Certificate of Data Quality

DQ인증

(데이터 품질인증)

과학기술정보통신부

WISESTONE

데이터품질인증

Certificate of Data Quality

과학기술정보통신부가 지정한 데이터 품질인증기관이
「데이터 산업진흥 및 이용촉진에 관한 기본법」 제20조 및
동법 시행령 제20조 5항(데이터 품질인증 대상 및 품질기준)에 의거
①데이터 내용 ②데이터 관리체계를 진단하고 수준을 평가해 **품질을 인증**합니다.

인증 대상

다양한 산업에서 생산·유통되는 모든 데이터를 대상으로 합니다.

제조 · 설비 · 의료 · 통신 · 공공 · 금융 · 이커머스 · 유통 · 자동차 · IoT

인증제도 효과

- 국가공인기관의 인증을 통해 데이터 품질 및 신뢰성 확보
- 데이터 제공자, 수요자 간 요구사항 조율 및 매칭
- 기구축 데이터 및 신규 계획 데이터의 품질 제고
- 마케팅 및 홍보자료 활용 등

와이즈스톤만의 강점

- 국내 최초 데이터 품질 분야 KOLAS 국제공인시험기관
- 국내 최다 데이터 품질 시험·인증 수행 경험
- 데이터 품질 분야 전담 조직 운영
- 다양한 데이터 품질 검증 도구 보유
- 국내 제1호 데이터 품질인증서 발급

인증 절차

인증 신청 → 상담 및 계약 → 인증 심사 → 인증심의위원회 → 인증서 발급

데이터 내용 인증 체계

■ 데이터 내용 인증 대상

정형 데이터(데이터 레코드, 필드에 입력되어 있는 문자, 숫자 등 값과 집합) 또는 비정형 데이터(텍스트, 이미지, 동영상, 음성 등의 원천 데이터와 속성값 등)

■ 데이터 내용 인증 유형 및 등급

• 인증 유형

데이터 구조•관리 특성에 따라 적용 가능한 필수 심사 항목(기본 중요 지표)과 선택 심사 항목(기준 정보에 따른 점검 지표)으로 구분되며,
데이터의 복잡도를 고려하여 적용되는 심사 항목에 따라 분류

인증 유형	기준
Complex-Type	필수 심사 항목 전체 적용 및 선택 심사 항목 3개 이상 적용
Normal-Type	필수 심사 항목 전체 적용 및 선택 심사 항목 3개 미만 적용
Simple-Type	필수 심사 항목 일부 적용 (불가피하게 데이터 특성에 따라 일부 항목 진행 불가 시)

• 인증 등급

데이터의 품질 수준에 따라 3개 등급으로 구분

인증 등급	판정기준
Class A	데이터 품질 점수 0.99 이상
Class B	데이터 품질 점수 0.97 이상
Class C	데이터 품질 점수 0.95 이상

※ 모든 개별 심사 항목이 최소 0.95 이상

■ 정형데이터 심사

• 정형데이터 심사 대상

정의된 구조(스키마)에 따라 체계적으로 구성된 데이터로, RDB(MySQL, Oracle, PostgreSQL 등), 스프레드시트(CSV 파일, 엑셀 파일 등), 구조화된 텍스트 데이터 등

• 정형데이터 심사 항목

과학기술정보통신부의 데이터 품질인증 품질기준을 기반으로 ISO/IEC 25024에서 정의된 품질 측정 지표를 참고하여 정형데이터 자체(inherent) 품질을 판정하기 위한 심사 항목으로 구성

구분	번호	데이터 품질기준	심사항목	설명
필수	1	완전성	데이터값 완전성	필수(NOT NULL)로 선언된 컬럼에 모든 데이터 값이 존재하는지 심사하는 항목
	2		데이터 레코드 완전성	레코드가 누락 없이 필요한 데이터를 모두 포함하고 있는지를 심사하는 항목
	3	유효성	구문 유효성	데이터 값이 컬럼에 정해진 구문 규칙을 따르는지 심사하는 항목
	4		의미 유효성	데이터 값이 의미적으로 유효하고 설정된 기준에 부합하는지 심사하는 항목
	5		범위 유효성	데이터 값이 유효한 범위 내에 있는지 심사하는 항목
	6		관계 유효성	데이터 값이 다른 컬럼과의 관계 규칙을 준수하는지 심사하는 항목
	7	일관성	참조 무결 일관성	데이터 값이 테이블 간의 참조 무결성 규칙을 지키는지를 심사하는 항목
선택	1	유효성	데이터값 정밀성	데이터 값이 정밀도 요구사항을 충족하는지 심사하는 항목
	2	일관성	데이터 포맷 일관성	데이터 파일 내에서 동일 속성의 구문이 일관되게 유지되는지 심사하는 항목
	3		공통 어휘 일관성	데이터 사전에 정의된 공통 용어를 사용하는지 심사하는 항목
	4	정확성	메타 데이터 정확성	메타 데이터의 데이터 값이 정확하고 설정된 요구사항에 부합하는지 심사하는 항목
	5		데이터값 정확성	컬럼 간 규칙에 따라 데이터의 값이 정확한지 평가
	6	접근성	표준기반 데이터 접근성	데이터가 요구되는 표준이나 협약을 얼마나 준수하는지를 심사하는 항목
	7	유일성	데이터값 유일성	데이터의 특정 값이 중복 없이 유일한지를 심사하는 항목

■ 비정형데이터 심사

• 비정형데이터 심사 대상

비정형데이터 : 일반적으로 객체(이미지, 비디오, 오디오, 텍스트) 데이터와 메타 데이터(객체 설명 또는 태깅-라벨링 등 작업 내용 설명)로 구성
반정형데이터 : 일부 구조화된 형태를 가지고 있는 데이터로 텍스트 파일, XML, JSON 등의 형식으로 저장되며, 필드와 값의 조합으로 구성

• 비정형데이터 심사 항목

과학기술정보통신부의 데이터 품질인증 품질기준을 기반으로 ISO/IEC 5259-2, NIA 학습용 데이터 품질관리 가이드라인, TTAK.KO-10.1344-Part2에서 정의된 품질 측정 지표를 참고하여 비정형데이터 품질을 판정하기 위한 심사 항목으로 구성

구분	번호	데이터 품질기준	심사 항목	설명
필수	1	완전성	데이터파일 레코드 완전성	데이터 명세서에 기술된 파일 수와 실제 파일 수가 일치하는지 심사하는 항목
	2		레코드 완전성	전체 데이터 중 필수 항목이 NULL이 아닌 데이터 레코드의 비율을 심사하는 항목
	3		메타 데이터 값 완전성	비정형 객체의 메타 데이터에서 필수 항목이 NULL이 아닌지를 심사하는 항목
	4	유효성	데이터 구조 구문 유효성	비정형 객체의 메타 데이터와 어노테이션 구조가 명세서에 정의된 규칙에 적합한지를 심사하는 항목
	5		데이터 포맷 유효성	비정형 객체가 명세서에 정의된 파일 포맷을 따르는지를 심사하는 항목
	6		객체 유효성	비정형 객체 중 실제로 활용 가능한 파일인지 여부를 심사하는 항목
	7	정확성	구문 정확성	비정형 객체에 대한 메타 데이터와 어노테이션 내 데이터의 구문이 명세서에 정의된 구문에 적합한지를 심사하는 항목
	8		메타 데이터 정확성	비정형 객체의 메타 데이터 값이 정확한지를 심사하는 항목
	9	유일성	객체 유일성	비정형 객체가 중복 없이 유일한지를 심사하는 항목
선택	10	완전성	어노테이션 완전성	학습용 데이터에서 객체 데이터에 어노테이션(라벨)이 존재하는 비율을 심사하는 항목 (학습용 데이터의 경우 필수 항목으로 적용)
	11		특징 완전성	데이터 셋에서 요구된 특징(feature)에 NULL 값이 없는 비율을 심사하는 항목
	12	유효성	관계 유효성	메타 데이터와 어노테이션 간의 관계 규칙을 준수하는지를 심사하는 항목
	13		데이터 값 정밀성	데이터 값이 명세서에 요구된 정밀도를 충족하는지를 심사하는 항목
	14		데이터 속성 유효성	비정형 객체의 속성이 데이터 명세서 기준에 적합한지를 심사하는 항목
	15		범위 유효성	메타 데이터와 어노테이션 값이 명세서에 정의된 범위 내에 포함되는지를 심사하는 항목
	16		시간 유효성	데이터 생성 또는 수집 시간이 명세서의 기준 시간을 만족하는지를 심사하는 항목
	17		어노테이션 유효성	학습 데이터에서 어노테이션 값(예: 바운딩 박스, 라벨링)이 기준을 만족하는지를 심사하는 항목
	18	일관성	공통 어휘 일관성	객체 데이터 파일명, 메타 데이터, 어노테이션이 공통 어휘 사전에 정의된 용어를 사용하는지를 심사하는 항목
	19		데이터 포맷 일관성	비정형 객체와 메타 데이터, 어노테이션 데이터가 동일한 도메인의 데이터 포맷을 유지하는지를 심사하는 항목
	20	정확성	어노테이션 정확성	학습 데이터의 어노테이션 값이 정확한지를 심사하는 항목
	21		주제 정확성	비정형 객체와 주어진 주제(컨텍스트)의 적절성을 심사하는 항목
	22	접근성	데이터 포맷 접근성	비정형 객체 데이터가 일반적인 데이터 포맷(예: jpg, png, bmp 등)으로 제공되는지를 심사하는 항목
	23		표준기반 데이터 접근성	데이터가 표준, 협약 또는 규정을 준수하는지를 심사하는 항목
	24	유일성	레코드 유일성	데이터 레코드가 중복되지 않고 유일한지를 심사하는 항목

Certificate of Data Quality

데이터 관리 체계

■ 데이터 관리 체계 인증 대상

- **데이터 기반 사업(프로젝트)** : 데이터를 생성, 가공분석, 활용하기 위해 수행되는 장기적인 데이터 기반 사업(프로젝트)지속성을 가진 데이터 기반 사업 단위의 데이터 관리체계
- **시스템** : 데이터를 처리하고 활용하기 위해 구축된 데이터 시스템 또는 플랫폼
- **관리 조직** : 데이터 기반 사업 또는 데이터 시스템을 운영하는 조직(팀, 부서, 전사)

■ 인증 등급 및 성숙도 수준

인증 등급	성숙도 수준	기준	인증여부
Lv. 1	수준 1 (기초적 관리)	운영 프로세스가 요구사항을 충족하는 데이터에 접근할 수 있음을 입증(데이터에는 적절한 보안 고려사항 적용)	**인증하지 않음**
		요구사항 관리 및 데이터 처리 활동에 대한 증빙 자료 미존재	
Lv. 2	수준 2 (실무적 관리)	요구사항을 관리하고 데이터 처리 수행 방법을 관리하게끔 운영 프로세스가 데이터를 사용한다는 것을 입증	**인증함**
		데이터가 요구사항을 충족하는 증빙 자료 존재	
		요구사항 관리, 데이터 처리 작업 관리, 품질 통제 수행	
Lv. 3	수준 3 (체계적 관리)	운영 프로세스가 조직이 데이터 품질관리를 수행하기 위해 공통적이고 반복 가능한 프로세스를 구현한 데이터를 사용한다는 것을 입증	
		데이터 관리 프로세스의 내용과 수행 방법을 체계적으로 정의	
		기획(QP), 통제(QC), 보증(QA), 개선(QI) 활동을 연계하여 수행	
Lv. 4	수준 4 (정량적 관리)	운영 프로세스가 데이터 품질관리를 수행하기 위해 예측 가능한 프로세스를 구현한 데이터를 사용한다는 것을 입증(예측 가능성에는 데이터 품질관리 성과 측정 수반)	
Lv. 5	수준 5 (관리 최적화)	운영 프로세스가 조직의 목표를 달성하는 데 지속 가능한 데이터 품질관리를 수행하기 위한 프로세스를 구현한 데이터를 사용한다는 것을 입증(지속 가능성에는 적절한 혁신 적용)	

■ 성숙도 수준에 따른 프로세스별 심사 항목

ID	프로세스 명	관리체계 성숙도 수준				
		1	2	3	4	5
DQP.1	요구사항 관리	–	●	●	●	●
DQP.2	데이터 품질 전략 관리	–	–	△	●	●
DQP.3	데이터 품질기준 문서 관리	–	–	●	●	●
DQP.4	데이터 품질 계획 수립	–	–	△	●	●
DQC.1	작업 지시 및 처리 지침 제공	–	●	●	●	●
DQC.2	데이터 처리 작업 관리	●	●	●	●	●
DQC.3	데이터 품질 모니터링 및 통제	–	●	●	●	●
DQA.1	데이터 품질 이슈 검토	–	–	–	●	●
DQA.2	측정 기준 제공	–	–	▲	●	●
DQA.3	데이터 품질 및 프로세스 성과 측정	–	–	▲	●	●
DQA.4	측정 결과 평가	–	–	–	●	●
DQI.1	원인 분석 및 해결 방안 도출	–	–	–	–	●
DQI.2	데이터 및 프로세스 수정	–	–	●	●	●
DQI.3	프로세스 개선	–	–	–	–	●
DRS.1	데이터 아키텍처 관리	–	–	●	●	●
DRS.2	데이터 흐름 관리	–	–	–	●	●
DRS.3	데이터 운용 관리	–	–	●	●	●
DRS.4	데이터 보안 관리	●	●	●	●	●
RPV.1	데이터 품질 조직 관리	–	–	●	●	●
RPV.2	인적 자원 관리	–	–	–	●	●

△ : Level 3 심사에서는 필수가 아님　▲: Level 3 심사에서는 데이터 품질 측정 관련 사항만 심사(기준 미준수 및 프로세스 약점에 대한 파악은 제외)　● : 해당 프로세스가 특정 성숙도 Level에서 심사 범위에 포함됨을 의미
예) 'DQC.2 데이터 처리 작업 관리'는 Level 1~5 심사에서 언제나 심사 범위에 포함'RPV.2 인적 자원 관리'는 Level 4, 5 심사에서만 심사 범위에 포함되며, Level 1~3 심사에서는 심사대상이 아님

DQ인증(데이터 품질인증)
주요 인증 사례

인증 대상 : 한화생명 통합고객 시스템

인증 대상 : GIVC 시스템

인증 대상 : 뉴스 기사 데이터

인증 대상 : 바이오 소재 데이터 플랫폼

인증 대상 : 무기체계 센서데이터

인증 대상 : IT 산업 뉴스 기사 데이터

인증 대상 : SW, HW 검색 및
비교 플랫폼 데이터

인증 대상 : 글로벌 실도로 주행 영상 데이터

인증 대상 : 사업자정보 수집 데이터

인증 대상 : SpAC9 Vault
(암 종류에 따른 당/단백질 데이터 리포지토리)

인증 대상 : 제약바이오 약물 프로파일 및
관련 비즈니스 계약 데이터

인증 대상 : 인구 인벤토리(치수 피해규모 정량산출)
제주도 홍수 및 침수 위험지수 산정 모델용 입력 데이터

인증 대상 : 가정용 AMI 스마트전력 데이터

인증 대상 : ConnectDI

인증 대상 : 스마트 기기 제품 데이터

인증 대상 : 전자의무기록
(EMR : Electronic Medical Record) 데이터

인증 대상 : 정신건강 상담 데이터셋

인증 대상 : 기계 통역 데이터

인증 대상 : 차량 종합 상세 데이터

인증 대상 : 인플루언서 채널 데이터

인증 대상 : 전력설비 방호 안티드론
통합운용시스템 데이터

인증 대상 : 원격 제어 이력 데이터

인증 대상 : 제작의뢰 신청 데이터

인증 대상 : 유통 품목 데이터

+ and more

문의 및 상담

와이즈스톤
경기 과천시 과천대로7길 33 디테크타워 B동 1407호
Tel : 02.6257.5958 E-mail : data@wisestone.kr

dq.wisestone.kr