다.	O
nullable	선택적 엔트리	O
typeSchema	(1)열이 내장 데이터 타입이 아니며 (2)명명된 데이터 타입이 이 열의 테이블과 동일한 스키마에서 정의되지 않은 경우에, 명명된 타입의 스키마	O
typeName	이 열의 고급 또는 구조화된 데이터 타입의 이름	O
fields	열이 배열 혹은 "udt" 카테고리의 구조화된 데이터 형식인 경우, 열의 필드 목록	O
defaultValue	열의 기본값	O

17) (NAK)원문: ~ are archived in ~

mimeType	BLOB 열이며, 모든 레코드가 이 열에 동일한 MIME 타입의 파일을 포함하는 경우, 이 열의 MIME 타입. 이 권고 요소는 이진 객체(binary object)에 대한 알맞은 뷰어 선택에 도움이 된다. 이는 수기 혹은 포맷 인식 매커니즘을 사용하는 다운로드 프로그램으로 채워넣을 수 있다.	O
description	열의 의미와 콘텐츠에 대한 기술	O
cardinality	배열인 경우 (최대) 요소 수(number of elements)	O

B.5.7 필드 메타데이터

ID	요구 조건의 기술	M/O
M_5.7-1	열 또는 필드의 필드 메타데이터는 metadata.xml 파일 안에서 보관될 수 있다. ¹⁸⁾	O

다음의 필드 메타데이터는 열 또는 필드가 배열 혹은 “udt” 카테고리의 고급 또는 구조화된 데이터 타입인 경우 metadata.xml 파일에 저장된다.

식별자	뜻	M/O
name	열 또는 필드의 필드 이름. 필드이름은 동일한 열에서 고유한 값이어야 한다. 권고 “udt” 타입 컨테이너(열 또는 필드)의 경우, 필드 이름은 대응하는 속성 이름과 동일해야 한다. 배열 컨테이너의 경우, 필드 이름은 컨테이너 이름 뒤에 대괄호 안에 배열 인덱스가 1부터 시작하여 기재되어야 한다. 예: “Punkt[1]”, “Punkt[2]”, 등등.	M
lobFolder	외부 파일 시스템 내의 상대 또는 절대 “file:” URI로서 작성된 LOB 폴더의 이름. 대형 객체(large objects)의 내부 및 외부 저장에 사용된다. 노트 이 엔트리는 필드가 LOB 필드(예: BLOB, CLOB, 또는 XML 타입)인 경우에만 의미가 있다. 누락된 경우, “”과 같다고 가정한다. 즉, 동봉한(enclosing)(?) 열 또는 필드 요소의 lobFolder 요소를 의미한다. 그렇지 않으면, 이 값은 이 LOB 필드의 파일이 저장될 폴더를 가리키는 “file:” - 되도록 상대 - URI 이어야 한다. 이 값이 상대 URI인 경우, 동봉한(enclosing)(?) 요소의 lobFolder 요소에 연계되어 있다고 가정한다. 이 열에 있는 셀의 상대 파일 속성은 이 폴더에 연계된 것으로 해석한다.	O
fields	필드가 배열 혹은 “udt” 카테고리의 구조화된 데이터 형식인 경우, 필드의 필드 목록	O
mimeType	BLOB 필드이며, 모든 레코드가 이 필드에 동일한 MIME 타입의 파일을 포함하는 경우, 이 필드의 MIME 타입. 이 권고 요소는 이진 객체(binary object)에 대한 알맞은 뷰어 선택에 도움이 된다. 이는 수기 혹은 포맷 인식 매커니즘을 사용하는 다운로드 프로그램으로 채워넣을 수 있다.	O
description	필드의 의미와 내용에 대한 기술	O

18) (NAK)원문: ~ can be archived in ~

B.5.8 기본 키 메타데이터

기본 키는 레코드를 식별하는 고유 키로 정의된다.

ID	요구 조건의 기술	M/O
M_5.8-1	테이블의 기본 키 메타데이터는 metadata.xml 파일에 보관될 수 있다. ¹⁹⁾	O

기본 키가 보관되는²⁰⁾ 경우, 다음의 기본 키 메타데이터가 metadata.xml 파일에 저장된다.

식별자	뜻	M/O
name	기본 키의 이름	M
column	기본 키의 열 목록	M
description	기본 키의 의미와 내용에 대한 기술	O

B.5.9 외래 키 메타데이터

ID	요구 조건의 기술	M/O
M_5.9-1	테이블의 외래 키 메타데이터는 metadata.xml 파일에 보관될 수 있다. ^{23,23)}	O

외래 키가 보관되는²⁴⁾ 경우, 다음의 외래 키 메타데이터가 metadata.xml 파일에 저장된다.

식별자	뜻	M/O
name	외래 키의 이름	M
referenced-Schema	참조된 테이블의 스키마	M
referencedTable	참조된 테이블 노트 참조된 외부 테이블 이름은 테이블 혹은 스키마 테이블 타입의 이름일 수 있다. 구분 식별자는 따옴표로 묶인다.	M
reference	열 및 참조된 열의 목록	M
matchType	매치 타입 (FULL, PARTIAL, 또는 SIMPLE)	O
deleteAction	작업 삭제, 예: CASCADE 노트 삭제 및 변경 작업은 SQL:2008 표준에서 허용하는 작업을 포함한다.	O
updateAction	작업 변경, 예: SET DEFAULT	O
description	외래 키의 의미와 내용에 대한 기술	O

19) (NAK)원문: ~ can be archived in ~

20) (NAK)원문: ~ is(be) archived:

B.5.10 참조 메타데이터

ID	요구 조건의 기술	M/O
M_5.10-1	외래 키에 사용되는 참조 메타데이터는 metadata.xml 파일에 보관될 ²¹⁾ 수 있다.	O

외래 키가 보관되는²²⁾ 경우, 다음의 참조 메타데이터가 metadata.xml 파일에 저장된다.

식별자	뜻	M/O
column	열의 이름	M
referenced	참조된 열의 이름	M

B.5.11 후보 키 메타데이터

후보 키는 기본 키의 후보가 되는 고유한 키로 정의된다. 기본 키 및 후보 키 타입은 metadata.xsd의 동일한 타입인 uniqueKeyType이다. 따라서, 후보 키에 대한 요구 조건은 기본 키에 대한 요구 사항과 동일하다. (M_5.8-1 참조)

ID	요구 조건의 기술	M/O
M_5.11-1	테이블의 후보 키 메타데이터는 metadata.xml 파일에 보관될 수 있다. ²⁵⁾	O

기본 키가 보관되는²⁶⁾ 경우, 다음의 기본 키 메타데이터가 metadata.xml 파일에 저장된다.

식별자	뜻	M/O
name	후보 키의 이름	M
column	후보 키의 열 목록	M
description	후보 키의 의미와 내용에 대한 기술	O

B.5.12 Check 제약조건(check constraint) 메타데이터

Check 제약조건은 검사할 조건으로 구성된다. 이는 SQL:2008 구문에서 BOOLEAN 표현식(*true*, *false* 또는 *unknown*)으로 표시된다.

ID	요구 조건의 기술	M/O
M_5.12-1	테이블의 check 제약조건 메타데이터는 metadata.xml 파일에 보존될 수 있다. ²³⁾	O

다음의 check 제약조건 메타데이터는 check 제약조건이 보존된²⁴⁾ 경우, metadata.xml 파일에 저장된다.

21) (NAK)원문: ~ can be archived in ~

22) (NAK)원문: ~ is archived:

23) (NAK)원문: ~ can be archived in ~

24) (NAK)원문: ~ is archived:

식별자	뜻	M/O
name	check 제약조건의 이름	M
column	check 제약조건의 열 목록	M
description	check 제약조건의 의미와 내용에 대한 기술	O

B.5.13 트리거 레벨 메타데이터

ID	요구 조건의 기술	M/O
M_5.13-1	테이블의 트리거 메타데이터는 metadata.xml 파일에 보존될 수 있다. ²⁷⁾	O

다음의 트리거 메타데이터는 트리거가 보존된²⁸⁾ 경우, metadata.xml 파일에 저장된다.

식별자	뜻	M/O
name	테이블의 트리거 이름	M
actionTime	BEFORE, AFTER 또는 INSTEAD OF	M
triggerEvent	INSERT, DELETE, UPDATE [OF <트리거 열 목록>]	M
aliasList	<이전 값 또는 새 값 별칭 목록>	O
triggeredAction	<트리거된 동작>	M
description	트리거의 의미와 내용에 대한 기술	O

B.5.14 뷰 레벨 메타데이터

ID	요구 조건의 기술	M/O
M_5.14-1	스키마의 뷰 메타데이터는 metadata.xml 파일에 보존될 ²⁵⁾ 수 있다.	O

다음의 뷰 메타데이터는 뷰가 보존된²⁶⁾ 경우, metadata.xml 파일에 저장된다.

식별자	뜻	M/O
name	스키마의 뷰 이름	M
columns	뷰의 열 이름 목록 노트 뷰의 열 메타데이터는 표의 열 메타데이터와 동일하게 구성된다.	M
query	뷰를 정의하는 SQL:2008 쿼리	O
queryOriginal	뷰를 정의하는 원본 SQL 쿼리 노트 SQL을 준수하는(SQLcompliant) 다양한 DB 프로그램들이 매우 다른 쿼리 구문을 허용하기 때문에, SQL:2008 쿼리 뿐만이 아니라 원본 쿼리가 여기에 나열된다. 독자적(proprietary) 쿼리 구문을 SQL:2008 타입으로 변환하는 것은 SIARD 포맷을 지원하는 각 DB 프로그램에 대응하는 어플리케이션에서 정의되고, 문서화되어야 한다.	O
rows	레코드 수	O
description	뷰의 의미와 내용에 대한 기술	O

25) (NAK)원문: ~ can be archived in ~

26) (NAK)원문: ~ is archived:

B.5.15 루틴 레벨 메타데이터

ID	요구 조건의 기술	M/O
M_5.15-1	스키마의 루틴 메타데이터는 metadata.xml 파일에 보존될 수 있다. ²⁷⁾	O

다음의 루틴 메타데이터는 루틴이 보존된²⁸⁾ 경우, metadata.xml 파일에 저장된다.

식별자	뜻	M/O
specificName	스키마의 루틴을 고유하게 식별하는 특정 이름 ²⁹⁾	M
name	스키마의 루틴 이름	M
description	루틴의 의미와 내용에 대한 기술	O
source	루틴의 원본 소스 코드 (VBA, PL/SQL, JAVA) 노트 많은 DB 프로그램이 SQL:2008-호환 쿼리로 변환될 수 없는 독자적 루틴을 가지고 있기 때문에, 루틴의 원본 코드(예: Oracle DB의 PL/SQL, MS Access 모듈의 VBA)가 여기에 보존될 수 있다.	O
body	루틴의 SQL:2008-호환 소스 코드	O
characteristic	루틴의 특징	O
returnType	루틴의 리턴 타입 (함수인 경우)	
parameters	파라미터 목록	

B.5.16 파라미터 메타데이터

ID	요구 조건의 기술	M/O
M_5.16-1	루틴에서 사용되는 파라미터 메타데이터는 metadata.xml 파일에 보존될 수 있다. ³¹⁾	O

다음의 파라미터 메타데이터는 파라미터가 보존된³²⁾ 경우, metadata.xml 파일에 저장된다.

식별자	뜻	M/O
name	파라미터의 이름	M
mode	파라미터의 모드(IN, OUT 또는 INOUT)	M
type	파라미터의 SQL:2008 내장 데이터 타입 노트 이 열의 데이터 타입이 내장 데이터 타입인 경우, 이 필드는 사용되어야 한다.	O
typeOriginal	원본 파라미터 타입 노트 SQL을 준수하는(SQLcompliant) 다양한 DB 프로그램들이 매우 다른 데이터 타입을 허용하기 때문에, SQL:2008 타입 뿐만 아니라 원본 타입이 여기에 나열된다. 독자적(proprietary) 타입을 SQL:2008 타입으로 변환하는 것은 SIARD 포맷을 지원하는 각 DB 프로그램에	O

27) (NAK)원문: ~ can be archived in ~

28) (NAK)원문: ~ is archived:

29) 객체 지향 요소를 SQL:1999에 도입함으로써, 파라미터 목록이 다른 두 개의 서로 다른 루틴(프로시저 또는 함수)이 동일한 이름을 가질 수 있는, "overloading"이 가능해졌다. 따라서 스키마에서 루틴이 고유해야 한다는 요구 조건은 성립할 수 없다. 대신 스키마의 루틴을 고유하게 식별하기 위하여, "특정 이름"이 도입되었다.

	대응하는 어플리케이션에서 정의되고, 문서화되어야 한다.	
typeSchema	파라미터가 내장 데이터 타입이 아니고 명명된 데이터 타입이 이 열의 테이블로서 동일한 스키마에 정의되어 있지 않은 경우, 명명된 타입의 스키마	O
typeName	이 파라미터의 고급 또는 구조화된 데이터 타입 이름	O
description	루틴의 의미와 기능에 대한 기술	O
cardinality	파라미터가 배열인 경우, 요소의 (최대) 개수	O

B.5.17 유저 레벨 메타데이터

ID	요구 조건의 기술	M/O
M_5.17-1	유저 메타데이터는 metadata.xml 파일에 보관될 수 있다. ³⁰⁾	O

다음의 유저 메타데이터는 metadata.xml 파일에 저장된다.

식별자	뜻	M/O
name	유저의 이름	M
description	유저의 의미와 기능에 대한 기술	O

B.5.18 역할 레벨 메타데이터

ID	요구 조건의 기술	M/O
M_5.18-1	역할 메타데이터는 metadata.xml 파일에 보존될 수 있다. ³¹⁾	O

다음의 역할 메타데이터는 metadata.xml 파일에 저장된다.

식별자	뜻	M/O
name	역할의 이름	M
admin	역할의 관리자 (유저 또는 역할)	M
description	역할의 의미와 기능에 대한 기술	O

B.5.19 권한 레벨 메타데이터

ID	요구 조건의 기술	M/O
M_5.19-1	권한 메타데이터는 metadata.xml 파일에 보존될 수 있다. ³⁵⁾	O

다음의 권한 메타데이터는 metadata.xml 파일에 저장된다.

식별자	뜻	M/O
type	부여된 권한 (예: SELECT)	M
object	권한이 적용될 객체	O
grantor	권한의 허가된 부여자	M
grantee	권한을 받는 대상 (유저 또는 역할)	M
option	권한 부여 옵션 (ADMIN 또는 GRANT)	O
description	권한 부여의 의미와 기능에 대한 기술	O

30) (NAK)원문: ~ can be archived in ~

31) (NAK)원문: ~ can be archived in ~

B.6 테이블 데이터 요건

위에서 기술한 것처럼, 보관된³²⁾ RDB의 테이블 데이터는 SIARD 아카이브의 문서 루트 안에 있는 content/ 폴더에 위치한다. 이들 테이블 데이터는 해당 스키마 및 테이블 폴더에 보관된다.

테이블 데이터는 항상 XML 파일에 저장된다. XML 스키마 정의는 각 테이블에 대해서 테이블 데이터의 XML 저장 포맷을 나타내기 위해 생성된다. 즉, 각 테이블에 대해서, table[number].xml에 스키마 정의 table[number].xsd 파일이 있다.

ID	요구 조건의 기술	M/O
T_6.0-1	모든 테이블 데이터(기본 데이터)는 SQL:2008의 일관성 요구사항을 반드시 충족해야 한다. 다양한 XSD에 대하여 구문적으로 유효성을 검사하지만, 의미 상으로 SQL 표준을 어기는 SIARD 파일은 이 포맷 기술에 부합하지 않는다. 특히 테이블 값은 메타데이터의 SQL 타입 제약 조건에 반드시 부합해야 한다. 또한 기본, 후보 및 외래 키 조건과 메타데이터에 저장된 null 허용 조건(nullability conditions)이 반드시 모두 만족해야 한다.	M
T_6.0-2	table[number].xml 파일은 스키마 정의 table[number].xsd 를 반드시 준수해야 한다. 즉, table[number].xml 은 table[number].xsd 에 대한 유효성 검사를 반드시 통과해야 한다.	M

B.6.1 테이블 스키마 정의

table[number].xsd 파일은 테이블에 대한 다음의 스키마 정의를 포함하고 있다.

ID	요구 조건의 기술	M/O
T_6.1-1	테이블 데이터의 XML 저장 포맷을 나타내는, 각 테이블에 대한 XML 스키마 정의가 반드시 있어야 한다.	M
T_6.1-2	이 스키마 정의는 테이블의 SQL 스키마 메타데이터를 반영하고, 테이블이 다양한 XML 타입의 열(column) 엔트리 시퀀스를 담고있는 일련의 line들로 저장된다는 것을 나타낸다. 열(column) 태그는 <i>c1, c2, ...</i> 로 하는 반면에, 테이블 태그의 이름은 <i>table</i> , 데이터세트 태그의 이름은 <i>tag</i> 이다. 열 태그는 항상 <i>c1</i> 으로 시작하여 1 씩 증가한다. NULL 값은 XML 파일에서 누락된 대응되는(missing corresponding)(?) 열로 표현되기 때문에, 공백이 없어야 한다. 예시 Appendix D.3a.의 table2.xsd 예시	M
T_6.1-3	테이블 스키마 정의에 사용되는 타입 매핑은 P 4.3-3에 명시되어 있다. XML 스키마 표준 타입 외에 다음의 특수한 타입이 사용된다: clobType, blobType, dateType, timeType, dateTimeType.	M

32) (NAK)원문: archived relational database

T_6.1-4	<p>고급 또는 구조화된 타입의 다수의 셀 값은 셀 태그 내에 개별 요소로 저장된다.</p> <p>ARRAY의 개별 요소 이름은 a1, a2, ... 이다. UDT의 개별 요소 이름은 u1, u2, ... 이다. 이름은 항상 각각 a1 또는 u1으로 시작하고, 1씩 증가한다. NULL 값은 XML 파일에서 해당 열의 누락으로 표현되기 때문에, 반드시 공백이 없어야 한다.</p> <p>예시 Appendix D.3c.의 table0.xsd 예시</p>	M
---------	---	---

B.6.2 대형 객체 데이터 셀

ID	요구 조건의 기술	M/O
T_6.2-1	<p>대형 객체는 table[number].xml 에 내부적으로 SIARD 아카이브의 분리된 파일 엔트리로서, 또는 외부적으로 파일시스템에 독립적 파일로서, 인라인으로(inline) 저장된다.</p>	M

대형 객체가 (내부적 또는 외부적으로) 개별 파일로 저장되는 경우, *file* 및 *length* 속성은 table[number].xsd 파일의 LOB 셀에 반드시 저장되어야 한다. 대형 객체가 인라인으로 저장되는 경우, 이 속성은 선택 사항이다.

식별자	뜻	M/O
file	<p>대형 객체가 인라인되지 않은 경우, 이 요소는 이 셀 또는 셀의 속성에 있는 대형 객체의 위치와 이름을 "file:" URI로 나타낸다. 상대 URI인 경우, 동봉된(enclosing)(?) 요소(열 또는 속성)의 <i>lobFolder</i>에 연계되어 있다고 해석한다.</p>	M ³³⁾
length	<p>길이 (BLOB의 경우 바이트, CLOB과 XML의 경우 문자)</p>	M ³⁷⁾
digestType	<p>무결성 정보 타입을 포함한다 (digest): "MD5", "SHA-1" 또는 "SHA-256"</p> <p>권고 <i>digest</i> 속성 뿐만 아니라, 이 속성은 개별 파일로 저장되는 모든 대형 객체에 대해 작성되어야 한다.</p>	O
digest	<p>CLOB 또는 BLOB의 무결성 정보를 포함한다.</p> <p>권고 <i>digestType</i> 속성 뿐만 아니라, 이 속성은 개별 파일로 저장되는 모든 대형 객체에 대해 작성되어야 한다.</p>	O

33) 대형 객체가 table[number].xml 파일에 인라인으로(inline) 저장된 경우, 선택적 요구조건(O)이다.

B.6.3 Date와 timestamp 데이터 셀

ID	요구 조건의 기술	M/O
T_6.3-1	Date와 timestamp는 SQL:2008 명세에 따라 0001-9999 년으로 반드시 제한되어야 한다. 이 제한은 <i>dateType</i> 및 <i>dateTimeType</i> 정의에 적용된다.	M
T_6.3-2	Date, time 및 timestamp는 Z로 끝나는 UTC로 저장되어야 한다. 이 제한은 <i>dateType</i> , <i>timeType</i> 및 <i>dateTimeType</i> 정의에 적용된다. 권고 모든 <i>date</i> , <i>time</i> 및 <i>timestamp</i> 는 Z로 끝나는 형태로 저장한다.	M

B.6.4 테이블 데이터

table[number].xml 파일은 이 테이블에 대한 다음의 테이블 데이터를 담고 있다:

ID	요구 조건의 기술	M/O
T_6.4-1	각 테이블의 테이블 데이터는 XML 파일에 반드시 저장되어야 한다.	M
T_6.4-2	테이블 파일은 다양한 열(<i>c1</i> , <i>c2</i> ...)로 세분화된 행 데이터를 포함하는 행 요소로 구성된다. 예시 Appendix D.4a.의 table2.xml 예시	M
T_6.4-3	열 또는 필드의 셀이 NULL인 경우, 이는 반드시 생략되어야 한다. 만약 길이가 0인 문자열 ""과 같으면, 이는 반드시 존재해야 하고, 비어 있어야 한다.	M
T_6.4-4	열의 셀이 복합 값(ARRAY, UDT)을 포함하고 있으면, 차례로 각각의 값을 가지고 있는 셀의 하위요소 시퀀스로 표현된다. 이 값들은 다시 복합화 될 수 있다. 예시 Appendix D.4c.의 table0.xml 예시	O
T_6.4-5	테이블이 대형 객체 타입(BLOB, CLOB, 또는 XML ...)의 데이터를 포함하는 경우, 개별 파일이 이에 대해 생성될 수 있으며 파일의 저장위치가 셀 콘텐츠 대신 파일의 저장위치가 저장된다. 대형 객체 데이터를 인라인하는 대신 각각의 파일로 저장할 때, 결정은 SIARD 아카이브를 생성하는 소프트웨어의 재량에 달려있다. 빈 폴더가 생성되지 않도록, 폴더는 필요할 때만 생성된다. (예: 데이터 포함) 대형 객체가 별도의 파일로 저장되는 경우, 셀 요소는 반드시 <i>file</i> , <i>length</i> 및 <i>digest</i> 속성을 반드시 가지고 있어야 한다. 예시 Appendix D.4b.의 table7.xml 예시 권고 하나의 열에 모든 대형 오브젝트를 인라인 하거나 none으로 설정하는 것이 강력하게 권장된다. lob 폴더 및 lob 파일을 표준화하고, 실제 이름 대신 예를 들어 lob4/ 및 record0.bin 또는 record0.txt를 사용할 것이 권장된다.	M

B.7 약어와 용어 정의

용어	뜻
AIP	보존정보패키지(Archival Information Package). AIP는 디지털 문서를 보관하는 과정에서 SIP로부터 비롯된다. 디지털 문서가 디지털 저장소에 저장된 정보 패키지의 형태를 나타낸다.
Archive	1. 보존기록의 목록 작성, 보관 및 보존, 이용을 담당하는 기관 2. 조직의 보관된 문서 3. 문서 보관 목적으로 설립된 건물 또는 기관 4. 다른 파일이 들어있는 파일에 대한 용어. Archive file 및 동의어 container file 참조.
Archive records	보관을 위해 아카이브에서 입수했거나 동일한 원칙에 따라 다른 기관에서 독립적으로 보관한 문서를 가리킨다.
Database	데이터베이스는 일반적으로 데이터베이스의 특정부분에 대한 개별 사용자 및 역할의 정의된 액세스 권한은 물론, 하나 이상의 데이터베이스 스키마로 구성된다. SQL:2008 사용자와 역할은 권한을 가질 수 있다. 따라서 관계형 데이터베이스는 다수의 구조화된 데이터베이스 객체(예: 스키마, 뷰) 및 테이블 콘텐츠로 구성된다. 데이터베이스 스키마는 일종의 네임스페이스 접두사이다. 데이터베이스 카탈로그는 카탈로그 내의 모든 스키마에 대한 메타데이터가 들어있다. SQL:2008의 카탈로그 레벨은 SIARD를 사용하여 보존 포맷으로 변환될 수 있는 데이터베이스 "문서"에 해당한다.
DIP	배포정보패키지(Dissemination Information Package); OAIS에 따르면, DIP는 주문 절차를 통해 사용자가 요청한 서류에 대한 컨테이너이다.
DNS	도메인 네임 시스템(Domain Name System)은 인터넷에서 네임스페이스를 관리하는 분산 데이터베이스이다.
Documents	문서는 매체에 관계없이, 정보를 이해하고 사용하기 위해 필요한 검색 지원도구 및 보충자료 뿐만 아니라 공무 수행 중 입수하거나 생산한 모든 기록정보이다.
Dossier	특정 업무와 관련된 모든 문서들. 하나의 Dossier는 기본적으로 하나의 업무와 관련이 있다. 하지만 유사한 사업을 결합하거나 dossier를 subdossier로 나누어, 이 기본 구조는 해당 요구를 충족시키기 위해 조정될 수 있다. Dossier의 산출(Compilation)은 분류 체계에 기초하여 수행된다.
Information package	선택적인 콘텐츠 정보와 선택적으로 구성된 보존 메타데이터로 구성된 개념적 컨테이너. 콘텐츠 정보와 패키지 기술을 서로 구분하고, 식별하고, 콘텐츠 정보를 검색할 수 있도록 하는 패키징 정보가 여기에 포함된다.
LOB	별도의 파일로 표시될 수 있는 CLOB, BLOB 또는 XML 열의 셀 콘텐츠에 대해 일반적으로 사용되는 "대형 객체".
Long-term archiving	일반적으로 시간의 제약없이, 디지털 정보를 저장하고 그것의 장기 가용성을 유지하는 것. 보관된 정보의 비트 스트림을 유지하는 것 외에도, 이를 사람이 항상 읽고 이해할 수 있는 형식으로 해석하고 표시할 수 있는 기능도 포함한다.
MD5	Message-Digest Algorithm 5
Metadata	메타데이터는 기술적인 특성을 가지고 있기 때문에 "기본 데이터에 대한 정보"(데이터에 대한 정보)로 설명될 수 있다.

OAIS	Open Archival Information System, ISO 14721:2003. 참조 모형. 아카이브를 정보를 보존하고 지정된 집단에서 사용할 수 있도록 사람과 시스템이 함께 작동하는 조직으로 설명한 참조모델.
Primary data	기본 데이터는 문서의 내용을 구성하는 데이터이다. SIARD 파일 내에서, 테이블 데이터가 기본 데이터의 기능을 수행한다.
Records creator	문서를 만들고 관리하는 기관 또는 조직단위를 가리킨다.
Routines	SQL 루틴(보관된 프로시저라고도 한다)은 부분적 표현이 가능한 뷰 쿼리를 이해하는 데 중요하다.
Schemas	스키마는 테이블, 뷰 및 루틴의 컨테이너이다.
SFA	스위스 연방 기록원(Swiss Federal Archives)
SHA1	Secure Hash Algorithm 1
SIP	제출정보패키지(Submission Information Package): OAIS에 따르면, SIP는 기록 생산 기관이 아카이브에 제출하는 정보 패키지이다. 디지털 문서(기본 데이터 및 메타데이터)가 포함된다.
Tables	테이블은 테이블의 각 열에 이름과 유형을 할당하는 필드, 실제 테이블 데이터를 포함하는 데이터세트, 선택적인 기본키, 참조 무결성을 보장하는 외래키, 데이터세트를 식별하는데 쓰이는 후보키, 그리고 일관성을 보장하는 제약사항에 대한 테이블 정의로 구성된다.
UTF	유니코드 변환 포맷(Unicode Transformation Format)
Views	뷰는 데이터베이스에 저장된 표준 쿼리이다. 또한 쿼리 결과는 필드와 데이터세트를 포함하는 테이블이다.
XSD	XML 스키마 정의

참고문헌

- [1] 국가기록원. 2007년. 「행정정보시스템 데이터세트 기록관리방안 연구보고서」.
- [2] 국가기록원. 2015년. 「데이터세트 구조분석 및 진본성 보장 기록관리 기능모델 연구」.
- [3] 국가기록원. 2016년. 「행정정보 데이터세트 기록관리 방안」.
- [4] 국가기록원. 2017년. 「차세대 기록관리모델 재설계 연구-행정정보 데이터세트 기록관리체계 설계-」.
- [5] 국가기록원. 2017년. 「행정정보 데이터세트 기록관리체계 실행방안(안)」.
- [6] 국가기록원. 2018년. 「행정정보 데이터세트 기록관리 매뉴얼(안)」.
- [7] 국가기록원. 2019년. 「행정정보 데이터세트 기록관리 체계 구축 시범사업 산출물」.
- [8] 국가기록원. 2020년. 「행정정보 데이터세트 관리방안 : 기록물관리지침」.