

데이터 품질관리 프레임워크와 비즈니스 시나리오

The Data Quality Management Framework and it's Business Scenario

이창수(Changsoo Lee)*, 김선호(Sunho Kim)**

초 록

e-비즈니스의 활성화로 기업과 조직에서 이해당사자 간의 데이터 교환이 활발해 짐에 따라, 신뢰성 있는 데이터의 확보 및 관리가 시급한 과제로 떠오르고 있다. 이러한 문제를 해결하기 위해, 본 논문은 데이터의 품질을 체계적으로 관리할 수 있는 프레임워크를 시나리오와 함께 제시한다. 데이터 품질 관리 프레임워크는 데이터 품질 모니터링, 데이터 품질 개선, 데이터 활용의 3단계로 구분되어 있으며 각 단계마다 3개씩, 총 9개의 프로세스로 구성되어 있다. 각 프로세스에는 필요성, 기능, 역할, 프로세스간의 관계가 명시되어 있다. 또한, 본 프레임워크를 현장에 직접 적용할 수 있도록, e-비즈니스에서 많이 사용되는 상품식별 및 분류 코드체계의 사례를 이용하여 업무 시나리오를 제시하였다.

ABSTRACT

As data exchange between business partners in e-business becomes more active, obtaining and managing reliable data is emerging as a pressing issue for corporations and organizations. For the resolution of data quality, this paper proposes a framework for data quality management with its scenario. The data quality management framework consists of three phases: data quality monitoring, data quality improvement and data application, each of which has three processes. In each process, necessity, functions, roles, and relationships among processes are specified. In order for users to directly apply the framework to the business field, a business scenario is given with examples of product identification and classification code systems widely used in e-business.

키워드 : 데이터품질관리, 프레임워크, 시나리오, e-비즈니스

Data Quality Management, Framework, Scenario, E-Business

* 강릉원주대학교 산업정보경영공학과 교수

** 교신저자, 명지대학교 산업경영공학과 교수

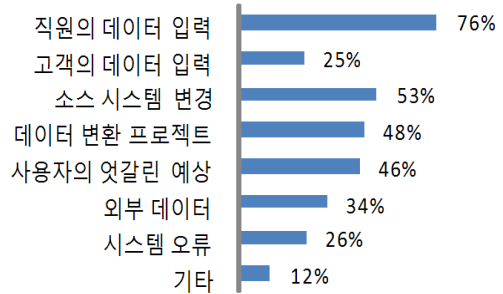
2010년 09월 27일 접수, 2010년 10월 17일 심사완료 후 2010년 11월 05일 게재확정.

1. 서론

저장장치의 기술 발달에 따른 가격의 합리화로 대용량 데이터 처리가 일반화되면서 데이터의 저장비용이 낮아지고 멀티미디어 데이터와 실시간 데이터의 증가로 저장된 데이터의 양이 기하급수적으로 늘어가고 있다. 또한 인터넷을 중심으로 정보 검색이 일상화되면서 다양한 데이터가 지속적으로 생성되고 있다[34]. 이러한 데이터의 양적 증가는 상대적으로 믿을만한 데이터에 대한 질적 욕구를 증대시키고 있다.

빈약한 품질의 데이터의 원인은 다양한 곳에서 발생하지만 주요 원인은 데이터 입력 과정에서 생긴다. TWDI에서 2002년 수행한 설문문에 따르면 설문 응답 기업은 데이터 결합의 가장 흔한 원인으로 직원의 데이터 입력 오류를 꼽았다. 철자 입력 오류, 숫자의 자릿수 착각, 코드 사용 실수, 엉뚱한 필드에 데이터 입력, 이름이나 약어의 인식 불가 등이 그 예이다. 이러한 데이터 품질 문제의 원인은 기업의 비즈니스가 웹 환경으로 옮겨가고 고객과 공급업자가 스스로 데이터를 운영 시스템에 입력하는 환경에서는 더욱 증가하게 된다<그림 1>.

기업들은 믿을만한 데이터를 확보하지 못해서 생기는 여러 가지 어려운 문제를 해결하려고 노력하고 있다. TWDI가 2006년 인터넷으로 실시한 설문에서 응답기업 741곳 중 83%가 품질이 빈약한 마스터 데이터로 인해 문제를 경험했다고 응답하였다. 이는 사용자의 대다수가 마스터 데이터 관리 문제가 현실적으로 상당히 심각하다는 것을 증명하고 있다[24].



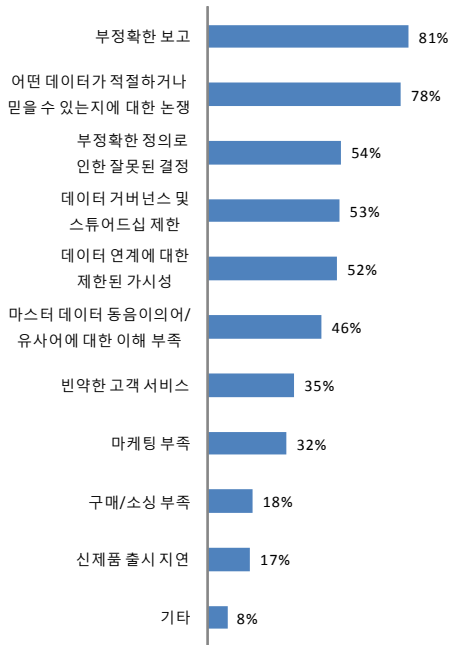
<그림 1> 데이터 품질 문제의 원인

빈약한 품질의 마스터 데이터는 데이터 웨어하우스와 비즈니스 인텔리전스 내부에 대부분 문제를 야기하지만 외부에도 발생한다. 가장 큰 문제들은 데이터 웨어하우징과 직접적 관련이 있는데 여기에는 부정확한 보고(81%), 어떤 데이터가 적절하거나 믿을 수 있는지에 대한 논란(78%), 부정확한 정의로 인한 잘못된 의사결정(54%) 등이 포함되어 있다<그림 2>.

가트너 조사에 따르면 데이터 관리를 위한 Hype Cycle은, 기술성숙 및 개발유인 측면에서 기술등장 단계, 도입기대 단계, 현실화단계, 개발도입 단계, 생산적용 단계의 5단계로 구분하고 있다. 또한, 현실적 기술적용 기대 연수를 2년 이내, 2~5년 이내, 5~10년 이내, 10년 초과, 적용전 퇴장으로 나누고 있다.

Hype Cycle은 가트너그룹이 IT 분야 기술의 트렌드와 비즈니스에의 영향력, 수용 상황, 미래에 대한 예측을 가늠하기 위해 개발한 과대광고곡선(hype cycle)으로서 가트너그룹에서 시간에 따라서 기술은 성숙한다는 점(maturity)과 이에 따라 대중들이나 언론에 회자되는 정도(visibility)가 일정한 규칙에 따라서 변한다는 사실에 근거해 기술에 대한 성숙도와 시장에서의 인식 상태를 5개 단계

로 구분해 보여주는 그래프이다.



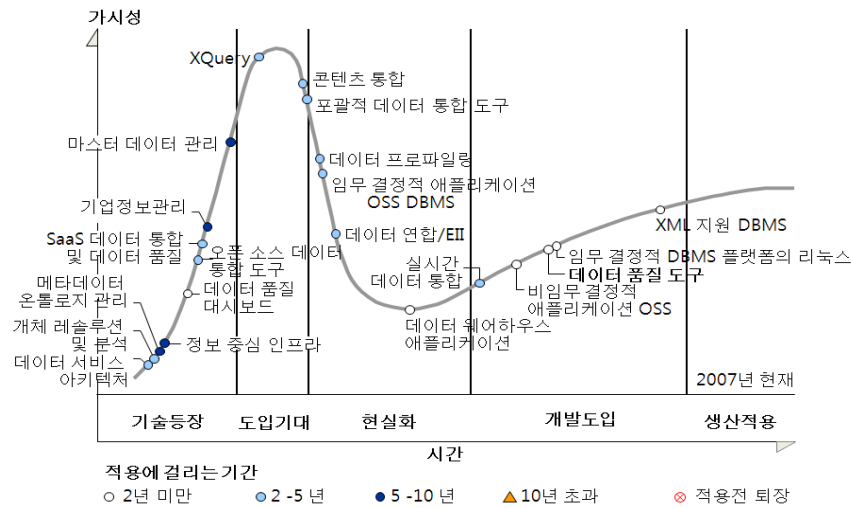
〈그림 2〉 데이터 품질 문제의 유형

데이터 품질 도구는 생산적용 단계(Plateau of Productivity) 직전의 개발도입 단계(Slope of Enlightenment)에 위치하고 있는 것으로 나타났다<그림 3>. 운영 통계 및 데이터 품질 통계에 대한 모니터링을 용이하게 하는 데이터 품질 대시보드(Data Quality Dashboard)와 마스터 데이터 관리는 기술 등장 단계에 위치하고 있다.

신뢰할만한 데이터는 전자거래의 기본이다. 전자거래에서 발생하는 저장 데이터와 거래 데이터의 품질 확보를 통해 사용자에게 안전한 거래에 대한 신뢰를 보증하는 것이 필요하다.

본 논문에서는 데이터 품질관리 프레임워크를 정의하며 이 프레임워크를 실제 적용할 수 있는 시나리오를 제공한다.

제 2장에서는 데이터 품질관리 프레임워크와 관련된 기존 연구를 살펴보았다. 데이터 품질관리 프레임워크는 ISO 9000형태를 지향



〈그림 3〉 Hype Cycle for Data Management, 가트너(2007)

하는 경우가 많았으며 산업별로 데이터 품질에 대한 요구가 많음을 알 수 있다.

제 3장에서는 데이터 품질관리 프레임워크를 제시한다. 프레임워크는 9개의 프로세스로 구성되며 각 프로세스의 필요성, 역할, 기능을 상술하였으며 각 프로세스간의 관계를 규명하였다. 제 4장에서는 제 3장에서 제시한 프레임워크에 맞추어 각 프로세스 별로 실제 적용 가능한 시나리오를 구성하였다.

2. 연구 동향

데이터 품질 관리를 효과적으로 수행하기 위해서는 체계적인 관리 절차를 마련하는 것이 바람직하다. 품질관리의 일반적 형식은 ISO 9000의 틀을 많이 따르고 있으며 여기서는 고객의 요구사항을 받아 측정/분석/개선, 관리책임, 자원관리, 제품 구현 등의 일련의 지속적 품질관리 개선활동을 통해 고객 만족을 달성하는 기본 흐름을 유지한다[14].

Batini and Scannapieco(2006)는 3개의 핵심 단계인 상태 재구성, 평가, 개선으로 구성된 프레임워크를 제시하고 있다. 이 프레임워크는 비즈니스 프로세스와 조직 서비스를 강조함으로써 비용이 많이 드는 부분을 집중 관리하려는 의도이다[2].

Otto(2007)는 데이터 품질의 이해당사자를 4가지 역할로 구분하여 Executive Sponsor, Chief Steward, Business Data Steward, Technical Data Steward를 다음 <표 1>과 같이 정의하고 있다[21].

Otto et al.(2007)은 기업 데이터 품질 관리 프레임워크를 제안하고 있다[22]. 이 프레임

워크의 구조는 비즈니스 엔지니어링(전략, 조직, 정보 시스템), 거버넌스, 실행의 세 계층으로 구성하고 있으며 거버넌스는 기업 데이터품질 전략 개발, 기업 데이터 품질 조직 설계, 기업 데이터품질 정보시스템 아키텍처 설계로 구성되어 있다. 실행은 기업 데이터 품질 전략과 의사소통 및 통제, 기업 데이터 품질 프로세스를 실행하고 모니터링, 기업 데이터품질 정보시스템 아키텍처를 운영하고 유지로 구성되어 있다.

<표 1> 데이터 품질의 이해당사자(21)

역할	내용
Executive Sponsor	데이터 품질관리에 대한 지원, 감독, 후원, 전략 방향, 자금지원을 제공
Chief Steward	위원회 결정을 실행에 옮기며 표준 채택을 강화. 데이터 품질 기준 및 목표 설정을 지원
Business Data Steward	비즈니스 관점에서 전자적 데이터 품질 표준과 정책의 세부사항을 결정
Technical Data Steward	표준화된 데이터 요소 정의, 형식, 프로파일 소스 시스템 세부사항, 시스템 간의 데이터 흐름을 제공

정보 제품 관점에서 데이터 품질 관리 모델에 접근하고 있다. 이는 데이터 및 데이터 처리를 정보 시스템의 데이터 품질을 시스템적으로 평가하고 관리하도록 고려하는 접근법이다[30]. 이 접근법을 이용하면 품질관리 기법을 데이터 품질 관리를 위해 물리적 제품 생산에 성공적으로 적용가능하다[31].

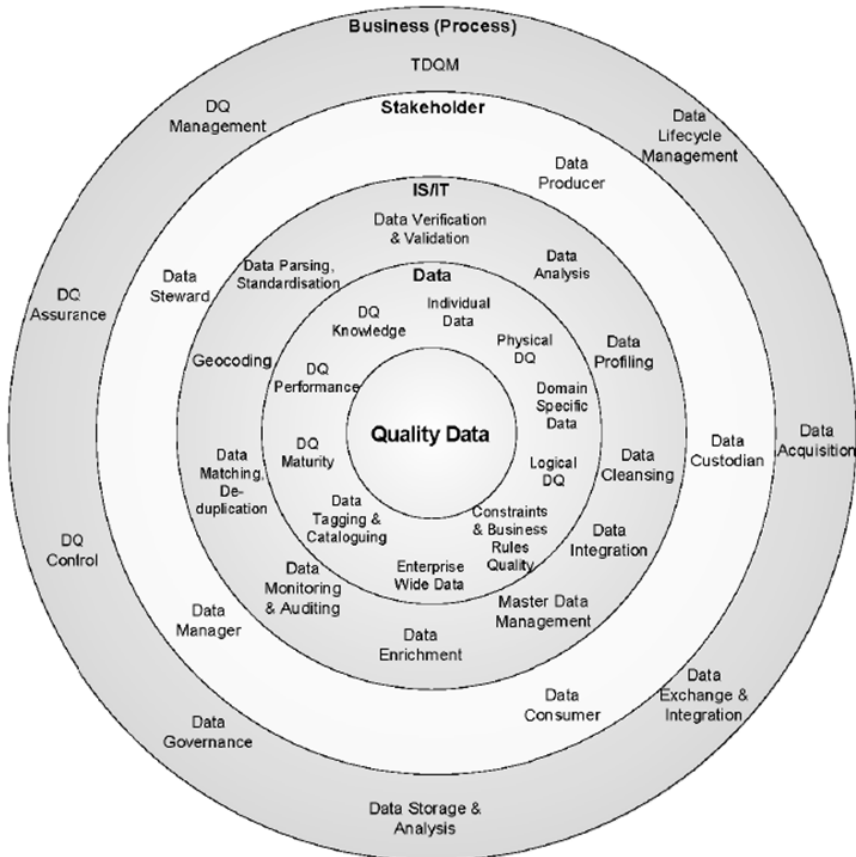
Shankaranarayanan and Cai(2005)는 B2B 환경에 대한 데이터 품질관리 표준은 다음 세 가지 요구사항을 만족해야 한다고 언급한

다: 조직은 내부적으로 데이터 품질을 관리한다(말하는 법을 알아야 한다). 조직은 데이터 품질을 평가할 체계를 갖추어야 한다(동일한 언어로 말해야 한다). B2B 관계에서 이해 당사자들은 교환 데이터와 관련된 메타데이터를 공유한다(서로 대화하며 이해해야 한다). 이 논문에서는 ISO 9000과 유사한 DQ 9000을 제안하고 있으며 여기에는 프로세스 정의, 레코드 유지, 관리 책임, 데이터 품질 전달 부서를 주요 요소로 포함하고 있다[27].

산업별로도 데이터 품질 지침이나 프레임워크를 만들어 운용하고 있다 : 의료[3, 6, 15],

유통[10], 시멘트 웹[8, 9], 자산관리[17], 토목[18]. 지방정부[20], 천문학[23], e비즈니스[25], 센서 네트워크[26], 정보시스템[28], 교통[29], 국방[4, 5].

이중에서 의료분야인 DEFRA는 유용한 템플릿과 다양한 데이터 품질 측정치에 대한 정의를 제공하고 있으며[6], 유통 분야의 GS1의 프레임워크는 데이터 품질 관리 시스템에 대한 요구사항, 제품 검사 절차, 기업용 자기평가 절차로 구성되어 있다[10]. 천문학 분야의 Radziwill and DuPlain(2006)은 데이터 품질 정책 설정, 데이터 생성 프로세스의 의존



〈그림 4〉 자산관리를 위한 데이터 품질 프레임워크(바깥 원부터 차례로 조직, 인력, 기술, 데이터)[17]

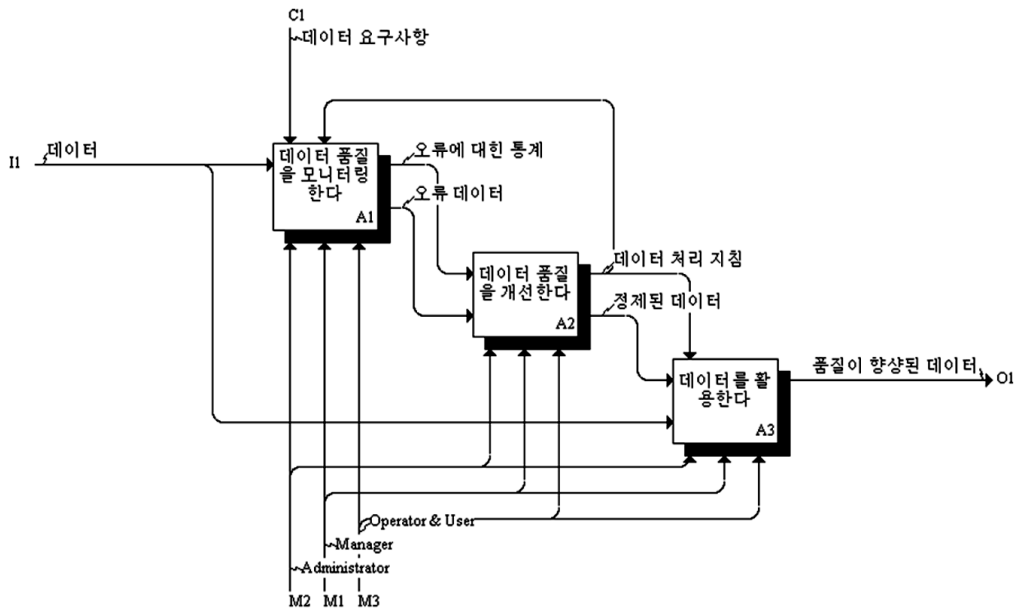
성 이해, 품질 정보에 대한 데이터 스토어 구현, 중요 인프라 및 통제 시스템 모니터 지점의 지속적 모니터링 제공, 데이터 생성 중 데이터 품질 평가에 대한 룰셋 적용, 도출된 데이터 품질을 평가하기 위해 알고리즘에 기반한 추가 룰셋 적용, 데이터와 생성 프로세스 간의 순환을 종료 등의 절차를 제공하고 있다[23]. Lin et al.(2007)은 자산관리의 데이터 품질을 보장하기 위한 프레임워크로 조직, 인력, 기술, 데이터를 포괄적으로 운용하는 안을 <그림 4>와 같이 제시하고 있다[17].

3. 데이터 품질관리 프레임워크

본 프레임워크는 한국데이터베이스진흥센터의 데이터 품질관리 지침[35]을 기반으로 하였으며 전자카탈로그와 마스터 데이터를

대상으로 품질관리 프레임워크를 제시한 기존 연구[32, 33]에 각 프로세스의 필요성, 기능 및 역할을 추가하였다.

본 프레임워크는 데이터 품질 모니터링, 데이터 품질 개선, 데이터 활용의 3단계로 구분하였다. 데이터 이해당사자로부터 데이터 요구사항을 받아 데이터 품질 계획수립, 기준 설정, 측정 등의 모니터링 활동을 통해 오류 데이터를 검출한다. 이어 오류 데이터에 대한 원인 분석, 수정, 데이터 권한 및 흐름 관리 활동을 수행하는 데이터 품질 개선단계를 통해 데이터가 정제되며 데이터 처리에 대한 지침이 만들어 진다. 마지막으로 3단계 데이터 활용에서는 전사적 데이터 아키텍처 관리, 데이터 설계, 데이터 처리 활동을 통해 데이터의 품질을 향상시킨다. <그림 5>에 단계별 흐름을 나타내었으며 단계별 프로세스는 <표 2>에 정리하였다.



<그림 5> 데이터 품질관리 프레임워크의 단계별 흐름

〈표 2〉 데이터 품질관리 프레임워크 단계별 프로세스

단계	프로세스
1. 데이터 품질 모니터링	1. 데이터 품질 계획 2. 데이터 품질 기준 설정 3. 데이터 품질 측정
2. 데이터 품질 개선	4. 데이터 권한 및 흐름 관리 5. 데이터 오류 원인 분석 6. 데이터 오류 수정
3. 데이터 활용	7. 전사적 데이터 아키텍처 관리 8. 데이터 설계 9. 데이터 처리

단계 1: 데이터 품질 모니터링

3.1 데이터 품질 계획

조직 내외에서 데이터 품질에 대한 요구사항이 있거나 품질에 대한 인식이 다를 경우 이러한 요구사항이나 상이한 품질 인식을 고려하여 통일된 품질 목표를 설정할 필요가 있다. 또한 수립된 데이터 품질 목표를 달성하기 위하여 구체적인 실행 방안을 설정할 필요가 있다.

3.1.1 기능 측면

- 통일된 목표 설정 및 관리 : 조직 내외에 존재하는 품질 요구사항과 다양한 품질 인식을 수렴하여 한 방향으로 결집된 품질 목표를 설정한다. 또한 이 품질 목표가 달성되도록 지속적으로 목표 관리를 수행한다.
- 품질관리 대상 파악 및 계획수립: 설정한 목표를 달성하기 위해 필요한 구체

적인 품질관리 대상을 식별하고 이를 개선하기 위한 실행계획을 수립한다. 구체적인 품질계획에는 과제도출, 일정계획, 자원 확보, 방법수립 등이 포함된다.

3.1.2 역할 측면

- 품질 관리 대상을 통제할 수 있는 위치. 데이터 품질 계획은 품질 관리 대상과 관련 자원을 통제할 수 있는 위치에서 수행되어야 한다.
- 최고경영층과의 소통을 통해 추진력을 확보할 수 있는 위치 : 이 위치에서는 최고경영층에 품질 목표 및 품질관리 활동의 진행 상황을 보고하여 조직전체가 데이터 품질 목표에 관심을 갖고 일관되게 계획을 진행하도록 해야 한다.

3.1.3 프로세스 간의 연계성

- ‘데이터 품질계획’과 ‘전사적 데이터 아키텍처 관리’ : 데이터 품질 계획의 결과가 전사적 데이터 아키텍처에 반영될 수 있다.
- ‘데이터 품질계획’과 ‘데이터 권한 및 흐름 관리’ : 데이터 품질 계획의 결과가 데이터 권한 및 흐름 관리에 반영될 수 있다.
- ‘데이터 품질계획’과 ‘데이터 품질 기준 설정’ : 데이터 품질 계획은 데이터 품질 기준 설정에 수행범위, 수행방법, 수행일정 등을 제공한다.

3.2 데이터 품질 기준 설정

데이터 품질 계획을 구체적으로 수행하기

위해서는 사용자들의 요구사항을 반영한 품질 측정 및 평가 기준을 제공할 필요가 있다. 기존에 간헐적으로 수행되고 여러 사람이 분산해서 가지고 있는 품질 기준을 체계적으로 수집 및 정리하여 정기적으로 정형화된 형태로 설정할 필요가 있다. 이를 통해 품질 이슈가 공식적으로 제기되고 필요한 자원이 배분되어 품질 측정 활동을 구체적으로 수행할 수 있다. 이 프로세스에서는 ISO 8000-1xx 시리즈가 품질측정 기준으로 적용될 수 있다.

3.2.1 기능 측면

- 품질 기준 설정. 품질 기준은 데이터 사용자들의 요구사항을 수렴하여 관련자들이 공감할 수 있는 실질적인 기준을 정의한다. 데이터 품질 기준의 충족을 확인하기 위하여 상세한 측정 대상 데이터, 측정 기준 및 방법을 구체적으로 설정한다.
- 기준별 품질 수준 점검. 데이터 품질 관리 대상 별로 데이터 품질 수준을 충족하였는지 점검한다. 필요에 따라 점검 결과를 데이터 품질 기준에 반영한다.

3.2.2 역할 측면

데이터 품질 기준 설정은 데이터 사용자의 품질 요구사항이 반영될 수 있는 위치에서 수행되어야 한다. 또한 해당 데이터의 관련자들 간에 충분한 협의가 이루어져서 공감대가 형성될 수 있도록 진행되어야 한다.

3.2.3 프로세스 간의 연계성

- ‘데이터 품질 기준설정’과 ‘데이터 품질계

획’ : ‘데이터 품질 기준 설정’은 ‘데이터 품질 계획’의 하부 작업으로 수행된다.

- ‘데이터 품질 기준설정’과 ‘데이터 품질 측정’ : 데이터 품질 기준을 바탕으로 데이터 품질 측정이 이루어진다.

3.3 데이터 품질 측정

데이터 처리과정에서 사용자의 오류, 응용 프로그램 상의 오류, 프로세스 상의 오류 등으로 인해 발견되지 않는 경우가 발생한다. 따라서 이것을 해결하기 위해서는 체계적인 품질 측정이 필요하다. 또한 특정 데이터는 시간이 경과함에 따라 의미가 없어지는 경우가 있으므로 이를 고려하여 데이터 품질 측정 주기를 설정해야 한다. 데이터 품질은 데이터 처리가 이루어진 후 빠른 시일에 진행하는 것이 바람직하지만 업무적인 특성에 따라 품질 일정을 조정할 수 있다.

3.3.1 기능 측면

- 데이터 품질 측정. 데이터 품질 대상, 기준 및 방법에 의해서 설정된 측정항목을 측정도구나 사람을 통해 측정한다. 단순 반복되는 데이터의 경우는 측정도구를 활용하는 것이 가능하지만, 복잡한 데이터의 경우는 측정자의 판단을 통해 오류 여부를 진단한다.
- 데이터 품질 측정 결과에 대한 통계 처리. 데이터 오류 원인분석이 가능하도록 측정 결과에 대한 저장 및 통계 처리를 수행한다.

3.3.2 역할 측면

데이터 품질 측정은 데이터에 대한 접근이

가능한 위치에서 수행된다.

3.3.3 프로세스 간의 연계성

- ‘데이터 품질 측정’과 ‘데이터 품질 기준 설정’ : ‘데이터 품질 측정’결과에 따라 ‘데이터 품질 기준’을 조정할 수 있다.
- ‘데이터 품질 측정’과 ‘데이터 오류 원인 분석’ : ‘데이터 품질 측정’결과 발견된 오류데이터와 통계자료를 ‘데이터 오류 원인분석’으로 전달한다.
- ‘데이터 품질측정’과 ‘데이터 오류 수정’ : ‘데이터 품질 측정’의 결과로 발견된 오류 데이터의 수정을 위하여 ‘데이터 오류 수정’으로 전달한다.

단계 2 : 데이터 품질 개선

3.4 데이터 권한 및 흐름 관리

데이터는 독자적으로 존재하는 것이 아니고 여러 부문과 연계되어 있다. 그러므로 데이터가 어떤 방법으로 활용되고 있는지를 파악하는 것은 데이터 품질 관리에서 매우 중요하다. 데이터와 사용자 간의 연계성이 파악되면, 사용자의 품질 요구사항을 명확하게 이해할 수 있고, 또한 데이터 품질 오류에 대한 책임 있는 처리가 이루어질 수 있다. 또한 데이터를 실질적으로 조작하는 어플리케이션 혹은 데이터 간의 흐름을 파악한다면, 데이터 오류로 인한 영향을 분석하고 관련 데이터를 추적하여 완벽하게 오류를 수정할 수 있다. 그 데이터를 다루는 사람이 여러 명일 경우 그 데이터의 품질을 책임질 수 있는 담당자

가 필요하다.

3.4.1 기능 측면

- 데이터 흐름 파악 : 여러 응용 시스템에 산재한 데이터가 동일한 속성 및 값을 갖도록 데이터간의 연계성을 파악하는 활동이다. 한 곳에서 변경이 발생하거나 데이터 오류가 발생하면 다른 응용 시스템에 어떤 영향을 미치는 지 파악한다.
- 권한(오너십 or stewardship, not authority) 부여 : 산재한 데이터를 통합 관리할 책임자를 지정한다. 데이터 값이 부분적으로 변경될 경우 해당되는 모든 응용 시스템에 반영되도록 한다.

3.4.2 역할 측면

- 동일한 데이터를 사용하는 현업을 통제하고 권한을 줄 수 있는 위치에 있어야 한다.
- 새로운 시스템이 도입되거나 데이터 흐름이 변경되면 조정하거나 통제할 수 있는 위치에 있어야 한다.

3.4.3 프로세스 간의 연계성

- ‘데이터 권한 및 흐름 관리’와 ‘전사적 데이터 아키텍처 관리’ : 데이터 권한 및 흐름관리가 이루어지기 위해서는 전사적 데이터 아키텍처 관리가 선행되어야 한다.
- ‘데이터 권한 및 흐름 관리’와 ‘데이터 품질계획’ : 데이터 권한 및 흐름 관리는 데이터 품질 계획에 맞춰서 권한과

흐름을 조율한다.

- ‘데이터 권한 및 흐름 관리’와 ‘데이터 오류 원인 분석’ : 데이터 권한 및 흐름 관리는 데이터 오류 원인 분석에 활용된다.

3.5 데이터 오류 원인 분석

발견된 오류 데이터의 값만 수정할 경우, 언제나 동일한 오류가 재발할 가능성이 있다. 그러므로 데이터 오류가 발생하는 근본적인 원인을 파악하고 필요한 조치를 취해야 한다. 데이터 오류에는 다양한 원인이 존재하므로, 단기간에 처리할 수 있는 것도 있지만 장기간의 계획에 의하여 단계적으로 제거해야 할 경우도 발생한다.

3.5.1 기능 측면

- 데이터 오류 원인 분석 및 제거. 데이터 오류를 효과적으로 제거하기 위해서는 데이터 구조나 데이터 표준, 데이터 흐름 등에서 근본적인 원인을 파악하여 동일한 오류가 재발하지 않도록 조치를 취한다. 과거 오류 원인과 해결 내역을 추적하면 오류 처리를 효과적으로 수행할 수 있다.
- 데이터 오류 예방 활동 : 동일한 유형의 오류가 다른 곳에서 발생하지 않도록 예방 활동을 수행한다. 이 활동을 위해서는 가능한 오류 원인(데이터 자체, 조직, 데이터 흐름 등)을 총체적으로 파악하여야 한다. 데이터 권한(오너십)이나 데이터 흐름 등을 조정할 필요가 있을 경우 담당자와 협의를 통하여 해결한

다. 또한, 재발 방지를 더욱 효과적으로 수행하기 위하여 오류를 원인별로 분류하여 재발 방지 지침을 작성하고, 관련자 교육을 수행하여야 한다.

3.5.2 역할 측면

- 데이터 오류의 원인 분석은 전사적인 입장에서 수행되어야 하며, 관련 데이터, 관련 시스템, 관련 사용자 등에 대한 추적이 가능한 위치에서 수행되어야 한다.
- 또한 데이터 오류 원인에 대한 조치를 수행할 수 있는 위치에서 진행되어야 한다.

3.4.3 프로세스 간의 연계성

- ‘데이터 오류 원인 분석’과 ‘데이터 권한 및 흐름 관리’ : ‘데이터 오류 원인 분석’을 위해서는 ‘데이터 권한 및 흐름관리’가 확보되어 있어야 한다. 왜냐하면, 원인 분석 결과에 따라 데이터 권한 및 흐름에 대한 조정을 요청할 수 있기 때문이다.
- ‘데이터 오류 원인 분석’과 ‘데이터 오류 수정’ : ‘데이터 오류 원인 분석’ 결과에 따라 ‘데이터 오류 수정’에 관련 데이터 값의 수정을 요구하게 된다.
- ‘데이터 오류 원인 분석’과 ‘데이터 품질 기준설정’ : ‘데이터 오류 원인 분석’ 결과를 ‘데이터 품질 기준’에 반영한다.
- ‘데이터 오류 원인 분석’과 ‘데이터 설계’ : ‘데이터 오류 원인 분석’ 결과에 따라 ‘데이터 설계’에서 데이터 구조 변경이 필요할 수 있다.

3.6 데이터 오류 수정

이 프로세스에서는 품질 측정 결과와 오류 원인 분석 결과에 따라 수정을 하게 된다. 이때 데이터 오류는 발견된 부분에 국한하여 수정하기 보다는 동일한 데이터나 연관되는 데이터가 다른 곳에도 중복되어 존재하므로 이를 모두 찾아서 수정할 필요가 있다. 또한 서로 연관되는 데이터를 파악하여 전체적인 일관성(consistency)이 유지될 수 있도록 수정할 필요가 있다.

3.6.1 기능 측면

- 데이터 오류 수정 : 품질 측정 결과 또는 오류 원인 분석 결과에 따라 발견된 오류를 수정하고 연관성 있는 데이터를 추적하여 관련 오류를 모두 수정한다.
- 수정 내역의 공유 : 오류 데이터가 수정 되면 관련자에게 수정 전과 후의 데이터 내용이 명확하게 알려야 하며 이를 통해 데이터 활용에 혼선이 없어야 한다.

3.6.2 역할 측면

데이터 오류 수정은 데이터 수정 권한이 있는 관련자와의 협의가 가능한 위치에서 수행되어야 한다. 또한 데이터 오류 수정에 대한 진행 사항을 관리할 수 있어야 한다. 특히 데이터 오류 및 수정 내역을 관련자들이 공유할 수 있는 환경을 제공해야 한다.

3.6.3 프로세스 간의 연계성

- ‘데이터오류수정’과 ‘데이터오류 원인분석’ : ‘데이터 오류 수정’의 결과가 축적

되어 ‘데이터 오류 원인 분석’에서 활용된다.

- ‘데이터오류수정’과 ‘데이터 처리’ : ‘데이터 오류 수정’의 결과가 ‘데이터 처리’에 대한 지침으로 반영될 수 있다.

단계 3 : 데이터 활용

3.7 전사적 데이터 아키텍처 관리

데이터가 조직 내에 분산되어 있으므로 이에 대한 체계적인 관리를 수행하지 않으면 데이터 품질을 확보할 수 없다. 전사적 데이터 아키텍처 관리는 조직 전체에서 활용하는 데이터는 어떤 것이 있으며 대내외적인 업무 수행에 적합한 품질을 확보하기 위해서는 어떤 형태를 갖추어야 하는지에 대한 정의가 포함된다. 또한 이들 데이터가 각 시스템에 어떤 형태로 분산되어 있는지를 파악하여 중복된 데이터 간에 불일치가 발생하지 않아야 한다.

3.7.1 기능 측면

데이터 활용에 속하며 전사적 개념 데이터 모델(conceptual data model)과 전사적 데이터 표준 관리가 두 가지 주요 활동이며 세부 내용은 다음과 같다.

- 전사적 데이터 개념 모델 : 전사적 데이터를 구성하는 것 중 공유되거나 관리가 필요한 항목들만을 선택하여 나타낸 모델로, 전사 데이터 모델 중에 대표성을 지닌 일부분을 말한다.
- 전사적 데이터 표준 : 모든 전사적 데이터 구조에서 지켜져야 할 규정으로 모든

데이터에 동일하게 적용되어야 한다.

3.7.2 역할 측면

관리자 역할에 속하며 다음과 같은 두 가지 중요한 역할을 수행한다.

- 전사적 조율 역할 : 여러 사람의 의견을 하나로 수렴하는 역할이 가장 중요하다. 또한 수렴된 의견을 모두에게(IT, 현업 모두를 포함하여) 강제할 수 있는 위치에 있어야 한다. 데이터 품질은 현업에 의해 좌우되는 요소가 많으므로 IT 분야 뿐만 아니라 현업 당사자를 조율하고 통제할 수 있는 권한이 반드시 필요하다.
- 전사적 공유 및 유지 역할 : 데이터 개념모델과 데이터 표준을 품질을 확보할 수 있는 방향으로 공유시키고 데이터 구조 변경시 개념모델과의 응용 시스템 간의 매핑을 유지하는 등 지속적인 관리가 이루어지는 역할을 수행한다.

3.7.3 프로세스 간의 연계성

- ‘전사적 데이터 아키텍처 관리’와 ‘데이터 품질계획’ : 전사적 데이터 아키텍처는 전사의 데이터 구성을 보여주므로 이 구성을 바탕으로 품질계획을 수립할 수 있다. 수립된 품질계획 중 중요사항을 전사 데이터 개념 모델의 관리 대상으로 포함시키고, 전사적 데이터 표준들도 관리해야 한다.
- ‘전사적 데이터 아키텍처 관리’와 ‘데이터 권한 및 흐름 관리’ : 전사적 데이터 아키텍처는 전사 데이터 구성을 보여주므로 이 구성을 바탕으로 오너십을 설

정할 수 있다. 실제적 데이터 오류 해결 과정에서 데이터 권한/흐름이 변경되면 이를 전사 데이터 아키텍처에 반영해야 한다.

- ‘전사적 데이터 아키텍처 관리’와 ‘데이터 설계’ : 전사적 데이터 아키텍처는 데이터 설계를 위한 데이터 표준과 개념 모델 등의 기준을 제공한다.

3.8 데이터 설계

품질오류는 사용자에게 의한 품질 오류와 구조적 품질 오류가 있는데, 사용자에게 의한 품질 오류는 시스템적으로 해결하는데 한계가 있으나 구조적 품질 오류는 시스템적으로 데이터 설계에서 해결해 줄 수 있다. 반면 구조적 품질 오류가 발생하면 사용과정에서 수정하기 쉽지 않으므로 초기 설계과정에서 품질에 대한 충분한 고려가 이루어져야 한다. 특히 특정 응용 시스템 입장에서만 데이터 설계를 하게 되면 전사적으로 공유되는 데이터에 대해서는 품질이 저하될 수 있다. 그러므로 전사적 관점에서 타 응용 시스템과의 데이터 연관성을 충분히 고려하여야 한다.

3.8.1 기능 측면

- 품질을 고려한 데이터 설계 : 데이터 설계 과정에서 데이터 품질을 확보하기 위하여 요구되는 형식이나 범위 등을 정의한다. 또한 데이터 설계는 사용자의 요구사항(user requirements)을 충분히 반영하여야 한다.
- 전사적 데이터 아키텍처와의 연계 : 데이터 설계를 바탕으로 데이터베이스를

구현하고, 데이터 베이스의 구조변경 요건이 발생하면, 전사적 데이터 아키텍처와의 연관관계를 유지하도록 데이터 설계 및 변경을 수행한다.

3.8.2 역할 측면

- 내부적 역할 : 응용 시스템 사용자 및 관련자와의 충분한 협의를 통하여 품질 요소를 반영할 수 있는 위치에 있어야 한다.
- 외부적 역할 : 타 응용 시스템과 데이터 연계를 협의할 수 있는 위치에 있어야 한다.

3.8.3 프로세스 간의 연계성

- ‘데이터 설계’와 ‘전사적 데이터 아키텍처 관리’ : ‘데이터 설계’의 결과는 ‘전사적 데이터 아키텍처 관리’에 반영된다.
- ‘데이터 설계’와 ‘데이터 처리’ : ‘데이터 설계’를 바탕으로 ‘데이터 처리’가 이루어진다.
- ‘데이터 설계’와 ‘데이터 품질 기준설정’ : ‘데이터 설계’를 근간으로 ‘데이터 품질 기준’이 정의된다.

3.9 데이터 처리

데이터 처리의 주요 실행자는 데이터 사용자(data end user)를 의미한다. 데이터 사용자의 부주의 및 이해부족이 데이터 품질에 직접적 영향을 미치는 주요 요인이며 이로 인하여 데이터 품질에 오류가 많이 발생한다.

따라서 데이터 사용자에게 데이터 품질관리 관점에서 활용 지침을 제공할 필요가 있으며 만약 잘못된 값을 입력하면 입력 당사자를 파악할 수 있어야 한다. 또한 시스템에 의한 데이터 처리 내역도 추적할 수 있도록 관리되어야 한다.

3.9.1 기능 측면

- 데이터 처리 : 데이터 품질 지침을 준수하여 데이터의 생성, 조회, 변경, 전달, 삭제 등을 수행하는 활동이다. 응용프로그램에 의해 데이터가 처리될 경우에도 데이터 품질 지침이 준수되도록 해야 하며, 데이터 처리 부분이 명확하게 인지되어야 한다.
- 데이터 로깅 : 데이터 처리 내역에 대한 추적이 가능하도록 데이터 사용자, 사용시간, 사용 내역 등이 기록되어 저장되어야 한다.

3.9.2 역할 측면

데이터 처리는 다양한 업무와 계층에서 많은 인원에 의해서 이루어지므로 체계적인 데이터 품질 관리가 쉽지 않다. 그러므로 데이터 처리 및 품질관리가 사용자별로 특화되어 이루어져야 한다. 또한 품질이 중요한 데이터에 대해서는 복수의 관련자에 의한 중복적인 데이터 품질 확인 작업이 필요하다.

3.9.3 프로세스 간의 연계성

- ‘데이터 처리’와 ‘데이터 설계’ : 데이터 처리 과정에서 추가적인 품질이슈가 발생하면 데이터 설계에 반영을 요청한다.

- ‘데이터 처리’와 ‘데이터 품질측정’ : ‘데이터 처리’의 결과에 대하여 ‘데이터 품질 측정’이 이루어진다.

4. 프레임워크 사용 시나리오

제 4장은 제 3장에서 설명하고 있는 데이터 품질관리 프레임워크의 실제 적용을 위한 업무 시나리오를 설정하였다. 여기서는 상품 식별코드와 상품분류 코드[36]를 중심으로 사례를 제시한다.

- 업무 환경 및 상황
조직의 정보화가 확대되면서 정보시스템이 조직 내에 다수 구축되고 있으며 데이터도 다수의 정보시스템에 분산되어 활용되고 있다. 이러한 데이터 중 일부는 다수의 내부 정보시스템에서 공통으로 사용되고, 일부는 외부 시스템과 공통으로 사용된다. 따라서 데이터에 품질 문제가 발생하면 내부적으로도 혼선이 발생하지만, 외부와의 관계에 심각한 영향을 미칠 수 있다. 따라서 데이터 품질 이해당사자들은 데이터의 일관성을 유지하기 위해 분산된 데이터에 대해 세심히 관리할 필요가 있다.

4.1 데이터 품질 계획

4.1.1 상황

모든 데이터에 대하여 모든 기준을 적용하여 데이터 품질을 점검하는 것은 불가능하다. 데이터 품질 관리 대상이나 기준은 절대적인

것이 아니고 상대적이기 때문이다. 즉, 조직의 사업 전략에 따라 데이터별로 중요성이 다르며, 데이터 품질의 기준이나 품질 수준도 상이하다. 그러므로 데이터 관리자는 조직의 사업 전략에 부응하는 데이터 품질 관리 계획을 수립해야 한다.

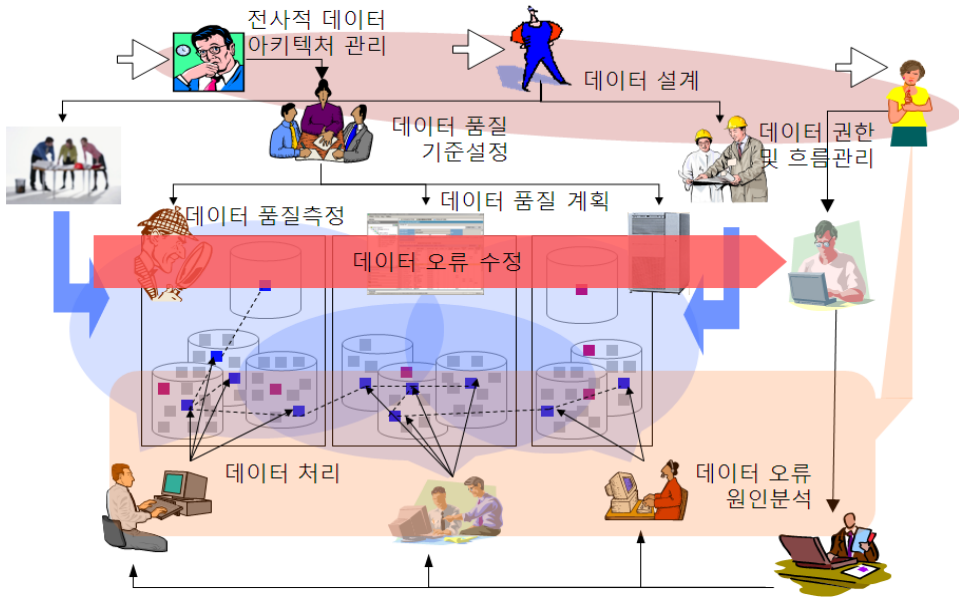
4.1.2 시나리오

UNSPSC 코드를 사용하고 있는 회사 A는 eCI@ss 코드를 사용하고 있는 회사 B에 제품을 납품하고 있다. 그러나 제품의 분류 코드가 상이하여 데이터 오류의 원인이 되고 있다. 따라서 회사 B는 회사 A에 대해 eCI@ss코드를 사용하여 제품을 납품하도록 요구하게 되었다. 결국, 회사 A는 두 개의 분류 코드를 사용하기로 결정하였다. 이에 따라 두 코드간의 통합 코드 관리 시스템을 마련하여 두 분류 코드 간의 매핑을 지원하고 코드 입력 오류를 사전에 예방하는 계획을 수립하였다.

4.2 데이터 품질 기준 설정

4.2.1 상황

데이터 품질 측정 항목, 대상 데이터, 특정 방식 등의 기준이 사전에 정의되지 않으면 데이터 품질 측정 프로세스가 체계적으로 수행될 수 없다. 조직의 관련자들은 충분한 협의를 통해 체계적이며 공식적으로 사용할 기준을 설정해야 한다. 데이터 품질관리 프레임워크가 사용되는 전체 환경을 <그림 6>에 나타내었다.



〈그림 6〉 데이터 품질관리 프레임워크 사용 시나리오

4.2.2 시나리오

자체 식별 코드의 품질을 측정하기 위하여 품질 기준으로 도메인 무결성과 참조 무결성을 택한다. 대상 데이터는 공장 A에서 사용되는 모든 정보시스템의 데이터이며, 측정 방법으로는 상용화된 프로파일링 도구나 직접 작성된 SQL 문을 이용한다.

4.3 데이터 품질 측정

4.3.1 상황

일반적으로 데이터 오류는 데이터 처리 과정에서 사용자에게 의해 우연히 발견되며, 발견된 데이터 오류는 사용자의 능력이나 업무 범위 내에서 수정되는 경향이 있다. 따라서

데이터 오류에 대한 수정을 사용자에게만 의존하게 된다면, 인식되지 못한 데이터 오류가 점진적으로 증가하게 될 것이다. 그러므로 데이터 오류를 찾아내기 위해서는 지속적이고 체계적인 데이터 오류 측정 및 점검이 필요하다. 이러한 측정 및 점검은 데이터 실행자(operator)에 의해서 이루어질 수도 있고, SQL과 같은 프로그램이나 데이터 품질 측정 도구에 의해서 수행될 수도 있다.

4.3.2 시나리오

공장 B에서 데이터 실행자가 과거에 입력된 모든 EAN 코드 및 해당 EAN 코드가 사용된 곳에서 오류가 있는지 점검하고 측정한다. 이때, 실행자는 데이터 베이스에서 오류를 찾기 위해 자체 개발한 SQL 프로그램을 사용한다.

4.4 데이터 권한 및 흐름 관리

4.4.1 상황

데이터 오류 발생 원인은 다양하므로 정확한 원인을 규명하기 위해서는 데이터의 발생에서부터 최종 사용까지의 데이터 흐름이 파악되어야 한다. 또한 데이터 사용에 대한 책임과 권한이 명확하게 규명되어야 한다. 데이터 관리자는 데이터 흐름과 데이터 권한을 분석하여 데이터 오류 발생과 관련 있는 담당자나 프로세스를 파악해야 한다. 그리고 데이터 관리자는 관련자들과 협의를 통하여 근본적인 문제에 대한 대응 방안을 마련해야 한다.

4.4.2 시나리오

공장 A는 UNSPSC 10.1버전을, 공장 B는 최신버전인 11.3을 사용하고 있으며 공장 A의 버전을 최신버전으로 갱신할 필요가 있다. 그러나 공장 A는 협력업체와 분류코드를 공유하고 있어서 공장 A만으로는 갱신을 쉽게 할 수 없는 상황이다. 이 경우 코드와 관련된 이해당사자(공장 A, B, 협력업체 등)간의 협의를 통하여 수정 책임자, 수정 범위, 기한 등을 정한다.

4.5 데이터 오류 원인 분석

4.5.1 상황

데이터 오류 수정은 눈에 보이는 데이터를 대상으로 이루어진다. 그러나 데이터 오류가

발생하는 근본적인 원인을 파악하고 제거하지 않으면, 시간의 흐름에 따라 유사한 종류의 데이터 오류가 반복적으로 발생하게 된다. 데이터 통제자는 데이터 오류 발생의 근본적인 원인을 추적하고 제거해야 동일한 오류 재발을 방지할 수 있다.

4.5.2 시나리오

제품 X에 대해서 공장 A와 공장 B에서 서로 다른 분류 코드가 빈번히 발견되고 있었다. 이에 대한 원인을 분석해보니 공장 A는 UNSPSC 10.1버전을, 공장 B는 11.3버전을 사용하고 있었다. 문제를 근본적으로 해결하기 위해서는 공장 A의 버전을 11.3버전으로 갱신해야 한다.

4.6 데이터 오류 수정

4.6.1 상황

데이터 실행자는 데이터 품질 측정 및 점검 과정에서 데이터 오류를 발견하면 수정해야 한다. 그런데 동일한 데이터가 다른 시스템에도 존재할 수 있고, 해당 데이터와 관련되는 데이터도 다수 존재할 수 있다. 따라서 데이터 실행자는 관련 데이터 간의 관계에 따라 해당 데이터를 모두 수정해야 한다.

4.6.2 시나리오

UNSPSC 버전 10.1에서는 제품 A의 분류 코드가 12345671이었는데 버전 11.3에서는 제품 A의 코드가 12345680로 변경되었다. 이 경우, 제품 A의 모든 분류코드와 관련 데이

터를 수정해야 한다.

4.7 전사적 데이터 아키텍처 관리

4.7.1 상황

데이터 관리자는 다수의 정보 시스템에 분산된 품질관리 대상 데이터와 데이터 구조를 일목요연하게 파악하기 위해 전사적 데이터 아키텍처를 관리한다. 전사적 데이터 아키텍처는 기존 데이터의 위치 및 연관 관계를 파악하는데 이용될 뿐만 아니라, 신규 데이터가 기존 데이터 구조와 일관성을 유지하는데도 이용된다.

4.7.2 시나리오

A공장은 자체 식별 코드, B공장은 EAN 식별 코드를 사용할 경우 전사적 차원에서 A공장과 B공장의 두 코드 체계를 통합적으로 관리해야 한다. 통합관리하기 위해서는 자체 식별코드와 EAN 식별 코드를 매핑하는 개념 데이터 모델이 마련되어야 한다.

4.8 데이터 설계

4.8.1 상황

구조적 문제로 인하여 발생하는 데이터 오류는 반복적으로 발생하는 경향이 있다. 이 문제를 해결하기 위해서는 전사적 관점에서 데이터 스키마를 다시 설계해야 한다.

4.8.2 시나리오

자사 상품 식별 코드 체계에서 check digit

기능이 없을 경우, 데이터 입력 오류 방지를 위한 체크 디짓 기능을 추가하여 상품 식별 코드 구조와 DB 스키마를 재설계한다.

4.9 데이터 처리

4.9.1 상황

데이터 구조가 잘 설계되어 있다고 해도, 사용자들이 데이터를 무성의하게 입력하거나, 관련 데이터가 모두 변경되지 않으면 동일해야 할 데이터 값이 서로 틀리는 결과가 발생한다. 그러므로 데이터 사용자가 데이터를 처리할 때 정해진 규칙을 준수하도록 해야 한다.

4.9.2 시나리오

EAN 코드 입력시 정보 시스템이나 사용자가 check digit 검사 규칙을 따라야 한다.

5. 결 론

본 연구는 데이터 품질 확보를 위한 체계적 절차를 3단계 활동과 9개의 프로세스로 프레임워크를 구성하여 각 프로세스별로 프로세스의 필요성, 프로세스의 기능 측면과 역할 측면에서의 수행 활동, 다른 프로세스와의 연관성을 정립하였다. 또한 e비즈니스 표준코드인 상품식별 코드와 상품분류 코드의 예를 통해 데이터 품질관리 프레임워크를 사용하는 시나리오를 제시하였으며 각 프로세스 별

로 데이터 사용 상황과 그 상황에 적합한 시나리오를 제공하였다.

본 논문에서 사용된 9개 프로세스의 일부는 2009년 조달청에서 실시한 ‘상품정보 체계 정비 및 품질제고 사업’에 적용되었다. 이 사업에는 본 프레임워크에서 제안된 관리, 통제, 구현이라는 3계층 개념을 그대로 구현하였다. 또한 국제표준화기구인 ISO에 8000-150표준으로 초안을 제출하여 투표를 앞두고 있어 국제표준으로 제정될 가능성이 큰 실정이다.

문헌 연구에서 살펴보았듯이 데이터 품질 관리 프레임워크는 e비즈니스 뿐만 아니라 의료, 국방 등 전 산업 분야에 걸쳐 적용되고 있다. 본 연구에서는 e비즈니스의 표준 코드를 중심으로 시나리오를 제시하였으나 향후 다양한 산업군에서 적용 가능한 시나리오가 개발되어 현장에 보급될 필요가 있다. 또한 본 프레임워크는 ISO TC184/SC4(산업 데이터 표준)에서 ISO 8000-150(마스터 데이터 품질관리 프레임워크) 표준으로 추진 중에 있으며 다양한 산업군에 적용됨으로써 표준의 효용을 극대화할 필요가 있다.

본 프레임워크를 구성하는 9개 프로세스는 상호 연관성을 가지며 단계별 순환 구조로 운영되도록 설계되어있다. 프레임워크의 전체 구성은 상위레벨에서 설명하고 있으나 하위 프로세스 간의 연계성을 보다 구체적이고 가시적으로 정의하기 위한 모델링이 필요하다. 각 프로세스별로 입력과 출력을 명확히 하여 프레임워크 사용자가 모호함을 느끼지 않도록 할 필요가 있으며 특히 시나리오 레벨에서는 산출물 수준의 가시성을 확보할 필요가 있다.

참 고 문 헌

- [1] Abbas, S. Q., Aggarwal, A., “Development of a Structured Framework To Achieve Quality Data,” *International Journal of Advanced Engineering and Application*, 2010.
- [2] Batini, C., Scannapieco, M., *Data Quality : Concepts, Methodologies and Techniques*, Springer, 2006.
- [3] Canadian Institute for Health Information. *The CIHI Data Quality Framework Version Vol, 1*, 2001.
- [4] Cykana, P., Paul. A., Stern, M., “DOD Guidelines on Data Quality Management” *Proceedings of the 1996 Conference on Information Quality*, Cambridge, MA, 1996, pp. 154-171.
- [5] Defense Information Systems Agency, *DOD Guidelines on Data Quality Management(Summary)*, <http://tricare.osd.mil/rm/documents/fa/DoDGuidelinesOnDataQualityManagement.pdf>, 2004.
- [6] DEFRA, *RADAR Data Quality Framework(Ver. No. 1)*, 2004.
- [7] Eppler, M. J., Wittig, D., “Conceptualizing Information Quality : A Review of Information Quality Frameworks from the Last Ten Years,” *Proceedings of the 2000 Conference on Information Quality*, 2000, pp. 83-96.
- [8] Fürber, C., Hepp, M., “Using SPARQL and SPIN for Data Quality Management on the Semantic Web,” *13th International Conference on Business Information Systems*, Springer LNBIP, 2010.
- [9] Fürber, C., Hepp, M., “Using Semantic

- Web Resources for Data Quality Management,” Proceedings of the 17th International Conference on Knowledge Engineering and Knowledge Management, Springer LNCS Vol. 6317, 2010, pp. 211-225.
- [10] GS1, Data Quality Framework, 2010.
- [11] ISO 8000-1 Information Quality (draft), 2008.
- [12] ISO/CD-TS 8000-110 Information quality-Part 110 : Master data quality : Syntax, semantic encoding, and conformance to customer requirements, 2008.
- [13] ISO/CD-TS 22745-1 Open Technical Dictionary-Part 1 : Overview, 2008.
- [14] ISO 9000 : 2005(E), Quality management systems-Fundamentals and vocabulary.
- [15] Kerr, K., The Development of a Data Quality Framework and Strategy for the New Zealand Ministry of Health, 2003.
- [16] Li, W., Lei, L., “An Object-Oriented Framework for Data Quality Management of Enterprise Data Warehouse,” Volume 4099 of Lecture Notes in Computer Science, 2006, pp. 1125-1129.
- [17] Lin, S., Gao, J., Koronios, A., Chanana, V., “Developing a data quality framework for asset management in engineering organisations,” *Int. J. Information Quality*, Vol. 1, No. 1, 2007, pp. 100-126.
- [18] Lucas, A., Reis, A., Caleira, M., “The quality of monitoring data in Civil Engineering works,” *Comunicação apresentada na 14th International Conference on Information Quality*, 2009.
- [19] Madnick, S. E., Wang, R. Y., Lee, Y. W., Zhu, H., “Overview and Framework for Data and Information Quality Research,” *Journal of Data and Information Quality(JDIQ)*, Vol. 1, No. 1, 2009.
- [20] Naqvi, M., Peter Dewsbury, P., *Data Quality Management Framework for Local Authorities*(Vol. 3, No. 3), Tribal, 2009.
- [21] Otto, B., “Data Quality Management : Framework and Approach for Data Governance,” 5th German Information Quality Management Conference, Bad Soden/Ts., 2007.
- [22] Otto, B., Wende, K., Schmidt, A., Osl, P., “Towards a framework for corporate data quality management,” *Proceedings of the 18th Australasian Conference on Information Systems*, 2007, pp. 916-926.
- [23] Radziwill, N. M., DuPlain, R., “A Framework for Telescope Data Quality Management,” *ADASS XV*, Madrid, 2005.
- [24] Russom, P., “Master Data Management: Consensus-Driven Data Definitions for Cross-Application Consistency,” *The Data Warehousing Institute*, 2006.
- [25] Scannapieco, V., Mirabella, M., Mecella, C. Batini, “Data Quality in E-Business Applications,” *WES 2002*, Springer-Verlag LNCS Volume : Vol. 2512, 2002, pp. 121-138.
- [26] Sha, K., Shi, W., “Consistency-driven data quality management of networked sensor systems,” *Journal of Parallel and Distributed Computing*, Vol. 68, No. 9, 1998, pp. 1207-1221.
- [27] Shankaranarayanan G., Cai, Y., “A Web Services Application for the Data Quality Management in the B2B Networked

- Environment,” Proceedings of the 38th Hawaii International Conference on System Sciences, 2005.
- [28] Tee, S. W., Bowen, P. L., Doyle, P., Rohde, F. H., “Factors Influencing Organizations to Improve Data Quality in their Information Systems,” Accounting and Finance, Vol. 47, No. 2, 2007, pp. 335-355.
- [29] Turner, S., “Defining and Measuring Traffic Data Quality,” Proceedings of the Traffic Data Quality Workshop, Washington, DC, 2002.
- [30] Wang, R. Y., “A Product Perspective on Total Data Quality Management,” Vol. 41, No. 2 Communications of the ACM, 1998, pp. 58-65.
- [31] Wang, R. Y., Lee, Y. W., Pipino, L. L. and Strong, D. M., “Manage Your Information as a Product,” Sloan Management Review, Vol. 39, No. 4, 1998, pp. 95-105.
- [32] 김선호, 이창수, “마스터 데이터 품질관리 프레임워크”, Entrue Journal of Information Technology, 제9권 제2호, 2010.
- [33] 김선호, 이창수, 이제현, “프로세스 기반의 전자카탈로그 데이터 품질관리”, 한국전자거래학회지, 제14권, 제3호, 2009.
- [34] 한국과학기술정보연구원, 글로벌동향브리핑(GTB), 2010.
- [35] 한국데이터베이스진흥센터, 데이터품질관리지침(Vol. 2, No. 1), 2006.
- [36] ECIF 15:2002 전자카탈로그 상품분류코드, 전자상거래표준화통합포럼, 2002.

저 자 소 개



이창수

(E-mail : cslee@gwnu.ac.kr)

강릉원주대학교 산업정보경영공학과 교수

서울대학교 산업공학과 (석·박사)

서울대학교 산업공학과 (학사)

관심분야

데이터 품질, 마케팅정보시스템, 검색엔진마케팅



김선호

(E-mail : shk@mju.ac.kr)

명지대학교 산업경영공학과 교수

Pennsylvania State University 산업공학과 (석·박사)

서울대학교 산업공학과 (학사)

관심분야

Data Quality, 전자거래 표준화, BPM,