

연구데이터 품질관리 및 품질평가 가이드라인

2021. 10.

<제목 차례>

1. 정확성
 - 1.1 입력값
 - 1.2 업무규칙
 - 1.3 범위 · 형식
 - 1.4 참조 관계
 - 1.5 계산식
2. 확실성
3. 완전성
 - 3.1 논리모델
 - 3.2 식별자
 - 3.3 물리구조
 - 3.4 속성의미
 - 3.5 고유성
4. 일관성
 - 4.1 속성
 - 4.2 표준
 - 4.3 연계값
 - 4.4 중복값
 - 4.5 이식성
5. 유용성
 - 5.1 충분성
 - 5.2 접근성
 - 5.3 활용성
6. 적시성
 - 6.1 최신값
 - 6.2 응답시간
 - 6.3 데이터 제공
7. 기밀성
 - 7.1 접근제한
 - 7.2 DB 보호
 - 7.3 오너쉽

- 8. 무결성
 - 8.1 개체 무결성
 - 8.2 참조 무결성
 - 8.3 도메인 무결성
 - 8.4 무결성 규칙
- 9. 적합성
- 10. 출처

<표 차례>

- <표 1> 데이터 품질관리 및 품질평가 가이드라인 지표 정의
- <표 2> 데이터 품질관리 및 품질평가 가이드라인 지표 특성 정의
- <표 3> 데이터 품질관리 및 평가 지표 측정 기준 및 배점표

지표는 총 9개, 지표 특성은 총 28개이다.

각 지표와 지표 특성에 대한 정의는 다음 표와 같다.

<표 1> 데이터 품질관리 및 품질평가 가이드라인 지표 정의

지표	정의
정확성	데이터의 정확성 측정. 정확한 데이터는 데이터의 원래 소스에서 의도했거나 정의한 내용을 나타내야 하며, 측정해야 하는 것을 측정함. 유효성이라고도 함. 정확한 데이터는 올바른 것으로 간주되며, 측정하려는 것을 측정함. 정확한 데이터는 오류(예시 : 녹음 또는 인터뷰어 편향, 전산 오류, 샘플링 오류)를 무시할 수 있을 정도로 최소화함
확실성	프로그램의 정보 시스템에서 생성되는 데이터는 누가, 언제, 얼마나 자주 사용하는지에 따라 변경되지 않는 프로토콜 및 절차를 기반으로 함. 데이터는 동일하게 정의, 측정 및 수집되기 때문에 신뢰할 수 있음
완전성	데이터의 예상 속성이 제공되는 정도. 모든 필수 데이터 요소는 데이터베이스 시스템에서 캡처됨. 완전성이란, 결과가 도출되는 정보 시스템이 적절하게 포괄적임을 의미할 수있으며, 이는 목록의 일부가 아니라 적격한 단위의 전체 목록을 나타냄
일관성	주어진 데이터 요소의 값이 애플리케이션과 시스템에서 동일할 때, 데이터는 일관성을 가짐
유용성	데이터가 당면한 작업에 적용가능하고 유용한 정도
적시성	데이터가 현재 상태이고, 지정된 대로 그리고 예상되는 시간 프레임에서 사용할 수 있는 정도 데이터는 최신(현재)이고 정보가 제때에 제공될 때 시기적절함. 적시성은 다음에 의해 영향을 받음 (1) 프로그램의 정보 시스템이 업데이트되는 속도 (2) 실제 프로그램 활동의 변화율 (3) 정보가 실제로 사용되거나 필요한 경우
기밀성	고객이 데이터에 대한 국가 및 국제 표준에 따라 데이터를 유지·관리한다는 것을 의미함. 이는 개인 데이터가 부적절하게 공개되지 않고 하드 카피 및 전자 형식의 데이터가 적절한 수준의 보안으로 처리됨을 의미함(예시 : 잠금 캐비닛 및 암호로 보호된 파일에 보관)
무결성	정보가 손상 또는 위조로 인해 손상되지 않도록 무단 액세스 또는 수정으로부터 정보를 보호하는 것을 의미하며, 데이터가 완전하고 정확하며 일관성 있고 맥락에 맞아야 함. 정밀성, 정확성, 완전성, 유효성의 의미로 사용되며, 데이터베이스의 정확성을 보장함. 데이터 무결성 기본 지표의 품질은 데이터의 존재, 구조, 내용 및 기타 기본적 특징의 척도임. (예: 데이터 무결성(data integrity)은 데이터를 보호하고, 항상 정상인 데이터를 유지하는 것)
적합성	형식이 이용에 적합한 정도

<표 2> 데이터 품질관리 및 품질평가 가이드라인 지표 특성 정의

지표	지표 특성	정의
정확성	입력값	데이터 입력 시 오류를 사전 검증(Edit)하고 수정(Imputation)하는 방법과 절차를 정의하고 준수해야 한다.
		법·규정에 의해 업무를 수행함에 있어 관리하는 데이터의 생성 규칙(산출식)으로 여러 테이블 간의 관계에 의해 의미적인 데이터의 정확성을 진단하는 로직임
	업무규칙	DB에 저장된 데이터 값은 유효한 범위와 형식을 가져야 한다.
	범위 · 형식	참조관계 데이터 간의 값은 참조되는 값의 범위 내에 존재해야 함
	참조 관계	계산 또는 집계된 값의 원값의 일부 혹은 전부는 DB에 이미 저장된 값이어야 한다.
완전성	계산식	계산 또는 집계된 값의 원값의 일부 혹은 전부는 DB에 이미 저장된 값이어야 한다.
	논리 모델	데이터 모델의 완전성은 데이터 품질관리의 가장 기본 사항이며, 대상되는 업무를 표현하는 엔티티와 엔티티간의 관계가 논리적으로 명확한지를 평가함
	식별자	모든 엔티티는 레코드들은 유일하게 구분하는 식별자가 존재해야 함
	물리 구조	논리 데이터 모델에서 정의한 데이터 규칙을 반영하여 물리구조(데이터 베이스)가 구현되어야 함
	속성 의미	DB에 저장된 속성의 데이터 값은 각기 독립된 의미를 가지고 있어야 한다.
일관성	고유성	보고된 데이터 중복이 없어야 하며, 각 데이터 레코드는 고유해야 함. 그렇지 않으면 오래된 정보에 액세스할 위험이 높아지므로, 데이터셋의 단일 뷰와 관련된 문제를 살펴보고 중복 사례가 없는지 확인하여야 함
	속성	속성 무결성(Domain Integrity)으로 기관 내 혹은 범정부 단위에서 같은 이름을 갖는 속성은 같은 의미로 사용되어야 하며, 같은 의미의 속성은 같은 이름과 형식으로 사용해야 한다.
	표준	기관 내 표준 데이터 항목을 정의해야 하며, 범정부 단위에서 준수해야 할 행정 표준 항목(용어, 코드)을 식별 및 준수해야 한다.
	연계값	흐름 데이터는 기관 간 혹은 기관 내 시스템 간 연계에 의해 DB간 이동이 일어나는 데이터로, 타 기관에 제공하는 데이터와 타 기관으로부터 제공받은 데이터의 값은 데이터 연계 절차와 규칙이 정의되고 준수되어야 한다.
	중복값	중복 데이터는 하나의 DB 내에서 원(Original) 값이 존재하고 이를 복사하여 사용하는 값들로, 같은 시점의 중복 데이터 값은 동일한 값으로 유지되어야 한다. (중복 값에는 테이블, 컬럼 포함)
유용성	이식성	기존 품질을 유지하면서, 한 시스템에서 다른 시스템으로 데이터를 설치, 교체 및 이동할 수 있는 정도, 다양한 상황에서 형식을 적용할 수 있는 정도
	충분성	사용자가 만족할 수 있는 충분한 데이터의 양을 수집 및 제공해야 한다. 또한, 충분한 정보 항목도 제공되어야 한다.
	접근성	데이터는 사용자의 다양한 요구에 유연하게 대응할 수 있는 형식과 구조여야 한다.
활용성	활용성	정보의 가치 측정을 위해 제공하는 정보의 사용자 만족도와 활용 현황을 측정하여야 한다.

적시성	최신값	실제 업무에서 변경된 값을 DB에 반영하기 위한 규칙과 절차가 정의되어 있어야 한다.
	응답시간	정보는 사용자가 만족하는 응답 시간 (성능) 내에 전달되어야 한다.
	데이터 제공	사용자에 의한 새로운 정보 요구가 있을 때, 서비스하기 위한 시간(데이터 수집, DB저장, 가공)은 최소화되어야 한다.
기밀성	접근 제한	중요 데이터에 대한 보기 접근 권한이 설정되어야 함
	DB 보호	데이터를 저장하고 관리하는 DBMS는 외부의 훼손, 변조, 유출, 시스템의 에러, 장애에 대비하여 충분히 안전하게 보호되어야 함
	오너쉽	데이터와 표준 데이터 품질관리를 위해 정보 라이프사이클(생성, 갱신, 삭제)에 따른 담당자가 지정되고 관리되어야 한다.
무결성	개체 무결성	행을 특정 테이블의 고유 엔터티로 정의한다.
	참조 무결성	개체 간 참조 관계가 존재할 때, 두 개체 간 데이터가 일관성을 가질 수 있도록 보증한다.
	도메인 무결성	열에 대한 데이터 무결성을 보장하기 위한 것이다.
	무결성 규칙	데이터를 추가, 수정, 삭제하는 과정에 부적절한 자료가 입력되는 것을 방지하기 위해서 테이블을 생성할 때 각 컬럼에 대해 정의한다. (위의 세 가지 무결성에 해당하지 않지만, 사용자가 정의한 규칙에 의한 무결성으로 다양한 응용의 필요에 따라 정의)

지표 특성이 없는 단일 지표에는 20점, 지표 특성에는 각 5점을 부여하였다.
총점은 180점이다.

<표 3> 데이터 품질관리 및 평가 지표 측정 기준 및 배점표

지표	지표특성	측정 기준	배점	
			세부 배점	지표 총점
정확성	입력값	○ 오류데이터 입력 방지를 위한 예방적 조치가 충분하게 설정되어 있다. ○ 오류데이터가 입력된 경우, 원인을 파악하고 이를 보정하기 위한 절차가 있다. ○ 비정형 데이터와 이에 대한 정보를 관리하는 메타데이터는 정확성을 유지하고 있다. ○ 사용자 요구에 맞는 오류가 없는 정확하고 최신의 고품질 데이터를 제공하고 있다.	5	25
	업무규칙	○ 기관 업무와 관련한 법, 규정, 지침(이하 근거규정)이 명시된 업무규칙 정의서가 있다. ○ 근거 규정이 변경된 경우, 이를 업무규칙 정의서에 반영하기 위한 절차가 있다. ○ 업무규칙 정의서를 기반으로 데이터를 수집하고, 데이터의 정확성을 검증하고 있다. ○ 업무규칙 정의서에 위배된 데이터 발견할 시 이를 보정하기 위한 절차가 있다.	5	
	범위·형식	○ 테이블의 속성에 저장된 데이터는 현실 세계에서 사용하는 범위에 맞게 저장되어 있다. ○ 테이블의 속성에 저장된 데이터는 도메인 표준에 따른 형식에 맞게 저장되어 있다. ○ 테이블의 속성에 저장된 데이터가 범위·형식을 위배한 경우 이를 보정하기 위한 절차가 정의되어 있다.	5	
	참조 관계	○ 데이터 간 종속관계가 있거나 참조관계가 있을 경우 데이터베이스에 제약 조건이 설정되어 있거나 프로그램을 통해 검증하고 있다. ○ 데이터 간 종속 관계 및 참조 관계가 있으나 실제 참조되지 않는 데이터가 있을 때, 원인을 파악할 수 있다.	5	
	계산식	○ 원천 데이터를 특정 계산식에 의해 가공하여 활용하는 경우, 계산식을 관리하고 있고 이를 통해 데이터 검증을 할 수 있다.	5	
확실성		○ 일반 데이터 : 리딩 제로(Leading zeros)에 해당되는 경우, 숫자 코드 및 식별자에 일관적 사용 및 메타데이터에 명시적으로 기술되었다. ○ 지리공간 데이터 - 데이터의 패키지화 크기가 적절하다. - 값 속성 테이블의 최신 업데이트 값을 반영하였다. ○테이블 데이터 :포함된 경우 좌표, 투영 및 데이터가 정의되었다. ○관계형 데이터베이스	20	20

		<ul style="list-style-type: none"> - '표 형식 데이터' 섹션의 요구사항을 충족한다. <p>데이터베이스 정규화 규칙의 메타데이터 내에서 사용하고 이를 문서화하였다.</p> <ul style="list-style-type: none"> - 데이터 내보내기 날짜 혹은 데이터 스냅샷 날짜의 메타데이터 기록이 있다. <p>○ 차트 및 그래프 작성 시, 해석적인 주석이나 정보가 추가되지 않았다.</p>		
완전성	논리 모델	<p>○ 데이터 구조 설계 시 논리 데이터 모델을 사용하고 있다.</p> <p>○ 논리 데이터 모델은 모델링 방법론을 적용하여 충분히 상세한 수준으로 작성되어 있다.</p> <p>○ 반드시 입력해야 할 데이터 항목엔 필수 입력 제약 조건 (NotNull)이 설정되어 있다.</p>	5	25
	식별자	○ 논리 데이터 모델의 엔터티와 물리 데이터 모델의 테이블에는 데이터를 유일하게 구별할 수 있도록 식별자와 Primary Key (PK)가 정의되어 있다.	5	
	물리구조	<p>○ 물리 데이터 모델 설계도와 데이터베이스 구조가 일치한다.</p> <p>○ 데이터베이스의 구조 변경 시, 논리 모델과 물리 모델을 변경한 뒤 데이터베이스에 적용한다.</p>	5	
	속성의미	○ 테이블의 속성에 입력된 데이터는 정의한 목적과 기준에 맞게 저장되고 있다.	5	
	고유성	<p>○ 데이터에 중복된 기록 또는 기능 포함 여부를 알 수 있다.</p> <p>○ “NoData“, “Null“ 또는 공백값이 데이터 필드의 형식 및 내용 정의에 적절한 코드를 사용하여 표시되었다.</p>	5	
일관성	속성	<p>○ 파일명에 일관적인 명명 규칙을 사용하고 있다.</p> <p>○ 약어 및 코드를 정의 및 사용하고 있다.</p> <p>○ 대문자와 구두점을 일관적으로 사용하고 있다.</p>	5	25
	표준	<p>○ 업무변경 및 표준변경 등에 의한 논리 모델의 변경 및 추가 시 정보시스템에 어떤 영향이 있는지 알 수 있다.</p> <p>○ 데이터 표준 및 표준화 절차가 정의되어 있으며, 이를 준수하고 있다.</p> <p>○ 도메인 표준에 따라 같은 의미의 속성은 동일한 유효 범위와 형식을 일관되게 사용하고 있다.</p> <p>○ 기관의 표준 코드에 등록된 코드를 일관되게 사용하고 있다.</p> <p>○ 명명규칙에 따른 데이터 표준 준수 현황을 파악하고 있다.</p> <p>○ 공공기관을 위한 국가 차원의 행정표준코드 맞용어를 인지하고 있으며, 이를 적용하고 있다.</p>	5	
	연계값	<p>○ 기관 내부 및 외부 시스템과 연계하는 데이터의 종류와 연계 주기, 방법을 파악 할 수 있다.</p> <p>○ 수집 및 제공되는 연계 데이터 관리 절차가 있으며, 이를 준수하고 있다.</p> <p>○ 수집 및 제공되는 연계 데이터에 대해 표준을 준수하고 있다.</p> <p>○ 수집 및 제공되는 연계 데이터에 대해 데이터 값에 대한</p>	5	

		<p>정합성을 검증하고 있다.</p> <p>○ 외부로부터 수집되는 연계 데이터에 대해 내부데이터 흐름 및 변경 영향을 파악할 수 있다.</p> <p>○ 외부에 제공하는 데이터의 변경이 발생할 시, 해당 기관에 통보하는 절차가 정의되어 있다.</p>		5
	중복값	<p>○ 데이터베이스 내 중복시킨 데이터를 구분하고 정합성을 검증할 수 있다.</p> <p>○ 중복된 데이터가 정합성을 위배한 경우, 이를 보정하기 위한 절차나 방법이 있다.</p>		
	이식성	<p>○ 파일명을 사용자가 해석 가능하다.</p> <p>○ 파일이 대부분의 운영체제 및 소프트웨어 응용프로그램에서 사용가능하다.</p>		
유용성	충분성	<p>○ 사용자가 만족하는 수준의 충분한 데이터를 제공하고 있다.</p> <p>○ 사용자가 이용가능한 데이터를 파악할 수 있도록 정보를 제공하고 있다.</p> <p>○ 주기적으로 사용자 교육을 실시하고 있다.</p>	5	15
	접근성	<p>○ 사용자가 어떤 경로를 통해 데이터를 접근하고 있는지 파악할 수 있다.</p> <p>○ 데이터 접근 경로를 파악하고 있다면, 이를 서비스 개선에 활용하고 있다.</p> <p>○ 사용자 편의성을 높이기 위해 데이터를 다양한 형태로 제공하고 있다.</p>	5	
	활용성	<p>○ 사용자의 요구사항을 수집 및 관리하고 있다.</p> <p>○ 사용자에게 제공한 데이터는 어떤 용도로 활용되고 있는지 파악할 수 있다.</p> <p>○ 서비스하고 있는 데이터에 대한 사용자 만족도를 조사하고 있다.</p> <p>○ 데이터 제공이나 오류 처리 등을 위한 처리 절차가 있으며, 이러한 처리 절차가 준수되고 있다.</p>	5	
적시성	최신값	<p>○ 사용자의 데이터 처리 요청은 처리기한 내에 반영되고 있다.</p> <p>○ 주요 데이터의 갱신 주기와 방법이 문서화되어있고, 이를 적용하고 있다.</p>	5	15
	응답시간	<p>○ 데이터에 대한 응답시간의 기준이 마련되어 있다.</p> <p>○ 응답시간 기준에 따라 성능을 측정하고 있다.</p> <p>○ 정보시스템의 응답 시간을 개선하기 위한 활동을 수행하고 있다.</p> <p>○ 물리데이터 모델 설계 시 데이터 용량 및 성능 등 확장성을 고려하여 설계하고 있다.</p>	5	
	데이터 제공	<p>○ 사용자에게 데이터를 제공하기 위한 데이터 제공 유형별 기한이 정의되어 있으며 이를 준수하고 있다.</p>	5	
기밀성	접근 제한	<p>○ 데이터 품질관리 표준 문서에 대한 사용자 접근 권한이 설정되어 있다.</p> <p>○ 데이터에 대한 접근 권한이 설정되어 있다.</p> <p>○ 데이터 접근에 대한 사용이력 (Log)을 관리하고 있다.</p>	5	15

	DB 보호	<ul style="list-style-type: none"> ○ 정전, 재해, 재난에 대비하여 데이터베이스관리 시스템은 안전한 장소에서 관리되고 있다. ○ 데이터베이스 백업 정책이 수립되어 있고, 이를 통해 백업되고 있다. ○ 개인정보 보호 대상 데이터가 파악되고, 접근제한 및 암호화 등을 적용하고 있다. ○ 데이터에 관한 보안 감사를 수행하고 발견된 취약점을 개선하고 있다. ○ 데이터별 폐기 기준을 정의하고, 폐기 절차에 따라 이관, 삭제 등을 수행하고 있다. ○ 비공개대상(저작권, 개인정보 등) 정보를 파악하여 관리하고 있다. 	5	
	오너쉽	<ul style="list-style-type: none"> ○ 데이터에 대한 관리 담당자가 지정되어 있다. ○ 데이터 소유자(오너쉽)에 의해 데이터의 통제(등록, 변경, 삭제)가 이루어지고 있다. 	5	
무결성	개체 무결성	○ 모든 테이블이 기본 키 (primary key)로 선택된 필드 (column)를 가져야 함. 기본 키로 선택된 필드는 고유한 값을 가져야 하며, 빈 값은 허용하지 않는다.	5	20
	참조 무결성	○ 관계형 데이터베이스 모델에서 참조 무결성은 참조 관계에 있는 두 테이블의 데이터가 항상 일관된 값을 갖도록 유지된다.	5	
	도메인 무결성	○ 테이블에 존재하는 필드의 무결성을 보장하기 위한 것으로 필드의 타입, NULL 값의 허용 등에 대한 사항을 정의하고, 올바른 데이터가 입력되었다.	5	
	무결성 규칙	○ 데이터베이스 전체에 공통적으로 적용되는, 데이터의 무결성을 지키기 위한 모든 제약 사항이 있다.	5	
적합성		<ul style="list-style-type: none"> ○ 지리 공간 데이터의 좌표가 적절하고 정의된 형식을 사용하고 있다. ○ 지리 공간 데이터/Raster 데이터가 파일 크기를 최소화하기 위해 데이터에 사용되는 적절한 비트/바이트 깊이로 데이터를 압축하였다. ○ 지리 공간 데이터/Vector 데이터가 shapefile, geopackage, GeoJSON, KML 또는 GML 파일로 제공되었다. ○ 데이터 품질 제어표시가 되어 있다. ○ 데이터의 품질 관리 및 품질 보증에 대한 정보에 대한 메타데이터가 있다. 	20	20
총점				180

5개의 세부지표(입력값, 업무규칙, 범위 · 형식, 참조 관계, 계산식)

총점 20점

가. 【 입력값 】

측정 방법 (5점)

- 오류데이터 입력 방지를 위한 예방적 조치가 충분하게 설정되어 있다.
- 오류데이터가 입력된 경우 원인을 파악하고 이를 보정하기 위한 절차가 있다.
- 비정형 데이터와 이에 대한 정보를 관리하는 메타데이터는 정확성을 유지하고 있다.
- 사용자 요구에 맞는 오류가 없는 정확하고 최신의 고품질 데이터를 제공하고 있다.

가이드라인

- 데이터 입력 시 필수 데이터 누락을 방지하기 위한 방법이 하나 이상 정의되고 이를 위한 사항들이 문서화되었는지 확인함
- 데이터 입력 시 데이터 값이 유효성 지표(범위와 형식 특성)에 맞는지 검증하는 방법과 사용자가 입력오류를 정정할 수 있는 프로그램이나 절차가 존재하는지 확인함
- 사진 등 이미지 데이터를 입력할 때의 규칙과 절차가 있는지 확인함
- 품질진단 대상 및 기준을 활용하여 데이터값 진단을 실시하고, 오류데이터를 식별하여야 함
- 데이터를 수집하거나 편집하기 전에 데이터를 검토하는 방법에 대한 계획을 세워 주어진 데이터셋에서 발생할 가능성이 있는 오류, 충돌 및 기타 데이터 문제의 종류에 대해 체계적으로 확인해야 함
- 결과 데이터 및 메타데이터와 연결될 때 도큐멘테이션된 품질 관리 절차는 데이터셋의 내용에 대한 완전한 이해가 가능한지 확인함
- 데이터 품질 보증 및 관리를 도큐멘테이션하는데 유용한 접근 방식은 데이터를 평가하기 위해 취한 조치, 문제 해결에 관한 결정이 내려진 방법, 해결하기 위해 취한 조치를 나열하는지 확인함
- 데이터 수명 주기의 각 단계에서 발생하는 문제 품질 관리 및 보증에는 다음과 같은 사항이 포함되어야 함
 - ✓ 잠재적으로 오류가 있는 데이터를 식별하는 방법
 - ✓ 잘못된 데이터를 처리하는 방법
 - ✓ 문제가 있는 데이터가 표시되는 방식(예시: 플래그 지정)

예시

- 데이터 입력 화면에서 필수 입력 항목에는 ‘*’가 표시되어있고 프로그램, DB에서 각각 검증함
- 사용자 데이터 입력 시 오류에 대비한 온라인 헬프데스크 또는 콜센터 구축

나. 【 업무규칙 】

측정 방법 (5점)

- 기관 업무와 관련한 법, 규정, 지침(이하 근거규정)이 명시된 업무규칙 정의서가 있다.

- 근거 규정이 변경된 경우 이를 업무규칙 정의서에 반영하기 위한 절차가 있다.
- 업무규칙 정의서를 기반으로 데이터의 수집하고, 데이터의 정확성을 검증하고 있다.
- 업무규칙 정의서에 위배된 데이터 발견할 시 이를 보정하기 위한 절차가 있다.

가이드라인

- 해당 기관의 고유한 업무 수행을 위한 법·규정의 근거 조항이 명시된 업무규칙 정의서 존재를 확인함
- 해당 기관 업무와 관련된 법·규정이 변경되었을 때, 이를 업무규칙 정의서에 반영하는 절차가 있는지 확인하고 정의된 업무규칙에 의해 데이터가 수집되었는지 확인함
- 업무규칙대로 데이터가 저장되었는지 검증하는 절차나 방법을 사용하였는지 확인함
- 업무규칙에 위배되는 데이터를 발견 시, 이를 정정할 수 있는 절차가 있는지 확인함
- 업무규정에 따라 개발된 품질진단 규칙을 설정하여 품질진단을 수행하여야 하며, 업무규칙을 활용하여 데이터 정합성 점검을 수행한 내역이 반드시 확인되어야 함
- 스크립트 및 코드에서는 다음과 같은 사항이 확인되어야 함
 - ✓ 일반 작성 지침 : 인라인 주석/의견 내용에는 USGS 데이터 배포에 적합하지 않은 비공식적인 구어체나 문구가 포함되지 않았는지 확인함
 - ✓ 소스 기여: 주어진 소프트웨어 언어의 프로그래밍 언어, 라이브러리 및 패키지 버전이 문서화되어 있는지 확인함
 - ✓ 품질 관리 작성: 스크립트/코드 및 인라인 주석에는 철자 또는 문법 오류가 포함되지 않았는지 확인함
 - ✓ 코드 주석: 인라인 주석이 제공되며 전문적이고 간결한 언어를 사용하여 코드를 이해하는 데 필요한 정보를 전달하는지 확인함
 - ✓ 쓰기 명확성 : 주석 및 기타 문서에 사용된 약어 및 전문 용어가 정의되어 있는지 확인함
 - ✓ 소프트웨어 고지 사항: 적절한 USGS 소프트웨어 고지 사항을 포함하는지 확인함
 - ✓ 데이터 추론: 모든 소스 데이터 및 버전에 대한 참조의 문서화를 확인함

예시

- 매일 정시에 기온은 측정되어야 한다. (DB에 저장된 측정 시간이 매 정시에 저장되었는지 확인하는 것은 업무규칙을 확인하는 것)
- 총 급여는 총 근로시간*시간당 최저임금 이상이어야 함

다. 【 범위 · 형식 】

측정 방법 (5점)

- 테이블의 속성에 저장된 데이터는 현실 세계에서 사용하는 범위에 맞게 저장되어 있다.
- 테이블의 속성에 저장된 데이터는 도메인 표준에 따른 형식에 맞게 저장되어 있다.
- 테이블의 속성에 저장된 데이터가 범위·형식을 위배한 경우 이를 수정보정하기 위한 절차가 정의되어 있다.

가이드라인

- DB에 저장되는 값은 실제 업무 값과 같은 범위를 가지는지 확인함
- 데이터베이스에 저장되는 값의 형식은 동일한 속성에 동일한 형식으로 저장되는지 확인함

예시

- 2010년 2월 30일은 유효하지 않은 범위 값임
- 핸드폰 전화번호는 010-9999-9999, 011-999-9999 모두 유효함
- 동일한 금액 속성을 A 테이블에서는 ‘원’ 단위로 저장, 통계테이블에서는 ‘천’ 단위로 저장하는 오류가 있음

라. 【 참조 관계 】

측정 방법 (5점)

- 데이터 간 종속 관계 혹은 참조 관계가 있을 경우, 데이터베이스에 제약 조건이 설정되어 있거나 프로그램을 통해 검증하고 있다.
- 데이터 간 종속 관계 및 참조 관계가 있으나 실제 참조되지 않는 데이터가 있을 때 원인을 파악할 수 있다.

가이드라인

- 논리적으로 관계가 있는 테이블들의 속성이 참조 무결성을 준수하는지 확인함 (같은 형식과 범위, 값 존재)
- 한 번도 참조되지 않는 값이 존재한다면 불필요한 데이터일 가능성이 있으므로, 유지를 위해서는 그 사유를 알 수 있어야 함
- 비정형 데이터를 관리하는 메타정보는 정형으로 관리하는지 확인함

예시

- 참조키 (Foreign Keys)가 없다면, 참조값은 참조되는 값의 범위 내에 존재해야 하며, ERD나 테이블 정의서에서 확인 가능함

마. 【 계산식 】

측정 방법 (5점)

- 원천 데이터를 특정 계산식에 의해 가공하여 활용하는 경우, 계산식을 관리하고 있고 이를 통해 데이터 검증을 할 수 있다.

가이드라인

- 원천 데이터를 가공한 데이터의 경우, 가공식을 이용해 재가공할 수 있는지 확인함 (재가동해도 동일한 값이 산출되어야 함)
- 집계 함수를 이용한 가공의 경우 원 레코드 건수에 대해 유효하게 확인이 가능한지 확인함

예시

- 급여의 총액에서 원 값과 세금 공제항목은 구분되고 이에 대한 계산식이 문서화되어야 함
- 레코드 수를 더한 값은 더하기 전의 원천 레코드 값들에 대해 추출이 가능해야 함

2

확실성

데이터 유형에 따라 측정 방법이 다름

측정 방법 (20점)

- 일반 데이터 : 리딩 제로(Leading zeros)에 해당되는 경우, 숫자 코드 및 식별자에 일관적 사용 및 메타데이터에 명시적으로 기술되었다.
- 지리공간 데이터
 - ✓ 데이터의 패키지화 크기가 적절하다.
 - ✓ 값 속성 테이블의 최신 업데이트 값을 반영하였다.
- 테이블 데이터
 - ✓ 포함된 경우 좌표, 투영 및 데이터가 정의되었다.
- 관계형 데이터베이스
 - ✓ '표 형식 데이터' 섹션의 요구사항을 충족한다.
 - ✓ 데이터베이스 정규화 규칙의 메타데이터 내에서 사용하고 이를 문서화하였다.
 - ✓ 데이터 내보내기 날짜 혹은 데이터 스냅샷 날짜의 메타데이터 기록이 있다.
- 차트 및 그래프 작성 시, 해석적인 주석이나 정보가 추가되지 않았다.

3

완전성

5개 세부지표(논리모델, 식별자, 물리구조, 속성의미, 고유성)

총점 20점

바. 【 논리모델 】

측정 방법 (5점)

- 데이터 구조 설계 시 논리 데이터 모델을 사용하고 있다.
- 논리 데이터 모델은 모델링 방법론을 적용하여 충분히 상세한 수준으로 작성되어 있다.
- 반드시 입력해야 할 데이터 항목엔 필수입력제약(Not Null) 조건이 설정되어 있다.

가이드라인

- 설계된 논리모델이 논리적으로 적절히 분할되어 있어 이를 가공하고 활용하는데 충분한 단위로 정의하였는지 확인함
- 논리모델의 엔터티에서 반드시 값이 채워져야 하는 속성들이 정의되었는지 확인함
- 모든 엔터티들은 업무 의미를 표현하기에 논리적인 모순이 없는 관계로 연결되었는지 확인함

예시

- ✓ 엔터티는 3차 정규화 수준으로 정의되어야 함
- ✓ 관계형 데이터 모델의 경우 엔터티, 속성, 관계의 정의가 명확해야 함

- ✓ 각 엔터티에는 하나 이상의 필수 속성이 정의되어야 함
- ✓ 모든 엔터티는 기수성(Cardinality, 엔터티 간의 관계)이 정의된 관계에 의해 연결되어야 함

사. 【 식별자 】

측정 방법 (5점)

- 논리 데이터 모델의 엔터티와 물리 데이터 모델의 테이블에는 데이터를 유일하게 구별할 수 있도록 식별자와 Primary Key (PK)가 정의되어 있다.

가이드라인

- 엔터티 내의 레코드를 유일하게 구분할 수 있는 식별자가 정의되었는지 확인함
- 엔터티의 식별자가 일련번호라면, 일련번호 외의 속성에 레코드 단위(의미)를 구분할 수 있는 속성들이 정의되었는지 확인함

예시

- 주문번호가 식별자라면 주문자, 주문일, 주문상품 등을 알 수 있어야 함
- 사람의 이름, 건물 이름, 전화번호, 우편번호 등은 이력 데이터의 식별자로 추천하지 않음

아. 【 물리구조 】

측정 방법 (5점)

- 물리 데이터 모델 설계도와 데이터베이스 구조가 일치한다.
- 데이터베이스의 구조 변경 시, 논리 모델과 물리 모델을 변경한 뒤 데이터베이스에 적용하고 있다.

가이드라인

- 논리 데이터 모델에서 정의한 내역들이 물리 데이터 모델 및 데이터베이스와 정합성이 맞게 생성 및 반영되었는지 확인함
- 데이터 검증 규칙(Foreign Key, Check Constraints)을 정의하였는지 확인함
- 데이터베이스가 데이터베이스에 저장되는 데이터의 양과 처리 성능을 고려하여 설계되었는지 확인함 (수직분할, 수평분할 적용)
- 산출물 : 데이터베이스 정의서, 테이블 정의서, 컬럼 정의서 및 ERD¹⁾(또는 관계정의서) 정의 및 관리가 되었는지 확인함

예시

- 필수적으로 데이터가 채워져야 하는 속성들은 Not null로 정의함
- 파티션(Partition)이 적용되었다면 이에 대한 사유가 문서화되어야 함
- 참조키, 체크 제약을 적절하게 사용함

자. 【 속성의미 】

측정 방법 (5점)

1) ERD : 물리 데이터모델 다이어그램

- 테이블의 속성에 입력된 데이터는 정의한 목적과 기준에 맞게 저장되고 있다.

가이드라인

- 엔티티의 속성은 고유한 특성이나 성질을 갖고 더 이상 분해할 수 없도록 정의되었는지 확인함
- 논리 모델에 필수 입력 항목으로 정의되었다면, 물리 모델 또는 데이터베이스에 이를 위한 제약 조건을 정의하거나, 데이터가 반드시 저장 또는 확인 가능한지 확인함
- 만일, 하나의 저장소에 두 개 이상의 의미를 갖는 값을 저장한다면 이를 구분할 수 있는 보조 컬럼과 값이 정의되었는지 확인함
- 만일, 하나도 값이 저장되지 않은 컬럼이 있다면 이에 대한 사유를 파악할 수 있는지 확인함

예시

- 민원 접수일자에는 해당 민원을 접수한 일자만 저장되어야 함
- 나이는 현재의 나이의 정수로 된 표현이어야 하며, 생년월일이 기록되어서는 안됨
- 논리적으로 필수 항목이라면, Not Null로 선언되지 않았어도 값은 채워져야 함
- '구분'으로 정의된 속성에 저장된 값이 다른 유형의 의미를 갖는 값이 저장되어 있다면 이를 구별할 수 있는 별도의 속성이 정의되어 있어야 함 (상태에 따라 구분 코드값이 변하는 경우)

차. 【 고유성 】

측정 방법 (5점)

- 테이블의 속성에 입력된 데이터는 정의한 목적과 기준에 맞게 저장되고 있다.

가이드라인

- 중복 데이터가 존재하는 경우, 이를 예방하기 위해 사용할 수 있는 방법은 아래와 같음
 - ✓ 데이터 스토리지 차원의 중복 방지(예 : DBMS 사용 시 중복 입력 불가)
 - ✓ 데이터 원자화(Atomize): 데이터를 가장 작은 단위로 나눈 것을 정보 조각(piece)이라고 칭함. 각 정보 조각의 변경 사항을 문서화하여 기록
 - ✓ 이중 입력 : 동일한 데이터를 2명 이상의 입력자 및 컴퓨터가 입력한 후, 입력한 데이터를 비교함. 비교 시 공백 및 추정치가 나타날 경우 문서화
 - ✓ 데이터 열람 기록 및 녹음본의 전사(transcribe)
 - ✓ 일관된 용어 사용
 - ✓ 중복 오류 검사 시행
 - ✓ 필요한 경우 실행 취소 허용
 - ✓ 반복적으로 입력해야 하는 항목의 횟수 최소화

4

일관성

5개의 세부지표(속성, 표준, 연계값, 중복값, 이식성)

총점 25점

카. 【 속성 】

측정 방법 (5점)

- 파일명에 일관적인 명명 규칙을 사용하고 있다.
- 약어 및 코드 정의 및 사용하고 있다.
- 대문자와 구두점을 일관적으로 사용하고 있다.

가이드라인

- 서로 다른 엔티티들 간에 속성이 같은 의미를 갖는다면, 같은 이름으로 정의되었는지 확인함
- 같은 의미의 속성 또는 컬럼은 동일한 유효 범위와 형식을 사용하는지 확인함

예시

- 주민등록번호는 행정 표준 단어 주민+등록+번호의 합성어이고 영문 약어명도 법무부 표준용어로 일치함
- 주민등록번호는 어느 엔티티에서 사용하든지 범위와 형식이 동일하게 정의되어야 함

타. 【 표준 】

측정 방법 (5점)

- 업무변경 및 표준변경 등에 의한 논리 모델의 변경 및 추가 시 정보시스템에 어떤 영향이 있는지 알 수 있다.
- 데이터 표준 및 표준화 절차가 정의되어 있으며 이를 준수하고 있다.
- 도메인 표준에 따라 같은 의미의 속성은 동일한 유효범위와 형식을 일관되게 사용하고 있다.
- 기관의 표준코드에 등록된 코드를 일관되게 사용하고 있다.
- 명명 규칙에 따른 데이터 표준 준수 현황을 파악하고 있다.
- 공공기관을 위한 국가 차원의 행정표준코드 및 용어를 인지하고 있으며, 이를 적용하고 있다.

가이드라인 (5점)

- 데이터 표준 (단어, 용어, 도메인, 코드)에 대한 적용 준수율을 주기적으로 점검함
- 비정형 데이터에 대하여 데이터 종류별 표준 포맷, 사이즈, 명명규칙을 적용하고 준수 사항을 확인함
- 범정부 행정표준용어 행정표준코드 및 데이터 표준을 참고하여 기관의 데이터 표준(표준 용어 및 표준코드)를 정의하고 준수 정도를 확인해야함.
- 데이터 구조 관련 산출물은 데이터베이스 정의서, 테이블정의서, 컬럼정의서 및 ERD(또는 관계정의서)를 가리킴. 진단대상 DB의 이러한 구조 산출물이 실제 DB의 구조에 맞게 현행화하여 관리하고 있는지 확인함
- 데이터 표준 및 진단대상 DB는 사전에 정의된 데이터 표준을 적용하여 구축해야 함
- 정량적 데이터 표준 적용률(%)을 확인하여, 품질관리 수준 진단 대상DB(컬럼)에 데이터 표준을 어느 정도 수준으로 적용하고 있는지 확인해야 함
- 범정부 데이터 품질진단 기준(12개)으로 오류율을 측정함
- 개방 표준 데이터셋을 업데이트 주기에 따라 현행화하여 제공해야 함
- 기관이 품질관리 수준 진단 대상 DB에 대하여 기관표준 및 범정부 표준을 적용하여 데이터 표준을 정의하였는지 확인하고, 품질관리 수준진단 대상 DB(컬럼)에 데이터 표준을 어느 정도 수준으로 적용하고 있는지 확인함
- 관계형 데이터베이스의 스키마 존재 여부와 ERD, 데이터 사전이 제공되는지 확인함

예시

- 기관 내에서 남녀 성별을 구분하는 표준 코드가 정의되어 있어야 함
- 범정부 단위에서 ‘주거지’ 는 같은 의미로 사용되어야 함
- 새로 개발하는 모든 신규 시스템은 표준 코드, 도메인, 단어, 용어를 100% 준수하여야 함

파. 【 연계값 】

측정 방법 (5점)

- 기관 내부 및 외부 시스템과 연계하는 데이터의 종류와 연계 주기, 방법을 파악할 수 있다.
- 수집 및 제공되는 연계 데이터 관리 절차가 있으며, 이를 준수하고 있다.
- 수집 및 제공되는 연계 데이터에 대해 표준을 준수하고 있다.
- 수집 및 제공되는 연계 데이터에 대해 데이터 값에 대한 정합성을 검증하고 있다.
- 외부로부터 수집되는 연계 데이터에 대해 내부 데이터 흐름 및 변경 영향을 파악할 수 있다.
- 외부에 제공하는 데이터의 변경이 발생할 시, 해당 기관에 통보하는 절차가 정의되어 있다.

가이드라인

- 내외부기관의 정보시스템(DB)과 연계되어 송수신하는 데이터의 목록을 정의하여 연계 데이터를 관리함
- 시스템 간 이동되는 데이터를 알고 연동주기와 방법이 문서화 되었는지 확인함
- 기관 간 연계되는 데이터 (제공하는 데이터, 제공받는 데이터)를 알고 연동주기와 방법을 문서화한 데이터 교환 기술서와 설계서를 작성하는지 확인함 (정보기술아키텍처 지침 중 DA 산출물로도 관리됨)
- 데이터 연동 중 발생하는 오류에 대한 정정 절차가 존재하고 이를 준수하고 있는지 확인함
- 기관 간 연계되는 데이터 중 행정 표준용어, 표준코드에 대한 준수율을 확인함
- 우리 기관을 포함하여 두개 이상의 기관에서 제공받은 데이터를 활용하는 방법이 정의되었는지 확인함

예시

- 기관 간 주고받는 건수나 수치인 경우는 합계가 일치함
- 주고받는 속성(칼럼)과 타입·길이가 변환 없이 동일한 표준을 사용함

하. 【 중복값 】

측정 방법 (5점)

- 데이터베이스 내 중복시킨 데이터를 구분하고 정합성을 검증할 수 있다.
- 중복된 데이터가 정합성을 위배한 경우 이를 보정하기 위한 절차나 방법이 있다.

가이드라인

- 동일 DB 내에서 두 개 이상의 테이블에 존재하는 중복 데이터의 중복된 값과 원값의 구분이 가능한지 확인함 (source와 target 데이터 추적)
- 중복되는 값들의 중복 사유와 갱신 주기, 원값과의 비교 방법이 문서화되었는지 확인함
- 원값과의 불일치가 발견되었을 때, 이를 교정하기 위한 규칙과 절차가 존재하는지 확인함

예시

- 같은 우편번호에 서로 다른 주소 명이 지정되어서는 안됨
- 주민원장 엔티티에 주민데이터 변경 이력의 최초등록일, 최종변경일이 중복저장되거나 이력관리 자체가 되지 않는 경우

거. 【 이식성 】

측정 방법 (5점)

- 대상 환경에 설치하여야 하는 소프트웨어는 설치 가능성을 수행하고 있다.
- 하나의 소프트웨어 구성 요소를 다른 구성 요소로 교체할 수 있다.
- 설치 가능성의 특성에는 설치를 위한 운영 체제 요구 사항, 애플리케이션 사용을 위한 브라우저 요구 사항, 메모리 또는 RAM 요구 사항, 설치 및 제거 절차, 설치 중단에 예외사항, 그리고 소프트웨어 설치를 위한 전제 조건 등이 있다.
- 시스템의 환경 적응성에 대한 특성으로 하드웨어 종속성, 소프트웨어 종속성, 표준 언어, 각 대상 환경과 시스템 통신, 종속성 캡슐화(encapsulation), 그리고 여러 시스템에 걸친 종속성 표현이 있다.
- 여러 시스템 간에 공통 통신 표준을 사용한다.

가이드라인

- 시스템이 각 대상 환경에 적응할 수 있는지 검증함
- 호환 가능한 시스템 구성 요소가 있는지 확인함
- 파일명을 사용자가 해석 가능한지 확인함
- 파일이 대부분의 운영 체제 및 소프트웨어 응용 프로그램에서 사용 가능한지 확인함
- 프로세서 속도, 디스크 공간 및 RAM, 모니터 해상도, OS 및 브라우저 버전과 같은 각 환경적 특성에 시스템을 이식할 수 있는지 확인함
- UI 및 기능적 특징과 관련하여, 애플리케이션 형태가 여타의 OS 및 브라우저와 유사한지 확인함
- 시스템의 모든 개별 구성 요소가 철저하게 테스트되고 서로 통합되어 하나의 큰 시스템을 형성하는지 필수적으로 확인하여야 함
- 두 개 이상의 구성 요소가 통신적인 문제 없이 서로 상호작용할 수 있는지 확인함
- 이전 구성 요소를 대체하는 구성 요소는 모든 대상 환경에서 이전 구성 요소와 동일한 결과를 생성해야 하며, 교체되는 구성 요소와 동일한 용도로 사용되는지 확인함
- 하위 시스템은 이상적으로 공통 스택 영역과 메모리를 공유하므로, 호환성은 여러 하위 시스템을 포함하는 대규모 시스템에서 특히 유용함
- 호환성 테스트 요구 사항에 실패한 구성 요소는 전체 시스템에 큰 영향을 미치므로, 각 구성 요소는 공통 리소스에 미치는 영향에 대해 철저히 검증하여야 함
- 이식성에 대한 검증은 여러 운영 체제 및 브라우저 버전, 그리고 소프트웨어 개발 라이프사이클 전반에 걸쳐 점진적이고 반복적인 방식으로 이루어져야 함

예시

- 안드로이드 운영 체제에서 사용할 수 있는 어플리케이션이 iOS 운영체제에서도 구동됨

5

유용성

3개 세부지표(충분성, 접근성, 활용성)

총점 15점

너. 【충분성】

측정 방법 (5점)

- 사용자가 만족하는 수준의 충분한 데이터를 제공하고 있다.
- 사용자가 이용 가능한 데이터를 파악할 수 있도록 정보를 제공하고 있다.
- 주기적으로 사용자 교육을 실시하고 있다.

가이드라인

- 사용자 만족을 위한 충분한 양의 정보를 적시에 수집하였는지 확인함
- 사용자가 활용하기에 충분한 양의 정보가 제공되었는지 확인함

예시

- 트랜잭션 데이터의 경우 예상 발생량을 알 수 있어야 함
- 연간 데이터의 월별 분석을 위한 충분한 데이터의 양이 수집되었음
- 연간 합계 자료로는 월별 세부 현황을 분석할 수 없음

더. 【접근성】

측정 방법 (5점)

- 사용자가 어떤 경로를 통해 데이터를 접근하고 있는지 파악할 수 있다.
- 데이터 접근 경로를 파악하고 있다면 이를 서비스 개선에 활용하고 있다.
- 사용자 편의성을 높이기 위해 데이터를 다양한 형태로 제공하고 있다.
- 내·외부기관의 정보시스템(DB)과 연계되어 송수신하는 데이터의 목록을 정의하여 연계데이터를 관리한다.
- 데이터는 기계가 읽을 수 있으며 개방형 형식으로 사용할 수 있다.
- 도큐멘테이션되었는지 확인한다.
- 지리공간 데이터는 십진법 형식으로 작성하고 있다.
- 데이터베이스가 공개 형식, 기계 가독형 및 비독점적 형식으로 되어 있다.

가이드라인

- 사용자가 접근 및 활용하기 쉬운 정보 단위(레코드 수준 분할, 속성 수준분할 등)로 제공되었는지 확인함
- 사용자가 접근 및 활용하기 쉬운 정보 형식(엑셀, 텍스트, 이미지, PDF, ...)으로 제공되었는지 확인함

- 데이터는 다운스트림 사용자가 접근할 수 있도록 구조화되고 패키징되어야 하며, 기계 가독형 및 개방적 형식으로 사용되는 것이 권장됨
- 지리공간 데이터는 심진법 형식으로 작성하여야 기계가 읽기에 용이하며, 다운스트림 사용에 적합함. 여러 형식(예: 다른 열의 도, 분, 초 및 소수 도)으로 좌표 정보를 제공하는 방법도 존재함
- 지리공간 데이터/Raster 데이터의 경우, GeoTiffs, HDF, netCDF, JPEG와 같이 널리 지원되는 공통 형식으로 제공되는 것이 권장됨
- 관계형 데이터베이스는 외래 키를 사용하여 테이블 간에 관계가 존재하는 테이블에 데이터를 저장함. 데이터베이스가 정적인 데이터 배포로 배포되는 경우, 데이터를 오픈 소스 응용 프로그램에 로드하는 방법과 지정된 소프트웨어 버전이 제공되어야 함.
- 관계형 데이터베이스의 스키마 존재 여부와 ERD, 데이터 사전이 제공되는지 확인함
- 관계형 데이터베이스의 데이터 유형이 메타데이터 및 데이터 문서에 기술되어 있는지 확인함
- 데이터가 관계형 데이터베이스를 반영하지 않지만 관계형 데이터베이스에서 추출되었을 수 있는 경우, 데이터를 일련의 플랫폼 파일(내재된 관계형 구조가 없는 테이블 형식 데이터)로 배포하는 것이 더 적절할 수 있음
- 차트 및 그래프의 기본 데이터가 데이터 배포 시에도 사용될 수 있는지 확인함

예시

- 새로 변경된 주소 체계는 도로명과 번지수가 분할되어 제공됨
- 수치로 제공되는 자료는 다양한 그래프로 표현 가능함
- 데이터를 csv, xls 포맷 등으로 제공하는 경우

러. 【 활용성 】

측정 방법 (5점)

- 사용자의 요구사항을 수집 및 관리하고 있다.
- 사용자에게 제공한 데이터는 어떤 용도로 활용되고 있는지 파악할 수 있다.
- 서비스하고 있는 데이터에 대한 사용자 만족도를 조사하고 있다.
- 데이터 제공이나 오류처리 등을 위한 처리 절차가 있으며, 이러한 처리 절차가 준수되고 있다.

가이드라인

- 데이터가 어떤 경로로 얼마나 자주 접근되고 있는지 조사함
- 데이터가 어떤 용도로 활용되는지 (정보 다운로드 목적) 조사함
- 제공하는 데이터의 사용자 만족도를 주기적으로 조사함

예시

- 우편번호가 얼마나 다운로드 되는지 조사함
- 기상청에 제공되는 날씨 데이터의 활용 유형을 파악함

3개의 세부지표(최신값, 응답시간, 데이터 제공)

총점 15점

머. 【 최신값 】

측정 방법 (5점)

- 사용자의 데이터 처리 요청은 처리기한 내에 반영되고 있다.
- 주요 데이터의 갱신 주기와 방법이 문서화 되어있고 이를 적용하고 있다.

가이드라인

- 주요 데이터에 대해 변경 주기가 관리되고 있는지 확인함
- 변경 주기에 따라 변경된 데이터를 반영하는 규칙이나 절차가 문서화되어 있는지 확인함
- 정의된 규칙과 절차에 따라 변경된 데이터 값을 반영하는지 확인함
- 업데이트 주기는 짧을수록 좋음

예시

- 대법원에서 이름이 변경되면 일 단위로 주민 정보에 반영되어야 함
- 동사무소에서 변경한 주민 정보는 가능한 빠른 시간 내 인터넷 증명서로 확인할 수 있어야 함

버. 【 응답시간 】

측정 방법 (5점)

- 데이터에 대한 응답시간의 기준이 마련되어 있다.
- 응답시간 기준에 따라 성능을 측정하고 있다.
- 정보시스템의 응답시간을 개선하기 위한 활동을 수행하고 있다.
- 물리 데이터 모델 설계 시 데이터 용량 및 성능 등 확장성을 고려하여 설계하고 있다.

가이드라인

- 화면이나 프로그램 단위에서 기대 응답시간과 제공 응답시간이 정의되고, 이에 대한 준수사항을 확인함
- 동영상 같이 비교적 용량이 큰 정보를 제공하기까지의 평균 기대시간에 대한 정의를 확인함
- 응답시간에 대한 사용자 만족도를 조사하고, 이를 개선하기 위한 활동을 확인함

예시

- 대개의 데이터 응답은 화면에서 3초 내로 처리됨
- 응답시간 지연에 대한 불만이 접수되면 이를 개선하기 위해 시스템을 튜닝함

서. 【 데이터 제공 】

측정 방법 (5점)

- 사용자에게 데이터를 제공하기 위한 데이터 제공 유형별 기한이 정의되어 있으며 이를 준수하고 있다.

가이드라인

- 요구사항에 대해 서비스하기까지 정의된 규칙과 절차가 있는지 확인함
- 요구사항 접수에서 서비스 제공까지의 평균 데이터 제공 시간을 측정하고 이를 줄이기 위한 활동 수행을 확인함
- 법령 변경 시 최종 업무 서비스에 적용하기까지 기간을 예측함
- 보안이 특별히 더 강조되는 데이터의 경우, 접근제한을 위한 사용자 계정 및 암호 관리 정책이 있으며 이에 대한 준수사항을 확인함

예시

- 감사기관이 요청한 자료는 감사 시작 이전에 제공되어야 유용함
- 센서스에서 수집된 연령별 인구수 정보는 공식 발표 일자에 공개되어야 함

7

기밀성

3개의 세부지표(접근제한, DB보호, 오너쉽)

총점 15점

어. 【 접근제한 】

측정 방법 (5점)

- 데이터 품질관리 표준 문서에 대한 사용자 접근 권한이 설정되어 있다.
- 데이터에 대한 접근 권한이 설정되어 있다.
- 데이터 접근에 대한 사용 이력(Log)을 관리하고 있다.

가이드라인

- 보안이 특별히 더 강조되는 데이터의 접근 제어를 위해 사용자 계정 및 암호 관리 정책이 있으며 이를 준수하는지 확인함
- 데이터 품질관리 기본 요소별 보기 접근 권한이 명시되어 있는지 확인함
- 유료로 제공되는 정보가 있다면, 이에 대한 권한 및 결제 과정이 투명하고 정확한지 확인함
- 사유지에서 캡처한 이미지는 토지 소유자의 동의 없이 공개해서는 안 됨
- 데이터에 개인 식별 정보(PII)를 포함한 독점 정보 또는 민감한 정보가 포함되어 있지 않은지 확인함
- 사진 및 그래픽, 이미지 자체 또는 포함된 메타데이터에 개인 식별 정보 또는 민감한 정보를 포함하고 있지 않은지 확인함
- 보안 및 관리 측면에서 스크립트 및 코드에 개인 식별 정보, 절대 파일 시스템 경로, 내부 서버 호스트명 또는 인터넷 프로토콜 주소 또는 사용자 이름/암호가 제거되었는지 확인함

- 정책 및 모범예시를 통해 정보의 기밀성을 보호함

예시

- DBMS에 역할(Role), 권한(Privilege)이 관리되어야 함
- DBMS 접근 계정과 암호가 주기적으로 갱신되어야 함

저. 【 DB 보호 】

측정 방법 (5점)

- 정전, 재해, 재난에 대비하여 데이터베이스관리 시스템은 안전한 장소에서 관리되고 있다.
- 데이터베이스 백업 정책이 수립되어 있고, 이를 통해 백업되고 있다.
- 개인정보보호 대상 데이터가 파악되고 접근제한 및 암호화 등을 적용하고 있다.
- 데이터에 관한 보안 감사를 수행하고 발견된 취약점을 개선하고 있다.
- 데이터별 폐기기준을 정의하고, 폐기 절차에 따라 이관, 삭제 등을 수행하고 있다.
- 비공개 대상(저작권, 개인정보 등) 정보를 파악하여 관리하고 있다.

가이드라인

- 공공데이터를 관리하는 DBMS는 통합 데이터 센터에서 관리되고 이의 상태를 관리하는 보안 담당자가 지정되어 있는지 확인함
- 데이터의 백업 주기나 백업 및 복구 방식을 정의하고 이를 준수하는지 확인함

예시

- DBMS를 위한 방화벽이 설치되어야 함
- DBMS가 적재된 서버는 충분히 안전한 장소에 설치되어야 함
- 데이터 센터에서 운용되는 DBMS에 대해서도 관리자가 지정되어야 함
- 보안 담당자는 DBMS 관리 환경 및 데이터의 백업, 복구 절차를 관리하여야 함

처. 【 오너쉽 】

측정 방법 (5점)

- 데이터에 대한 관리 담당자가 지정되어 있다.
- 데이터 소유자(오너쉽)에 의해 데이터의 통제(등록, 변경, 삭제)가 이루어지고 있다.

가이드라인

- 주요 업무 데이터에 대해 현업 담당자(데이터 오너쉽)가 배정되었는지 확인함
- 지정된 담당자에 의해서 데이터가 생성, 갱신 및 삭제되는지 확인함
- 업무 담당자가 변경되었을 때, 데이터 관점의 인계 절차 존재를 확인함

예시

- 데이터의 오류 문제 발생 시, 이에 대한 의사결정을 하고 정정할 담당자가 지정되어야 함

8

무결성

4개의 세부지표(개체 무결성, 참조 무결성, 도메인 무결성, 무결성 규칙)

총점 20점

커. 【 개체 무결성 】

측정 방법 (5점)

- 모든 테이블이 기본 키 (primary key)로 선택된 필드 (column)를 가져야 함. 기본 키로 선택된 필드는 고유한 값을 가져야 하며, 빈 값은 허용하지 않는다.

☐ 모든 행에 값이 있는지 확인함

터. 【 참조 무결성 】

측정 방법 (5점)

- 관계형 데이터베이스 모델에서 참조 무결성은 참조 관계에 있는 두 테이블의 데이터가 항상 일관된 값을 갖도록 유지된다.

참조하는 부모/마스터 값 목록이 있는 경우, 모든 자식(파생) 값이 일관성이 있는지 확인함

퍼. 【 도메인 무결성 】

측정 방법 (5점)

- 테이블에 존재하는 필드의 무결성을 보장하기 위한 것으로 필드의 타입, NULL 값의 허용 등에 대한 사항을 정의하고, 올바른 데이터가 입력되었다.

가이드라인

- 제한된 수의 값만 있는 열이 있는지 확인 및 검증함
- 해당 열의 데이터가 얼마나 완전한지 확인함
- 이상값이 있는지 확인함

허. 【 무결성 규칙 】

측정 방법 (5점)

- 데이터베이스 전체에 공통적으로 적용되는, 데이터의 무결성을 지키기 위한 모든 제약 사항이 있다.

가이드라인

- 완전성/작성률(fill rate), 유효성, 값 목록 및 빈도 분포, 패턴, 범위, 최대값 및 최소값, 참조 무결성과 같은 필수 측정값 등을 포함함
- 데이터 무결성의 이해를 위하여 데이터를 프로파일링하여야 함. 데이터 프로파일링은 데이터가 어떻게 생겼는지 상세하게 볼 수 있는 수단임
- 모든 데이터에 대해, 데이터가 일관된 방식으로 수집, 유지 및 사용할 것을 입증함

9

적합성

측정 방법 (5점)

- 지리 공간 데이터의 좌표가 적절하고 정의된 형식을 사용하고 있다.
- 지리 공간 데이터/Raster 데이터가 파일 크기를 최소화하기 위해 데이터에 사용되는 적절한 비트/바이트 깊이 및 데이터를 압축하였다.
- 지리 공간 데이터/Vector 데이터가 shapefile, geopackage, GeoJSON, KML 또는 GML 파일로 제공되었다.
- 데이터 품질 제어 표시가 되어있다.
- 데이터의 품질 관리 및 품질 보증에 대한 정보에 대한 메타데이터가 있다.

가이드라인

- 데이터 검토 또는 품질 보증의 일부로 잠재적인 문제를 체계적으로 분류하기 위하여 데이터 품질 제어를 표시함
- 잠재적인 데이터 사용자가 데이터 품질을 평가할 수 있어야 프로젝트 초기에 문제를 식별할 수 있음
- 데이터 품질 및 품질 관리에 사용되는 모든 방법은 다른 사람들이 독립적으로 데이터를 평가할 수 있도록 전달되어야 함
- 메타데이터에 적용된 품질 관리 방법 및 가정을 설명함
- 실용적인 경우, 코드를 포함하여 품질 분석을 수행할 때 사용된 소프트웨어를 설명하고, 품질 관리 분석을 수행한 사람, 수행한 시기 및 데이터 세트에 대한 변경사항을 메타데이터에 포함함
- 품질 분석에 사용된 기준 또는 시험자료를 기술함
- (한정자 플래그가 있는 데이터가 요약되어 파생 데이터 세트를 생성하는 경우) 파생 데이터 파일의 메타데이터에 플래그가 지정된 데이터 백분율과 누락 데이터 백분율을 포함함
- 고주파 관찰은 종종 다운 샘플링되며, 기본 데이터에서 얼마나 많은 데이터가 거부되었는지 아는 것이 중요함

예시

- 실행 가능한 경우 교정 곡선을 만드는 데 사용된 데이터를 포함할 수 있음

내용		출처	비고
정확성	입력값	preview data release checklist 2020	USGS
		Data Quality Assessment handbook	DataOne
		공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
	업무규칙	공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
	범위 · 형식	공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
	참조 관계	공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
	계산식	공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
확실성		preview data release checklist 2020	USGS
		Data Quality Assessment handbook	DataOne
완전성	논리 모델	preview data release checklist 2020	USGS
		공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
	식별자	공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
	물리 구조	공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
	속성 의미	공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
일관성	표준	preview data release checklist 2020	USGS
		공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
	연계값	Data Quality Assessment handbook	UNGM
		Dimensions of Data Quality(DDQ)	DAMA NL
		공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
	중복값	공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
	이식성	공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
		https://www.reseat.or.kr/portal/bbs/B0000261/view.do?nttlId=203298&menuNo=200019&optn18=&searchCnd=1&searchWrd=%EC%8B%9C%EC%8A%A4%ED%85%9C+%EC%9D%B4%EC%8B%9D%EC%84%B1&pageIndex=1	논문 ²⁾
유용성	충분성	preview data release checklist 2020	USGS
		공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
	접근성	preview data release checklist 2020	USGS
		공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
적시성	최신값	공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
		공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
		공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
	응답시간	공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
	데이터 제공	공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
기밀성	접근 제한	공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털

	DB 보호	공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
	오너쉽	공공데이터 품질관리 매뉴얼 v2.0 (2018)	공공데이터 포털
무결성	개체 무결성	https://www.precisely.com/blog/data-integrity/what-is-data-integrity	precisely
		https://terms.naver.com/entry.naver?docId=2839810&cid=40942&categoryId=32840	두산백과
	참조 무결성	https://www.precisely.com/blog/data-integrity/what-is-data-integrity	precisely
		https://terms.naver.com/entry.naver?docId=2839810&cid=40942&categoryId=32840	두산백과
	도메인 무결성	https://www.precisely.com/blog/data-integrity/what-is-data-integrity	precisely
		https://terms.naver.com/entry.naver?docId=2839810&cid=40942&categoryId=32840	두산백과
	무결성 규칙	https://www.precisely.com/blog/data-integrity/what-is-data-integrity	precisely
		https://terms.naver.com/entry.naver?docId=2839810&cid=40942&categoryId=32840	두산백과
적합성	preview data release checklist 2020		USGS
	Data Quality Assessment handbook		DataOne

2) A. Abran et al. (2013). 시스템 이식성 요구사항을 위한 표준 기반 참조 프레임워크. Computer Standards & Interfaces, 35, 380-395.

