

프로세스 기반의 전자카탈로그 데이터 품질관리

Process-based e-Catalog Data Quality Management

김선호(Sunho Kim)*, 이창수(Changsoo Lee)**, 이제현(Jehyun Lee)***

초 록

전자거래가 보편화 되고 전자카탈로그 데이터의 양이 증가하면서 데이터 품질을 관리할 수 있는 체계적인 접근 방법이 필요해지고 있다. 이러한 필요성에 따라 여기서는 프로세스 기반의 전자카탈로그 데이터 품질관리 프레임워크를 제시한다. 이것은 기업이나 조직의 전자카탈로그 시스템의 기대를 만족시키기 위해 지속적으로 수행하는 데이터 관리 및 개선 활동에 대한 방법이다. 이 프레임워크에서는 품질관리를 위한 대상이 데이터, 데이터 품질관리 항목, 데이터 품질관리 프로세스의 세 가지로 구분된다. 세 가지 관리 대상은 다시 관리조직의 레벨인 총괄데이터 책임자, 데이터 관리자, 사용자에 따라 세분화 되어 제시된다.

ABSTRACT

As electronic commerce becomes more common and the data volume of e-catalog increases, a systematic approach to data quality management is being required. Upon the necessity, we propose a process-based framework for e-catalog data quality management. This is the methodology for data management and improvement activities continuously performed to satisfy the expectation of industry to e-catalog systems. In the framework, contents for quality management consist of data, quality management items, and quality management processes. These are again subdivided according to organization levels, i.e, user, data administrator, and chief information officer.

키워드 : 전자카탈로그, 데이터품질관리, 프로세스, 프레임워크
e-Catalog, Data Quality Management, Process, Framework

* 교신저자, 명지대학교 산업경영공학과 교수

** 강릉원주대학교 산업정보경영공학과 교수

*** 명지대학교 산업경영공학과

2009년 04월 10일 접수, 2009년 07월 24일 심사완료 후 2009년 07월 31일 게재확정.

1. 서 론

오늘날 경쟁에서 살아남기 위해서 기업은 축적된 데이터를 정보로 가공 변환하여 의사결정에 도움을 줄 수 있는 능력이 있어야 한다. 따라서 기업들이 데이터 웨어하우징, 지식관리, CRM 등에 투자하여 보유한 대규모 데이터로부터 기업이 필요로 하는 지식 정보를 생성하려고 노력하고 있다. 특히, 전자거래 분야에서는 활성화를 위하여 e-마켓플레이스나 쇼핑몰에서 전자카탈로그 데이터의 양을 증가시켜왔다. 그러나 데이터가 증가할수록 관리하기 힘들어지고 부정확한 데이터가 많아지고 있다. 부정확한 데이터를 기반으로 비즈니스 의사결정이 이루어질 경우 기업에 미치는 영향은 매우 클 수 있다. 예를 들면, 데이터 품질 저하로 인해 발생된 데이터 손실 비용, 재작업 비용 등이 기관이나 기업의 연간 예산이나 수익의 10~25%에 이르거나, 데이터 품질 저하로 인한 고객 불만, 제품 관련 소송, 재작업 등의 비용이 평균 매출액 25~30%에 이르기기도 한다[1].

이러한 문제점을 해결하기 위해 현장에서는 데이터 품질관리를 위한 방법이나 활동을 필요로 하고 있다. 이러한 요구에 부응하여 프레임워크 분야에서 데이터 거버넌스 모델, 데이터 품질관리 모델, 데이터 품질관리 성숙 모델 등에 대한 연구가 진행되어 왔다. 데이터 거버넌스는 조직정보를 관리, 사용, 개선, 보호하는 프로세스에 새로운 규칙을 추가하기 위한 품질 통제 규칙이다. 데이터 거버넌스 모델은 정보의 일관성, 유용성, 보안성, 사용성, 품질 관리에 중점을 두고 있으며, 데이터 오너십(ownership)과 스튜어드십(stewardship)

의 개념과 밀접하게 연결되어 있다[19, 14]. 데이터 품질관리는 수시로 발생하는 데이터 품질에 대한 업무활동을 의미하며[17, 18, 20, 22, 2], 데이터 거버넌스의 하위 개념으로 간주된다[11, 13, 23, 24]. 데이터 품질관리 성숙 모델은 데이터 품질관리에 대한 체계적인 접근을 가능하게 하는 객관적인 기준에 의하여 데이터 품질관리 수준을 측정하고 더 나은 수준을 확보하기 위한 가이드라인을 제시하고 있다[14, 21, 3]. 이것은 단지 현재의 품질 수준을 측정하기 위한 것 뿐만 아니라 품질향상을 위해서 어떤 프로세스를 어떻게 개선해야 하는 지에 대한 방향을 제시하고 있다.

최근에는 ISO TC184/SC4/WG13에서 마스터 데이터와 엔지니어링 데이터에 대한 품질향상을 위하여 ISO8000 데이터 품질 표준을 개발하고 있다[15, 16]. 그러나 아직 ISO 8000에 대한 전체 프레임워크가 제시되지 않고 제한적으로 마스터 데이터 분야에 대해서만 개발되고 있다. 마스터 데이터 분야도 ISO22745 기술사전[17] 기반 하에 전자카탈로그 데이터 교환을 위한 데이터 품질에 초점을 맞추고 있어, 데이터 품질관리를 위한 프로세스 접근 방식은 고려되지 않고 있다. 특히, 전자거래 분야에서 양질의 전자카탈로그 데이터를 관리할 수 있는 체계적인 접근 방법이 제시되어 있지 않은 실정이다.

이러한 필요성에 따라 이 논문에서는 프로세스 기반의 전자카탈로그 데이터 품질관리 프레임워크를 제시한다. 이것은 기업이나 조직의 전자카탈로그 시스템의 기대를 만족시키기 위해 지속적으로 수행하는 데이터 관리 및 개선 활동에 대한 방법이다. 이 프레임워크는 전자카탈로그 데이터 품질 관리 대상을

관리 조직별로 구별하고 있다. 품질관리를 위한 대상은 데이터, 데이터 품질관리 항목, 데이터 품질관리 프로세스의 세 가지로 구분된다. 세 가지 관리 대상은 다시 관리조직의 레벨인 총괄데이터 책임자, 데이터 관리자, 사용자에 따라 세분화 되어 제시된다.

제 2장에서는 품질관리 대상을 데이터, 데이터 품질관리 항목, 데이터 품질관리 프로세스로 하는 전자카탈로그에 대한 데이터 품질관리 프레임워크를 제시하였다. 제 3장에서는 데이터의 종류를 표준데이터, 업무 데이터로 구분하여 설명하였다. 제 4장에서는 각 데이터별 품질관리 항목을, 제 5장에서는 데이터 품질관리 프로세스와 조직의 담당자별 책임과 역할을 제시하였다.

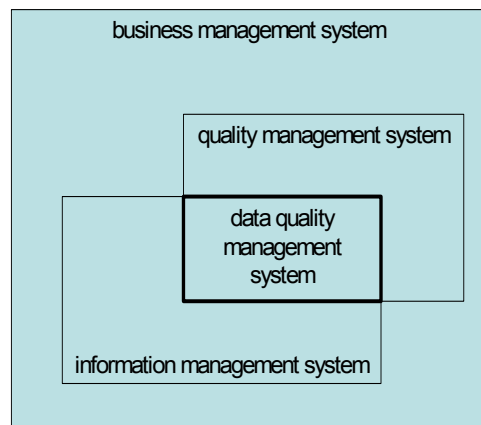
2. 전자카탈로그 데이터 품질관리 프레임워크

2.1 전자카탈로그 데이터 품질관리 프레임워크의 필요성

ISO 8000은 데이터 품질을 확보하기 위한 방안으로 출발하여 초창기에는 카탈로그 데이터의 품질, 특히 개방형 기술사전 표준인 ISO 22745에 의해 운영되는 기술사전 및 카탈로그 데이터에 대한 품질을 염두에 두고 제정되기 시작하였다. 그러나 표준의 내용이 점차 일반화 되면서 데이터 품질(Data Quality)에서 정보 품질(Information Quality)로 개념이 확대되고 대상도 마스터 데이터로 옮겨가면서 2006년 중반에 마련되었던 표준의 구조가 조금씩 변경되고 있다. 또한 ISO 8000과

품질경영 표준인 ISO 9000의 관계를 명확하게 하기 위하여 데이터 품질관리 시스템(data quality management system), 품질관리 시스템(quality management system), 정보관리시스템(information management system) 간의 관련성을 정의하였다. 일반적으로 데이터 품질관리는 정보관리 시스템의 일부분에 해당한다. 이러한 데이터의 품질은 기업 제품의 품질에 영향을 미치므로 데이터 품질을 다루는 ISO 8000은 제품의 품질관리를 다루는 ISO 9000의 범주 내에서 정의되어야 할 필요성이 제기되었다. 따라서 다음 <그림 1>과 같이 데이터 품질관리 시스템은 기업의 품질관리 시스템과 정보시스템의 공통부분으로 정의되어야 한다.

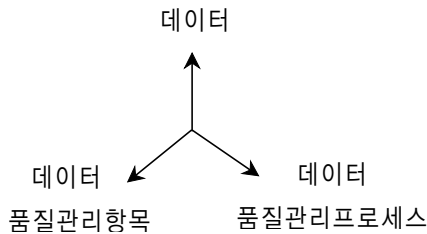
이러한 데이터 품질관리 시스템을 정의하기 위해 시스템 구성요소를 파악하고 데이터 품질의 정의, 대상, 관리항목, 관리 프로세스 등을 정립할 필요성이 제기되었다. 본 논문에서는 전자카탈로그 데이터를 대상으로 데이터 품질관리 시스템을 구성할 수 있는 프레임워크를 제공한다.



<그림 1> 데이터 품질관리 시스템의 영역

2.2 전자카탈로그 데이터 품질관리 프레임워크의 개념

전자카탈로그 데이터의 품질관리 대상은 <그림 2>와 같이 크게 데이터, 데이터 품질관리항목, 데이터 관리 프로세스의 세 요소로 구성된다.



<그림 2> 전자카탈로그 데이터 품질관리 구성 요소

데이터는 데이터 관리 조직 계층에 따라 표준 데이터와 업무 데이터로 구분된다. 또한, 데이터별로 관리해야 할 품질관리항목이 달라지며 데이터 관리 프로세스는 데이터 품질관리항목에 따라 결정된다. 따라서 여기서는 <표 1>과 같은 프레임워크를 제시할 수 있다.

이 프레임워크는 한국데이터베이스진흥센

터에서 제시된 모델[2]을 기반으로 하였으나 데이터 구조와 데이터 모델 관리자 및 DB 관리자를 고려하고 있지 않다. 데이터 구조는 DB 구조의 문제로서 DBMS의 기능을 이용하여 속성간의 관계를 통제하기도 하며 더 나아가면 구조를 재설계하고 궁극적으로 소프트웨어 재개발 문제까지 다루어야 한다. 따라서, 여기서는 데이터 구조와 관련성이 적으면서 전자카탈로그 콘텐츠 품질을 직접적으로 다루는 데이터 및 데이터 관리 프로세스를 고려하고 있다. 또한, 데이터와 데이터 품질관리항목 사이에 데이터 품질관리 항목을 추가하여 데이터 관리 프로세스가 만들어지는 근거를 제시하였다.

데이터 관리 조직에서는 총괄 데이터 책임자(Chief Information Officer), 데이터 관리자(Data Administrator), 사용자(User)로 구성된다. 이와 같은 분류는 EA(Enterprise Architecture) 개념에서 보면 개괄적(contextual) 관점, 개념적(conceptual) 관점, 그리고 운용적(operational) 관점에 해당된다. 먼저, 총괄데이터책임자는 전자카탈로그를 위한 데이터 관리의 총괄, 데이터 관리 정책 및 지원 마련, 데이터 관리자 간 이슈 사항 조정 등을 다루고 있다. 데이터 관리자는 전자카탈로그 관련

<표 1> 전자카탈로그 데이터 품질관리 프레임워크

관리조직 \ 관리대상	데이터	데이터 품질관리 항목	데이터 품질관리 프로세스
총괄 데이터 책임자 (개괄적 관점)	-	-	정책 관리
데이터 관리자 (개념적 관점)	표준 데이터	표준 데이터별 품질관리항목	데이터 표준 관리
사용자 (운용적 관점)	업무 데이터	업무 데이터별 품질관리항목	데이터 활용 관리

표준데이터에 대한 개발 및 관리, 검증 및 표준화 절차를 수립하고 운영한다. 또한 전자 데이터 모델을 통합하고 데이터 요구사항을 정리하며 기능별 데이터 관리자를 지원한다. 사용자는 전자카탈로그를 위한 업무 데이터, 즉 데이터 소스, 운영 데이터, 분석 데이터를 활용한다. 또한 데이터에 대한 추가 요건을 요청하고 데이터 활용도를 높여나간다.

3. 전자카탈로그 데이터 종류

데이터 품질 표준을 정의하고 있는 ISO 8000에서는 데이터 유형을 마스터 데이터(카탈로그 데이터), 제품 데이터, 거래 데이터로 크게 구분하고 있으며 본 논문은 마스터 데이터를 대상으로 프레임워크를 제시한다. 마스터 데이터가 가지고 있는 특성(예, 식별, 분류, 속성)을 중심으로 데이터 품질을 고려하며 데이터 품질을 위한 프로세스를 고려하고 있기 때문에 표준 대상인 카탈로그 데이터 자체 뿐만 아니라 품질관리를 수행하기 위한 절차에서 발생하는 업무 데이터도 고려한다.

다음 <표 2>에서는 본 논문에서 대상으로 하고 있는 데이터를 직접 데이터인 표준데이터와 관리 과정에서 발생하는 간접 데이터인 업무 데이터로 구분하여 설명하고 있다.

3.1 표준 데이터

표준 데이터는 정보시스템에서 사용하는 용어, 도메인, 코드 및 기타 데이터 관련 요소에 대해 공통된 형식과 내용으로 정의하여 사용하는 표준 관련 데이터를 의미한다. 전자 카탈로그에서 사용하는 표준 데이터는 식별 체계, 분류체계, 속성체계, 단어 및 용어 등 다양하게 존재한다[4]. 여기서는 ECIF에서 채택된 전자카탈로그 표준을 집중적으로 언급한다. 식별체계 표준 데이터인 GTIN(Global Trade Item Number, 국제무역물품번호)은 GS1에서 개발한 유통 물품에 대한 식별자이다 [8]. 분류체계 표준 데이터인 UNSPSC(United Nations Standard Products and Services Code, 국제연합 표준 제품 및 서비스 코드)는 전자상거래에서 제품 및 서비스를 분류하는데 사용되는 코드 시스템이다[5]. 공통속성 표준 데

<표 2> 전자카탈로그 데이터의 종류

구분	데이터 유형	참조 표준	비고
표준 데이터	식별표준 데이터	GTIN	데이터 기반
	분류표준 데이터	UNSPSC	
	공통속성표준 데이터	GDD	
	개별속성표준 데이터	ISO 22745	
	단어 및 용어사전 표준 데이터	ECIF 27	
업무 데이터	소스 데이터	데이터 품질 관리지침	프로세스 기반
	운영 데이터		
	분석 데이터		

이터인 GDD(Global Data Dictionary, 국제 데이터 사전)는 유통 물품과 거래 당사자 정보를 기술하기 위한 속성들의 집합이다[7, 9]. GDD는 GCI(Global Commerce Initiative)에 의해 만들어진 GDAS(Global Data Alignment System)를 기반으로 제시되었다. 개별속성 표준 데이터인 개방형 기술사전(ISO 22745)은 속성과 값을 한쌍으로 하는 방법을 사용하여 개인, 기구, 위치, 상품 및 서비스를 기술하는데 사용되는 개념의 용어, 정의 및 이미지를 포함하는 기술사전이다[10]. 이 외에도 전자상거래에 관한 주요 용어의 혼용을 방지하기 위한 단어 및 용어사전 표준 데이터로서 e-비즈니스 용어 v1.0이 있다[6]. 이것은 용어, 약어, 정의, 원전, 용어의 주석, 비고로 구성되어 있다.

3.2 업무 데이터

업무 데이터란 전자카탈로그와 관련하여 기관이나 기업의 업무 및 비즈니스를 수행하는데 필요한 데이터를 의미하며, 일반적으로 데이터 흐름에 따라 소스 데이터, 운영 데이터, 분석 데이터로 구분할 수 있다.

3.2.1 소스 데이터

소스 데이터는 운영 업무 데이터의 소스가 되는 데이터로 전자카탈로그를 구성하는 기초 데이터를 의미한다. 예를 들어, 제품 규격 데이터, 일반 문서, PC에 저장된 데이터 소스 파일, 이메일 및 팩스 등이 있다.

3.2.2 운영 데이터

운영 데이터는 기업이나 기관의 목표 달성

및 업무 수행을 위해 DB에서 저장·관리하여 활용하는 데이터를 의미하며 좁은 의미로는 전자카탈로그 운영 데이터를 의미한다. 여기에는 단순한 입·출력 작업 처리를 위해 일시적으로 필요한 임시 데이터는 제외한다.

3.2.3 분석 데이터

분석 데이터는 운영 데이터의 추출(extraction), 변환(transformation), 적재(loading) 등의 과정을 통해 생성되는 데이터를 말한다. 주로 전자카탈로그 시스템 또는 CMS(Content Management System)을 통해 구축된 데이터가 분석 대상이다. 분석 데이터가 기관이나 조직의 업무나 제반 활동을 신속하게 지원할 수 있도록 하기 위해서는 최신성과 정확성을 갖춰야 한다.

4. 전자카탈로그 데이터 품질관리 항목

전자카탈로그 데이터 품질관리 프레임워크 가운데 데이터 부문의 관리 대상인 표준 데이터 및 업무 데이터의 품질 관리 항목을 제시한다. 데이터 관리 정책은 데이터로 관리되는 것이 아니므로 제 5장 관리 프로세스에서 다루도록 한다.

4.1 전자카탈로그 데이터 품질관리 기준 항목과의 결합

전자카탈로그 데이터에 대한 품질관리 항목은 한국데이터베이스진흥센터의 데이터 품질관리성숙모형에서 언급하고 있는 데이터 품

〈표 3〉 데이터 유형과 품질관리 항목 체크리스트

구분	데이터 유형	품질관리 기준항목					
		정확성	일관성	유용성	접근성	적시성	보안성
표준 데이터	분류표준 데이터						
	공통속성 표준 데이터						
	개별속성 표준 데이터						
	단어 및 용어사전 표준 데이터						
업무 데이터	소스 데이터						
	운영 데이터						
	분석 데이터						

질의 6가지 기준인 정확성, 일관성, 유용성, 접근성, 적시성, 보안성을 원용하였다[3]. 이 6가지 기준을 본 프레임워크에서 제시하고 있는 표준 데이터와 업무 데이터에 결합하여 각 기준 항목별로 다음 <표 3>과 같은 점검표를 만들 수 있다. 본 점검표는 사용기관이나 사용대상에 따라 해당되는 품질관리 기준을 선택적으로 운용할 수 있으며 품질관리 기준을 보다 세분화시킬 수도 있다. 품질관리 기준의 실제 운용은 적용기준을 등급화하거나(예를 들어, 상, 중, 하 등의 구분) 정성적 적용 지침을 마련하는 것도 가능하다. 본 논문에서는 품질 관리 프레임워크에 포함될 데이터 유형을 제공하는 것이 범위이므로 기준 항목에 따른 세부 운용방안을 제시하는 것은 개별 운용 주체에게 위임한다.

4.2 표준 데이터 품질관리 항목

표준 데이터에 대한 품질관리 항목을 정하기 위해 5개의 데이터 유형을 크게 세부분으로 나눈다. 식별과 분류는 넓게 보면 공통 속

성에 포함 가능하므로 식별, 분류, 공통 속성을 한 단위로 묶는다. 나머지 개별 속성과 용어사전은 각각 한 단위로 구분한다.

4.2.1 식별/분류/공통속성 품질 관리 항목

식별, 분류, 공통 속성의 표준 데이터는 정성적인 기준보다는 정량적인 기준이 적용되는 데이터이다. 품질 관리 항목으로는 앞서 언급한 6가지 기준을 적용할 수 있으며 다음 <표 4>에는 이 중에서 일부만 선별하여 관리 지침과 적용 예를 설명하였다.

4.2.2 개별속성 품질 관리 항목

ISO 22745를 중심으로 한 개별 속성 표준 데이터의 품질관리 항목은 정보 제품 명세의 품질, 정보 콘텐츠의 품질, 정보 표현 품질로 나눈다. 이는 개방형 기술사전 표준인 ISO 22745가 제품 명세 및 표현 방법을 포함하기 때문에 콘텐츠의 품질과 함께 개별속성 품질관리 항목으로 설정하였다. ISO 22745의 내용에 따라 정보 제품명세의 세부기준을 정하였으며 정보콘텐츠는 6가지 기준관리 항목을

〈표 4〉 식별/분류/공통속성 품질 관리

대 상	관리항목	관리 지침	적용 예
식별/분류/공통속성 표준 데이터	유용성 (재사용성)	표준화 기구나 정부·공공기관에서 정의한 코드 활용	ISO 22745, UNSPSC 코드 사용
	일관성	하나의 값으로 통일	Y/N 또는 1/0
	일관성 (유일성)	동일한 내용의 코드 중복 사용방지	분류코드를 부여할 경우에 입시방편으로 유보코드를 남발하면 제품의 정확한 분류를 방해함
	유용성 (정보분석성)	가능한 범위의 데이터는 모두 코드화	전자카탈로그 데이터 속성값 입력시 필드 텍스트 보다는 코드값 체크리스트 기법을 사용

준용하였으며 정보표현은 표현성 한 가지만을 기준으로 정하였다.

4.2.3 단어 및 용어사전 품질 관리 항목

본 항목은 데이터 품질관리 지침[2]과 데이터 품질관리 성숙모형[3]에서 용어 표준관리의 내용을 전자카탈로그의 특성에 맞추어 관리 지침을 설정하였다. 또한 ECIF에서 발표된 e-비즈니스 용어를 대상으로 한다.

- 표준 용어 사전은 용어 사용이나 의미의 차이로 인해 발생하는 전사차원의 혼란을 최소화할 수 있어야 한다. 이를 위해 가능하면 유사어 사전, 동의어 사전, 이의어 사전을 구축할 필요가 있다.
- 용어가 지나치게 업무 관점으로 정의되어 직관적으로 이해가 힘들거나 의미상의 혼란을 초래해서는 안 된다. 일반적인 의미와 다르게 사용된 용어는 다른 용어로 대체해야 한다.
- 업무를 반영하여 용어 약어를 사용하거나 내부에서 별도로 정의된 용어를 사용하는 것은 자제한다. 단, 불가피하게

약어나 별도 정의된 용어를 사용할 경우는 각주와 용어사전을 통해 충분히 고지하도록 한다.

- 표준 용어는 전사적으로 사용되고 있는 개체와 속성의 용어를 대상으로 일반적으로 사용되는 표준 단어 사전의 정의되어 있는 단어를 조합하여 생성한다. 예를 들어, TTA의 정보통신 용어사전 등을 이용한다.
- 용어 사전은 개체 용어 사전과 속성 용어 사전으로 구분하여 정의·관리하며, 각각의 용어는 논리명(한글명)과 물리명(영문명)을 가지며 용어의 범위 및 자격 형식 등에 대한 설명을 포함하고 있어야 한다.

4.3 업무 데이터 품질 관리 항목

전자카탈로그 관련 업무데이터 품질관리 항목은 앞서 언급한 6가지 기준관리 항목과 결합하여 다음과 같은 프레임워크를 제시할 수 있다.

〈표 5〉 개별속성 품질관리 항목

개별속성 품질유형	세부 기준	고려 사항
정보 제품 명세	정보 표준 품질	전사적 관점 비즈니스 이해당사자의 승인필요 단일 목적형
	정보 이름 품질	비즈니스 용어 이름 품질 개체 유형 이름 품질 속성 이름 품질
	정보 정의 품질	비즈니스 용어 정의 품질 개체 유형 정의 품질 속성 정의 품질
	비즈니스 규칙명세품질	비즈니스 행동과 정보에 적용 정책, 규제, 기타 통제사항을 표현
정보 콘텐츠	정확성	소스에 대한 정확성, 현실에 대한 정확성 유지
	일관성	데이터 값이 정보 제품 명세 적합해야 값 유효성, 비즈니스 규칙 유효성, 파생 유효성 보장 데이터 값과 속성 정의의 일치 레코드의 유일성
	유용성	레코드 완전성 값 완전성
	접근성	데이터 값의 상세 수준
	적시성	정보 시간 지연의 최소화
	보안성	정보의 품질 보장
정보 표현	표현성	접근성 확득성 적절한 매체 업무 적절성 표현 방법의 표준화 교환성 구조의 표준화 가독성

전자카탈로그를 다루는 업무는 주로 CMS 이용, 카탈로그 배포, 제품 정보 제공 및 제품거래 등이며 이에 따른 소스 데이터, 운영 데이터, 분석 데이터의 정의 및 설명은 제 3.2

절에 언급하였다. <표 6>은 이러한 전자카탈로그 업무 관련 데이터가 기준 관리 항목을 적용하고 세부 고려사항을 제시함으로써 향후 체크리스트 형태로 각 항목을 점검할 수

〈표 6〉 업무데이터 품질관리 항목

업무 데이터 유형	관리 항목	고려 사항
소스 데이터	정확성	전자카탈로그의 효율적 검색 및 관리
	일관성	전자카탈로그 통합관리, 수명주기 및 권한 정의
	유용성	전자카탈로그 소스 데이터 관리 절차 제정 및 준수
	접근성	Fault Tolerant한 설계 및 실행 규칙
	보안성	시스템 보안 규정 준수
운영 데이터	정확성	전자카탈로그 소스 데이터와 동일
	일관성	전자카탈로그 속성 및 데이터 누락 방지
	유용성	전자카탈로그 운영 데이터 관리 기준
	접근성	전자카탈로그 사용자 인터페이스 제공
	적시성	전자카탈로그 데이터의 최신성, 데이터 책임제
	보안성	전자카탈로그 콘텐츠의 일치
분석 데이터	정확성	전자카탈로그 데이터 요구 수준
	일관성	전자카탈로그 표준 준수
	유용성	전자카탈로그 관리 규칙
	접근성	전자카탈로그 주제 및 영역별 분류
	적시성	전자카탈로그 분석 주기, 변환주기
	보안성	전자카탈로그 데이터 관리의 엄격한 절차

있는 토대를 제공한다.

5. 전자카탈로그 데이터 품질관리 프로세스

전자카탈로그의 데이터 품질관리 프로세스는 고품질 전자카탈로그 데이터의 운영·관리를 위한 데이터 관리 정책 수립에서부터 데이터 표준관리, 데이터 활용관리에 이르는 일련의 데이터 관리 활동들을 체계적인 흐름으로 표현한 프로세스이다<그림 3>. 여기서 전

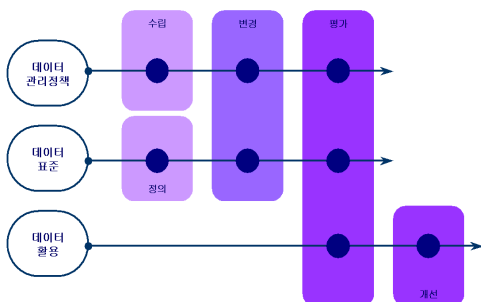
자카탈로그 데이터 관리 정책에 대한 관리 프로세스는 전자카탈로그 데이터 표준관리 및 데이터 활용관리의 기반이 되는 정책에 대한 수립, 변경, 평가 프로세스로서 데이터의 형태가 ‘비정형 텍스트’이며 업무적인 정책 및 규칙을 기술하고 있다. 따라서 전자카탈로그 데이터와 데이터 품질관리항목에서 언급하지 않았으나 본 장에서는 데이터 정책관리 프로세스에 대해서 언급하기로 한다. 여기서 포함되는 프로세스의 내용은 각각 다음과 같으며 구체적으로 세부 관리 대상, 업무 흐름도, 조직 담당자별 책임과 역할을 제시한다 :

- 전자카탈로그 데이터 정책 관리 : 수립, 변경, 평가
- 전자카탈로그 데이터 표준 관리 : 표준 데이터 정의, 변경, 평가
- 전자카탈로그 데이터 활용 관리 : 업무 데이터 활용 평가, 개선

전자카탈로그 데이터 정책 및 표준 관리에서는 ‘정책 수립’ 및 ‘정의’ 프로세스 이후에도 데이터를 변경할 수 있으므로 ‘평가’ 프로세스에 앞서 ‘변경’ 프로세스를 추가하였다. 평가 후에는 변경 프로세스로 돌아갈 수 있다. 전자카탈로그 데이터 활용관리에서는 데이터의 품질을 평가한 후에 개선하는 프로세스로 이어진다. 여기서 개선은 데이터 정책 및 표준 관리에서의 변경과 동일한 개념이나 프로세스의 순서를 나타내기 위해 다른 표현을 사용하였다.

5.1 전자카탈로그 데이터 관리 정책

전자카탈로그 데이터 관리 정책은 기업의 비전과 목표를 달성하기 위해 필요한 전자카탈로그 데이터 확보 계획과 확보한 데이터를

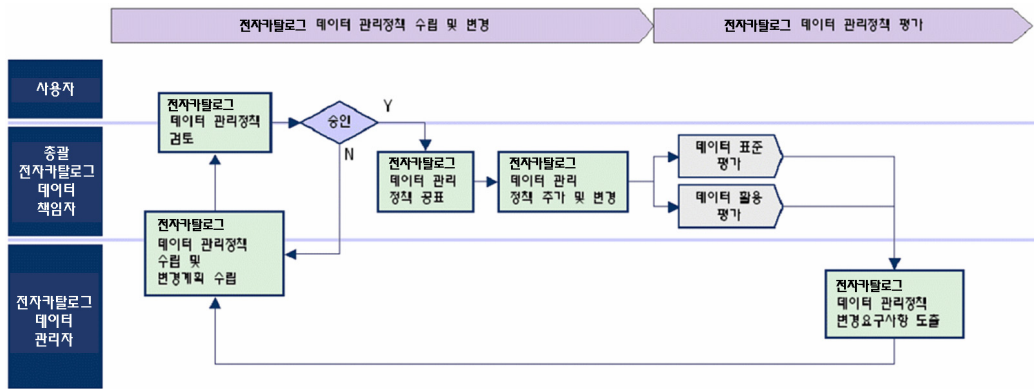


〈그림 3〉 전자카탈로그 데이터 품질관리 프로세스 범위

효과적으로 유지·관리하기 위한 체계 및 계획을 정의하는 작업을 말한다<그림 4>. 기업은 전자카탈로그에 대한 데이터 관리 정책을 수립함으로써 기업의 비전과 목표에 맞는 데이터를 확보하고, 확보된 데이터를 사용자가 원하는 시간에 원하는 데이터로 안정적으로 서비스할 수 있는 기본 원칙 수립과 관리체계를 구축할 수 있다. 그러나 전자카탈로그 데이터 관리 정책 자체는 일상적이원하반복을 사활동한 아니므로, 프로세스를 관리하는 것보다 데이터 관리 의해 관리활동이 잘 이루어지는지를 점검하는 것이 필요하다. 즉, 정기적·비정기적인 데이터 품질 평가 활동을 통한 검토 작업이 필요하다.

세부 관리 대상은 다음과 같다;

- ① 전자카탈로그 데이터 관리 원칙 : 전자카탈로그 데이터의 효과적인 확보, 유지, 관리를 위해 수립된 규정이나 계획, 지침 등에 포함된 데이터 관리 방향을 의미하며, 문서화(statement, rationale, implication)하여 관리하고, 변경은 데이터 관리와 관련된 담당자 및 사용자에 의해 이루어지도록 한다.
- ② 전자카탈로그 데이터 품질관리 메인 프로세스 : 고품질의 전자카탈로그 데이터를 지속적이고 안정적으로 서비스하기 위해 각 기관의 특성에 맞게 정의하고 정의한 프로세스 간의 연관관계를 정의한 프로세스를 의미하며, 전자카탈로그 데이터 관리 원칙에 준하여 전자카탈로그 데이터 관리 프로세스 목록을 도출하고 도출된 프로세스 간의 상호운용성을 고려하여 메인 프로세스를 정의한다.



〈그림 4〉 전자카탈로그 데이터 관리정책 업무 프로세스

〈표 7〉 전자카탈로그 데이터 관리정책 책임 및 역할

프로세스 조 직	수립 및 변경	평 가
사용자	<ul style="list-style-type: none"> 제안 정책 협의회 참여·검토 전자카탈로그 데이터 관리 정책 변경 협의 참여 및 검토 	-
총괄 데이터 책임자	<ul style="list-style-type: none"> 전자카탈로그 데이터 관리 정책 수립 전자카탈로그 데이터 관리 정책 변경계획 수립 전자카탈로그 데이터 관리 정책 수립 및 변경사항에 대한 공표 전자카탈로그 데이터 관리 정책 변경 요구사항의 타당성 검토 전자카탈로그 데이터 관리 정책 추가 및 변경 	<ul style="list-style-type: none"> 전자카탈로그 데이터 표준 평가(총괄) : 데이터모델 매핑, 데이터 표준 준수 체크, 변경 영향도 분석, 전자카탈로그 데이터 표준 미준수 원인 분석, 데이터 정제 전자카탈로그 데이터 활용 평가(총괄) : 핵심 데이터 수집, 활용도 측정기준 수립, 데이터 활용 측정, 활용저하요인 분석
데이터 관리자	<ul style="list-style-type: none"> 전자카탈로그 데이터 관리 정책 수립 지원 전자카탈로그 데이터 관리 정책 변경계획 수립 지원 정책에 따른 관련 프로세스의 수립 및 개선 변경된 정책을 위한 프로세스 수립 및 관리 감독 	<ul style="list-style-type: none"> 전자카탈로그 데이터 표준 평가 : 전자카탈로그 데이터 표준-개념 데이터 모델의 표준 준수 평가 전자카탈로그 데이터 활용 평가 : 전자카탈로그 데이터 활용관점의 사용자 대상 활용도 조사 평가 전자카탈로그 데이터 표준 평가 및 데이터 활용 평가 사항을 기반으로 전자카탈로그 데이터 관리 정책 변경 요구사항 도출 전자카탈로그 데이터 표준 평가 및 데이터 활용 평가 사항을 기반으로 전자카탈로그 데이터 관리 정책 변경 요구사항 도출

- ③ 전자카탈로그 데이터 관리 조직 : 각 기관에서 정의한 전자카탈로그 데이터 관리 프로세스를 지원하고 담당할 담당자와 조직을 정의하는 것을 말하며, 각 기업은 체계적으로 정의한 전자카탈로그 데이터 관리 메인 프로세스를 수행하기 위해 필요한 역할을 정의하고 정의된 역할을 수행할 수 있는 담당자를 선정해야 한다.

전자카탈로그 데이터 관리정책 업무 흐름도와 조직의 담당자별 책임과 역할이 <그림 4> 및 <표 7>에 요약되어 있다.

5.2 전자카탈로그 데이터 표준 관리

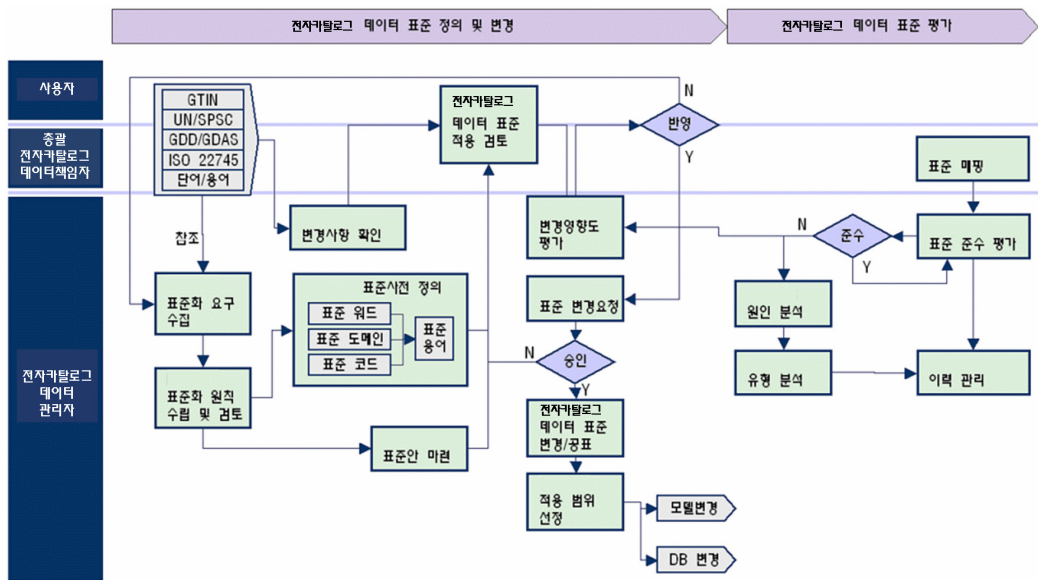
이 업무 프로세스는 전자카탈로그 데이터 표준화 원칙에 따라 정의된 표준 단어 사전 및 도메인 사전, 표준 용어 사전, 표준 코드, 데이터 관련요소 표준 등을 기관에 적합한 형태로 정의 및 변경관리하고, 전자카탈로그 데이터 표준 준수 평가 등을 통한 데이터 표준 제반활동을 말한다<그림 5>. 여기서는 DB의 설계와 개발을 지원하고 전사적인 데이터 표준의 사용 및 재사용을 통해 시스템 간 상호운용성, 데이터 공유, 시스템 통합, 비즈니스 프로세스 개선 등을 지원한다. 전사적으로 공통된 표준을 사용하게 함으로써 데이터의 일관성과 정합성을 유지할 수 있다.

전자카탈로그 데이터 표준관리 업무 흐름도와 조직의 담당자별 데이터 표준관리 책임과 역할이 <그림 5> 및 <표 8>에 요약되어 있다.

앞서 정의한 전자카탈로그 데이터 표준관

리를 세분화 하면 식별 코드관리, 분류 코드관리, 공통속성 관리, 개별속성 관리, 용어 표준관리로 분류가 되지만, 실질적인 전자카탈로그 데이터 표준 관리 프로세스 및 관리 문서 생성을 위해서는 세부 관리 대상이 표준 단어 사전(용어 표준관리 포함), 표준 용어 사전(용어 표준관리 포함), 표준 도메인 사전, 표준 코드(식별 코드관리, 분류 코드관리, 속성 코드관리, 개별속성 코드관리 포함), 데이터 표준 요소로 그룹핑 및 상세화 하여 재분류되며 각 관리대상의 설명은 아래와 같다. 전자카탈로그 데이터 활용관리 업무 흐름도와 데이터 활용관리 책임과 역할이 <그림 6> 및 <표 9>에 요약되어 있다.

- ① 표준 단어 사전 : 기업이나 기관에서 업무상 사용되며 일정한 의미를 갖는 최소 단위의 단어(word)를 정의한 사전을 말한다. 표준 단어를 정의함으로써 업무상 편의나 관습에 따라 동일한 단어를 서로 다른 의미로 사용하는 경우, 혹은 하나의 단어에 다양한 의미를 부여하여 사용하는 등의 문제를 방지할 수 있다.
- ② 표준 도메인 사전 : 도메인(domain)이란 속성에 정의된 조건을 만족시키는 값의 범위를 의미하며, 표준 도메인은 전사적으로 사용되고 있는 데이터 가운데에 논리적·물리적으로 유사한 유형의 데이터를 그룹화하여 해당 그룹에 속하는 데이터의 유형과 길이를 정의한 것을 말한다.
- ③ 표준 용어 사전 : 표준 용어는 업무에서 빈번하게 사용되며 표준 단어를 조합해 정의한 용어(term)를 의미한다. 단어는



〈그림 5〉 전자카탈로그 데이터 표준관리 업무 프로세스

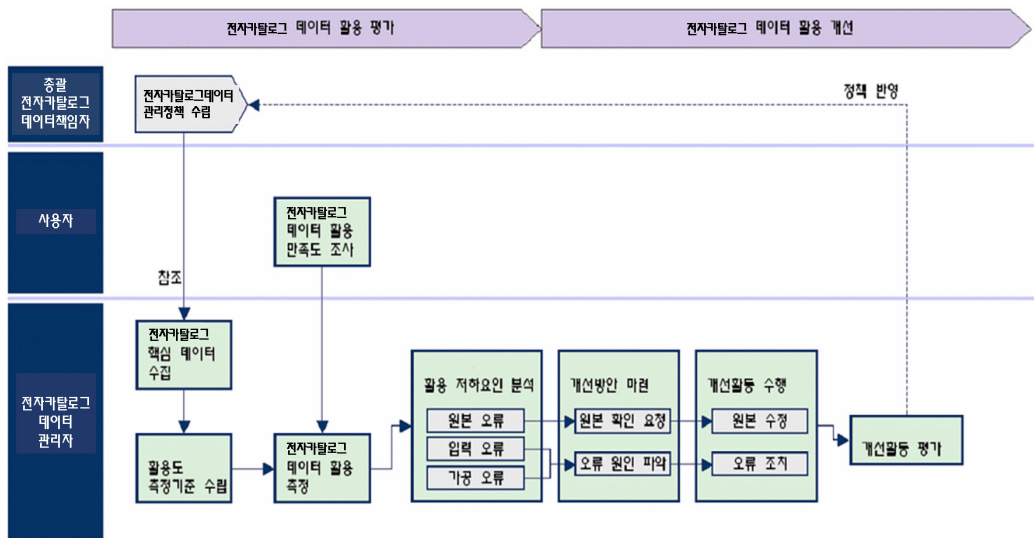
〈표 8〉 전자카탈로그 데이터 표준관리 책임과 역할

구분	프로세스	정의	변경	평가
사용자		전자카탈로그 데이터 표준 검토 및 승인	-	-
총괄 데이터 책임자		전자카탈로그 데이터 표준 검토 및 승인	-	전자카탈로그 데이터 표준 평가 : 데이터 모델 매핑
데이터 관리자		<ul style="list-style-type: none"> 전자 표준화 요구사항 수립 전자카탈로그 데이터 표준화 원칙 수립 표준 단어 사전, 도메인 사전, 코드 사전, 용어 사전 정의 전자카탈로그 데이터 표준 공표 	<ul style="list-style-type: none"> 표준화 요구사항 검토 표준 단어 사전, 도메인 사전, 코드 사전, 용어 사전, 전자 카탈로그 데이터 관련요소 변경 전자카탈로그 데이터 표준 등록 및 공표 	<ul style="list-style-type: none"> 전자카탈로그 데이터 표준-전자 카탈로그 데이터 모델 매핑 (시행) 전자카탈로그 데이터 표준 준수 체크 : 전자카탈로그 데이터 표준-개념 데이터 모델의 표준 준수 체크 변경 영향 분석 : 전자카탈로그 데이터 표준 변경 영향 분석

개별적이거나 용어는 업무와 조직의 성격에 따라 조합이 달라질 수 있다. 표준 용어 사전을 정의함으로써 기관이나 기업 내부에서 서로 상이한 업무 간의 의사소통이 필요한 경우 용어에 대한 이

해 부족으로 유발되는 문제점을 최소화할 수 있다.

④ 표준 코드 : 다양하게 나타날 수 있는 데이터 값을 정형화하기 위해, 정의된 기준에 따라 제한된 범위 내의 기호로 대



〈그림 6〉 데이터 활용관리 업무 프로세스

〈표 9〉 전자카탈로그 데이터 활용관리 책임과 역할

프로세스 조 직	평 가	개 선
사용자	◦ 데이터 활용 의견 제시	-
총괄 데이터 책임자	◦ 데이터 활용 만족도 조사	◦ 데이터 활용도를 개선하기 위한 활동 중 필요 사항을 데이터 관리 정책에 반영
데이터 관리자	◦ 핵심 데이터 수집 및 활용 측정 기준 수립 ◦ 데이터 활용 측정 ◦ 활용 저하 요인 분석	◦ 데이터 활용도 개선 활동 평가 ◦ 데이터 활용도 개선 방안 수립 ◦ 데이터 활용도 개선 활동 수행

치한 것을 의미한다. 표준 코드에는 각 산업별로 법적·제도적으로 부여하여 공통적으로 사용되는 코드뿐만 아니라 기관이나 기업 내부에서 정의하여 사용하는 코드가 대상이 된다.

5.3 전자카탈로그 데이터 활용관리

이 업무 프로세스는 전자카탈로그 데이터의 활용여부를 점검하거나 활용도를 높이기

위해 측정대상 데이터와 품질지표를 선정하여 품질을 측정하고 분석하여 품질을 충족시키지 못하는 경우 원인을 분석하여 담당자로서 하여금 조치하도록 하는 작업을 말한다. 이 프로세스를 통하여 애플리케이션에서 활용되지 않는 데이터를 점검하여 DB의 사용 환경을 개선하고 업무적 중요도가 높은 데이터에 대한 품질의 평가와 개선으로 데이터의 활용도를 높인다<그림 6>. 세부 관리 대상은 다음과 같다.

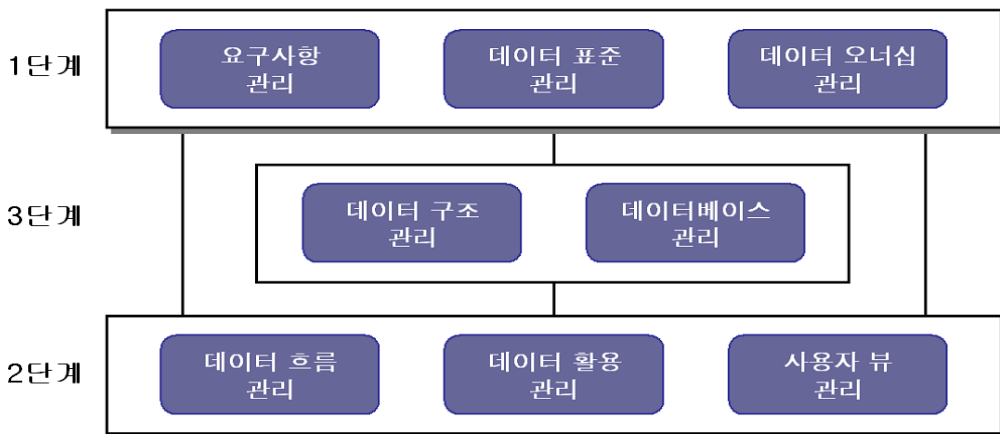
- ① 핵심 데이터 : 회사의 고객, 프로세스, 시장 환경, 재무정보 등에 직접적으로 영향을 미치는 중요성이 높은 데이터를 말한다. 핵심 데이터는 업무 프로세스상의 중요성, 재무적 관점에서 관리의 필요성, 최종적인 사용자의 활용성 등을 기준으로 도출하여 관리한다.
- ② 측정방법 : 데이터의 업무적인 규칙 및 특성(domain, validation 등)을 반영한 데이터 품질 측정 기준을 말한다. 업무적인 규칙과 특성이 정확하게 반영이 되도록 하며, 핵심 데이터별로 측정 방법을 관리한다.

6. 결론 및 향후 체계 확대 전략

6.1 프로세스를 고려한 전자카탈로그 데이터 품질 관리의 장점

데이터 품질을 고려할 때 대상이 되는 데이

터는 대부분 정적인 데이터를 다룬다. 거래 데이터가 동적인 데이터의 성격을 일부 가지고 있다. 데이터 품질관리 시스템 전체를 고려할 때는 데이터 자체와 함께 이 시스템 내부에 있는 데이터를 관리하는 프로세스도 함께 포함 되어야 완전한 시스템으로서의 역할을 수행한다. 프로세스라 함은 데이터 품질을 향상시키기 위한 일체의 행위를 가리킨다. 본 논문에서는 특별히 데이터 정책관리, 데이터 표준관리, 데이터 활용 관리로 구분한 프로세스를 제안하고 있다. 또한 프로세스의 주체로는 사용자, 총괄 데이터 책임자, 데이터 관리자를 주요 역할자로 제안하고 있다. 각 프로세스와 역할자는 데이터 품질 향상을 위한 지침에 따라 세부 프로세스별로 역할과 책임을 수행하게 되며 데이터 속성별로 데이터 품질을 점검하게 된다. 이와 같이 본 논문에서는 데이터와 데이터를 관리하는 프로세스가 결합된 프레임워크를 제시함으로써 데이터 품질관리 시스템의 균형을 도모하고 있다.



<그림 7> 데이터 품질관리 프로세스 확대 범위

6.2 전자카탈로그 데이터 품질관리 방법론으로 확대

본 방법론은 전자상거래 관점에서 전자카탈로그의 데이터 품질관리 프로세스를 정의하기 위한 것으로서, 데이터 관리정책 하의 데이터 표준관리 및 데이터 활용관리에 초점을 맞추었다. 좀 더 완벽한 데이터 품질관리를 위해서는 콘텐츠에 많은 영향을 미치는 프로세스뿐만 아니라 간접적으로 영향을 미치는 프로세스도 고려하는 것이 바람직하다. 예를 들면 데이터 품질관리 성숙모델에서는, <그림 7>에서 보는 바와 같이, 데이터 표준관리와 데이터 활용 관리뿐만 아니라 요구사항 관리, 데이터 구조 관리, 데이터 흐름 관리, 데이터베이스 관리, 데이터 오너십 관리, 사용자 뷰 관리도 고려하고 있다[3]. 이러한 관리 프로세스들을 모두 동시에 구현시키는 것은 쉽지 않다. 따라서 데이터 품질관리 프로세스 체계를 단계적으로 확대 구현하는 것이 바람직한데, 중요성에 따라 제 1단계에는 요구사항 관리, 데이터 오너십 관리를, 제 2단계에는 데이터 흐름 관리, 사용자 뷰 관리를, 제 3단계에는 데이터 구조 관리, 데이터베이스 관리를 구현할 것을 권장한다.

본 논문에서 다루고 있는 전자카탈로그 중심의 방법론은 일반적인 시스템에서 다루는 콘텐츠를 대상으로 확대하여도 기본적으로는 변화가 없다. 그러나 방법론을 적용해야 할 시스템에서 다루고 있는 데이터의 유형과 성격, 데이터의 속성 중에 특별히 강조가 필요한 영역, 데이터 계층과 거버넌스 계층의 역할 분담, 데이터 품질의 수명주기와 품질관리 프로세스의 관점 등이 보다 깊이 다루어져야 할 것이다.

참 고 문 헌

- [1] 마이크로소프트웨어, “체계적인 데이터 품질관리를 위한 대안을 찾아서”, 제1월호, 2005.
- [2] 한국데이터베이스진흥센터, 데이터 품질관리 지침(Ver 2.1), 2006.
- [3] 한국데이터베이스진흥센터, 데이터 품질관리 성숙모형(Ver 1.0), 2006.
- [4] 한국전자거래협회, 전자카탈로그 표준화 가이드라인, 2003.
- [5] ECIF 15 : 2002 전자카탈로그 상품분류 코드, 전자상거래표준화통합포럼, 2002.
- [6] ECIF 27 : 2002 e-비즈니스 용어, 전자상거래표준화통합포럼, 2002.
- [7] ECIF 67 : 2003 전자카탈로그 공통속성 기술규칙, 전자상거래표준화통합포럼, 2003.
- [8] ECIF 69 : 2003 전자카탈로그 상품식별표준(GTIN)산업공통 활용지침, 전자상거래표준화통합포럼, 2003.
- [9] ECIF 100 : 2004 글로벌 데이터 사전(GDD) 상품 데이터 모델 및 연계지침, 전자상거래표준화통합포럼, 2004.
- [10] ECIF 0 : 2005 개방형 전자카탈로그 기술 사전 명세, 전자상거래표준화통합포럼, 2005.
- [11] Bitterer, A. and Newman, D., Organizing for Data Quality, Gartner Research, Stamford, CT, 2007.
- [12] Cykana, P., Paul. A., and Stern, M., “DOD Guidelines on Data Quality Management,” Proceedings of the 1996 Conference on Information Quality, Cambridge, MA, 1996, pp. 154-171.

- [13] Dyché J. and Levy, E., Customer Data Integration, John Wiley and Sons, 2006.
- [14] IBM, IBM Data Governance Council Maturity Model, 2007.
- [15] ISO 8000-1 Information Quality (draft), 2008.
- [16] ISO/CD-TS 8000-110 Information quality-Part 110 : Master data quality : Syntax, semantic encoding, and conformance to customer requirements, 2008.
- [17] ISO/CD-TS 22745-1 Open Technical Dictionary-Part 1 : Overview, 2008.
- [18] Otto, B., Wende, K., and Schmidt, A, Osl, P., "Towards a framework for corporate data quality management," Proceedings of 12th International Conference on Information Quality, Cambridge, USA, 2007.
- [19] Berson, A. and Dubov, L., Master data management and customer data integration for a global enterprise, Mc Graw Hill Osborne, 2007.
- [20] Richardson, A., Data quality assessment guidelines, www.nrimns.nsw.gov.au, 2003.
- [21] Ryu, K. S., Park, J. S., and Park, J. H., "A data quality management maturity model," ETRI Journal, Vol. 28. No. 2, 2006.
- [22] Wang, R. Y., "A product Perspective on Total Data Quality Management," Communications of the ACM, Vol. 41, No. 2, 1998, pp. 61-65.
- [23] Wende, K., "A model for data governance-organizing accountabilities for data quality management," Proceedings of 12th International Conference on Information Quality, Cambridge, USA, 2007.
- [24] Wende, K. and Otto, B., "A Contingency Approach to Data Governance," Proceedings of 12th International Conference on Information Quality, Cambridge, USA, 2007.

저 자 소개



김선호
1992년~현재
1985~1989년
1979년
1989~1992년
1979~1984년
관심분야

(E-mail : shk@mju.ac.kr)
명지대학교 산업경영공학과 교수
Pennsylvania State University 산업공학과 석·박사
서울대학교 산업공학과 (학사)
한국기계연구원 자동화연구부 선임연구원
국방과학연구소 MIS실 연구원
Data Quality, 전자거래 표준화, BPM



이창수
1992년~현재
1994년
1989년
1987년
관심분야

(E-mail : cslee@nukw.ac.kr)
강릉원주대학교 산업정보경영공학과 교수
서울대학교 산업공학과 (박사)
서울대학교 산업공학과 (석사)
서울대학교 산업공학과 (학사)
데이터 품질, 마케팅정보시스템



이제현
2008년~현재
2008년
관심분야

(E-mail : want21st@hotmail.com)
명지대학교 산업경영공학과 (석사과정)
명지대학교 산업경영공학과 (학사)
데이터 품질, SCM