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 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

데이터 레코드, 필드에 입력되어 있는 문자, 숫자 등 값과 집합 또는 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 지표(기본 중요 지표)와 선택 지표(기준 정보에 따른 점검 지표)로 구분 데이터의 복잡도를 고려하여 적용되는 지표에 따라 인증 유형 기준 적용

구분	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)

※ 모든 개별 심사 항목이 0.95 이상

• 심사지표

데이터의 품질 수준에 따라 3개 등급으로 판정 기준을 구분

구분	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

■ 정형데이터 심사

• 정형데이터 심사 대상

정해 놓은 형식과 구조에 따라 저장되도록 구성된 데이터로 테이블과 컬럼으로 구성되어 있음

• 정형데이터 심사 지표

과학기술정보통신부 데이터품질인증 운영 지침을 기준으로 ISO/IEC 25024 평가 항목을 적용하여 정형데이터 자체(inherent) 품질을 판정하기 위한 심사 지표 구성

구분	번호	데이터 품질기준	심사항목	설명
필수	1	완전성	데이터값 완전성	데이터 파일의 데이터값에 대한 값의 완전성
	2		데이터 파일 레코드 완전성	데이터 파일의 데이터값에 대한 값의 완전성
	3	유효성	구문 유효성	구문적으로 정확한 값을 갖는 데이터값의 비율을 측정
	4		의미 유효성	의미적으로 정확한 데이터값의 비율을 측정
	5		범위 유효성	지정된 범위 안에 포함된 데이터값의 비율을 측정
	6		관계 유효성	컬럼 간의 관계규칙이 정의된 데이터 중 관계규칙을 만족하는 데이터의 비율
	7	일관성	참조 무결 일관성	테이블의 한 속성의 다른 테이블에서 같은 속성의 같은 값으로 존재하는 비율을 측정
선택	1	유효성	데이터값 정밀성	명세서의 정밀도를 만족하는 데이터값의 비율을 측정
	2	일관성	데이터 포맷 일관성	데이터 포맷 일관성이 요구되는 데이터 중 포맷 일관성을 만족하는 데이터 비율
	3		공통 어휘 일관성	데이터 사전에 정의된 공통 어휘를 사용하는 데이터 비율
	4	정확성	메타 데이터 정확성	요구사항 명세서에 정의된 메타 데이터 중 적절한 정보는 제공하는 메타 데이터 비율
	5		데이터값 정확성	데이터 파일 내 데이터값에서 기대 값을 만족하는 데이터값의 비율을 측정
	6	접근성	표준기반 데이터 접근성	표준, 협약 또는 규정이 요구되는 데이터 중 해당 표준, 협약 또는 규정에 적합한 데이터 비율
	7	유일성	데이터값 유일성	데이터값에 중복이 존재하는 데이터값의 비율을 측정

■ 비정형데이터 심사

• 비정형데이터 심사 대상

비정형 데이터 : 비정형데이터는 일반적으로 객체(이미지, 비디오, 사운드, 텍스트) 데이터와 메타데이터(객체 설명 또는 태깅·라벨링 등 작업 내용 설명)로 구성됨
반정형 데이터 : 일부 구조화된 형태를 가지고 있는 데이터로 텍스트 파일, XML, JSON 등의 형식으로 저장되며, 필드와 값의 조합으로 구성됨

• 비정형데이터 심사 지표

과학기술정보통신부 데이터품질인증 운영 지침을 기준으로 ISO/IEC FDIS 5259-2, NIA 학습용 데이터 품질관리 가이드라인, TTAK.KO-10.1344-Part2 평가 항목을 적용하여 비정형데이터 품질을 판정하기 위한 심사 지표 구성

구분	번호	데이터 품질기준	심사 항목명	설명
필수	1	완전성	데이터파일 레코드 완전성	데이터 파일의 데이터값에 대한 값의 완전성
	2		레코드 완전성	데이터 세트에 빈 데이터 아이템이 없는 데이터 레코드의 비율
	3		메타 데이터값 완전성	비정형 객체에 대한 메타 데이터 내 필수 항목이 Not Null인 데이터의 비율
	4	유효성	데이터 구조 구문 유효성	비정형 객체에 대한 메타데이터나 라벨링 데이터의 구조가 정의된 구조에 적합한 데이터의 비율
	5		데이터 포맷 유효성	비정형 객체가 정의된 파일의 포맷에 적합한 데이터 비율
	6		파일 유효성	비정형 객체 중 실제 활용할 수 있는 파일의 비율 ㅇ실제 활용이 어려운 예 - 공통 : 파일을 열어 정상적으로 파일이 동작하는지 확인 - 이미지 : 이미지가 깨진 경우 - 텍스트 : 텍스트가 깨져서 내용 확인이 어려운 경우 - 비디오 : 영상이 깨져 내용 확인이 어려운 경우 - 오디오 : 요구 사항에 없는 노이즈 또는 사운드 데이터가 비어 내용 확인이 어려운 경우
	7	정확성	구문 정확성	비정형 객체에 대한 메타데이터나 라벨링 데이터 내 속성(컬럼)값이 명세서 기준 구문적으로 정확한 값을 갖는 데이터 아이템의 비율
	8		메타 데이터 정확성	비정형 객체의 메타 데이터의 값이 정확한 비율(예: 파일명, 해상도, 확장자 등)
	9	유일성	객체 유일성	동일한 객체 데이터가 존재하는 비율(파일명이 아닌 데이터의 내용의 동일성 측정)
선택	10	완전성	어노테이션 완전성	학습용 데이터의 경우, 객체 데이터에 라벨이 존재하는 비율(학습용 데이터의 경우 필수 항목으로 적용) ㅇ 유형별 어노테이션 예 - 텍스트 : 내용 요약, 변역, 질의응답, 말뭉치 태깅 등 - 이미지/비디오 : 라벨링, 바운딩박스, 키포인트, 세그멘테이션, 태깅 등 - 오디오 : 전사 등
	11		특징 완전성	데이터 세트의 지정(요구)된 특징에 대하여 null 데이터값이 없는 데이터 아이템의 비율
	12	유효성	관계 유효성	비정형 객체의 속성과 메타데이터/어노테이션 데이터의 속성들 간의 규칙이 존재하는 경우, 규칙에 맞는 데이터 아이템의 비율
	13		데이터값 정밀성	데이터 명세서에 요구되는 정밀도 만족하는 데이터값의 비율
	14		데이터 속성 유효성	비정형 객체 데이터의 속성이 기준에 적합한 비율 ㅇ 유형별 속성 예 - 이미지 : 밝기, 해상도, 컬러심도, 픽셀당 비트수 등 - 동영상 : 해상도, 컬러심도, 프레임레이트, 재생시간 등 - 오디오 : 주파수, 재생시간, 평균 데시벨 등 - 텍스트 : 어절수, 어휘 개수, 문장의 길이 등급
	15		범위 유효성	메타데이터 또는 어노테이션 데이터의 속성 값이 요구되는 범위 안에 포함되는 비율
	16		시간 유효성	비정형 객체 데이터의 생성 또는 수집 시간이 기준 시간을 만족하는 비율
	17		어노테이션 유효성	학습용 데이터의 경우, 어노테이션 데이터의 어노테이션 값이 기준을 만족하는 비율 (예: 바운딩박스 크기, 정해진 카테고리 또는 라벨링 기준 등)
	18	일관성	공통 어휘 일관성	데이터 사전에 정의된 용어(terms)로 사용되는 공통 인식 어휘 비율
	19		데이터 포맷 일관성	동일한 데이터 아이템의 데이터 포맷을 준수하는 데이터 아이템 비율
	20	정확성	어노테이션 정확성	학습용 데이터의 경우, 어노테이션 데이터가 정확한 비율 ㅇ 유형별 어노테이션 예 - 텍스트 : 내용 요약, 변역, 질의응답, 말뭉치 태깅 등 - 이미지/비디오 : 라벨링, 바운딩박스, 키포인트, 세그멘테이션, 태깅 등 - 오디오 : 전사 등
	21		주제 정확성	비정형 객체가 주어진 주제(컨텍스트)를 반영하는 비율
	22	접근성	데이터 포맷 접근성	일반적인 데이터 포맷으로 의도된 사용자가 데이터나 정보에 접근할 수 비율(예 : 이미지의 경우 jpg, png, bmp 등)
	23		표준기반 데이터 접근성	표준, 협약 또는 규정에 부합해야 하는 전체 데이터 아이템의 비율
	24	유일성	레코드 유일성	데이터 세트의 중복 레코드 비율

데이터 관리 체계

■ 데이터 관리 체계 인증 대상

- **프로젝트(또는 고유목적 사업)** : 지속성을 가진 데이터 기반 사업 단위의 데이터 관리체계
- **시스템(또는 플랫폼)** : 특정한 데이터 시스템 또는 플랫폼에 적용되는 데이터 관리체계
- **조직(또는 조직 단위)** : 데이터 기반 사업을 추진하는 조직(팀, 부서, 전사 포함)의 관리체계

■ 인증 수준

인증 수준	성숙도 수준		특징	인증여부
Lv. 1	1	도입	데이터 품질 인식은 존재하지만 체계적인 관리 활동 미흡	인증하지 않음
Lv. 2	2	관리	기본적인 관리 활동 존재(성과 관리 및 산출물 관리)	인증함
Lv. 3	3	체계화	공식화된 관리 활동이 체계적으로 수행	
Lv. 4	4	예측화	계량적이고 통계적인 방법으로 성과를 안정적으로 유지	
Lv. 5	5	혁신화	데이터 품질 관리체계를 지속적으로 고도화	

* 데이터 품질 관리체계 인증은 Lv.2 (관리 수준)부터 부여

■ 인증 수준별 심사 범위

ISO 8000-61, 62 및 국내 실무를 고려하여 5개 인증 수준(Lv.)별 심사 대상 프로세스는 다음과 같음.

인증 수준	심사 대상 프로세스	
Lv.2	DQP.1 요구사항 관리 DQC.1 데이터 명세 및 작업지시서 제공 DQC.2 데이터 처리 작업 관리	DQC.3 데이터 품질 모니터링 및 통제 DRS.4 데이터 보안 관리
Lv.3	DQP.1 요구사항 관리 DQP.2 데이터 품질 전략 관리 (필수 아님) DQP.3 데이터 품질 기준 문서 관리 DQP.4 데이터 품질 계획 수립 (필수 아님) DQC.1 작업 지시 및 처리 지침 제공 DQC.2 데이터 처리 작업 관리 DQC.3 데이터 품질 모니터링 및 통제 DQA.2 측정 기준 제공(데이터 품질 측정 기준 제공만 해당)	DQA.3 데이터 품질 및 프로세스 성과 측정(데이터 품질 수준 측정만 해당) DQI.2 데이터 오류 및 프로세스 수정 DRS.1 데이터 아키텍처 관리 DRS.3 데이터 운용 관리 DRS.4 데이터 보안 관리 RPV.1 데이터 품질 조직 관리
Lv.4	DQP.1 요구사항 관리 DQP.2 데이터 품질 전략 관리 DQP.3 데이터 품질 기준 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.2 데이터 오류 및 프로세스 수정 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
Lv.5	DQP.1 요구사항 관리 DQP.2 데이터 품질 전략 관리 DQP.3 데이터 품질 기준 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 오류 및 프로세스 수정 DQI.3 데이터 관리 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

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 : 02.2039.3155 E-mail : data@tecel.kr

dq.tecel.kr