

데이터품질관리 성숙도모델에 대한 연구

A Study of Data Quality Management Maturity Model

김찬수(Chan Soo KIM)*, 박주석(Joo Seok Park)**

초 록

오늘날 정보화 사회에서 경쟁하는 기업들에 있어서 데이터품질 저하는 기업경쟁력 하락과 새로운 비용창출이라는 부정적인 영향요인으로써 작용하고 있다. 이러한 데이터품질 저하의 문제를 해결하기 위해 데이터품질에 대한 많은 선행연구들이 진행되어 왔으며, 데이터품질의 측면 중 결과적이고 현상적인 품질개념인 데이터값의 품질과 데이터서비스의 품질에 대해 주로 연구되어 왔다. 이에 반해 본 연구에서는 원인적인 데이터품질 개념인 데이터의 구조적 품질을 메타데이터 관리의 관점에서 연구하였으며, 이를 통해 평가와 개선을 위한 관리의 관점이 적용된 데이터품질관리 성숙도모델을 제시하였다. 또한 본 연구에서 제시한 데이터품질관리 성숙도모델의 타당성 검증을 위해 데이터품질 관리단계가 성숙될수록 데이터 품질수준이 높아지게 된다는 것을 실증적으로 검증하였다.

ABSTRACT

In companies competing for today's information society, Data quality deterioration is causing a negative influence to generate company competitiveness fall and new cost. A lot of preceding study about data quality have been proceeded in order to solve a problem of these data quality deterioration. Among the sides of data quality, it has been studied mainly on quality of the data value and quality of data service that are the results quality concept. However, this study studied structural quality of the data which were cause quality concept in a viewpoint of metadata management and presented data quality management maturity model through this. Also empirically this study verified that data quality improved if the management level matured.

키워드: 데이터품질, 메타데이터, 성숙도모델, Data Quality, Metadata, Maturity Model

* 경희대 MIS 박사과정(keytoyou@empal.com)

** 경희대 MIS교수(joseok@khu.ac.kr)

■ 논문 접수일 : 2003. 11. 25

■ 게재 확정일 : 2003. 12. 4

1 서 론

오늘날 정보화사회에서 경쟁하는 기업들에 있어서 정보품질은 기업경쟁력의 첨병으로서의 역할을 하고 있으며, 이로 인해 정보품질의 근간이 되는 정보시스템의 데이터품질은 그 중요성이 점점 높아지고 있다. 이러한 상황에서 데이터품질의 저하는 기업의 여러 부문에 대해 부정적인 영향요인으로 작용하고 있다. 즉 품질이 낮은 데이터로 인해 고객만족도 하락, 높은 운영비용, 비효율적인 의사결정 및 기업전략수행 지연 등의 기업경쟁력 저하의 요인으로써 작용하고 있을 뿐만 아니라, 종업원의 사기저하, 조직 간의 불신으로까지 이어져 조직문화에도 부정적인 영향요인으로써 작용하고 있다.

이러한 데이터품질 저하로 인한 문제점들을 해결하기 위해 데이터품질에 대한 많은 선행연구들이 진행되어 왔다. 그러나 선행연구들은 대부분 데이터품질의 측면 중 데이터값의 품질과 데이터서비스의 품질이라는 결과적으로 나타나는 품질요인에 대해 연구되어 왔으며, 주로 품질수준평가의 관점에서 연구되어 왔다. 본 연구는 데이터품질의 요소 중 데이터의 구조적인 품질과 구조적인 품질의 관리단계에 주안점을 두어 연구를 진행하였다. 잘못된 데이터값과 데이터서비스는 데이터의 구조적

인 문제점에 기인하여 발생하게 된다. 데이터의 구조적 품질을 제대로 관리하고 평가하기 위해서는 데이터의 데이터인 메타데이터에 대한 관리가 필수적으로 요구된다.

즉 품질 높은 데이터의 구조를 유지하기 위해 관리되어야 하는 메타데이터는 무엇인가를 정의하고, 이러한 메타데이터를 관리해 나가기 위한 데이터품질관리의 성숙도모형을 제시하고자 한다. 데이터품질관리를 위한 성숙도모형의 제시를 통해 기존연구의 한계점인 결과론적인 현상품질에 대한 평가개념을 발전시켜, 관리하고 개선해 나갈 수 있는 데이터의 구조적품질 유지방안을 제시하는 것이 본 연구의 목적이다. 또한 본 연구에서 제시하고자 하는 데이터품질관리 성숙도모형을 통해 기업들이 현재 자사의 데이터품질관리 수준의 단계를 파악하고, 나아가 보다 높은 데이터품질수준을 획득함으로써 기업경쟁력을 제고할 수 있는 지침을 제공하는 것이 본 연구의 목적이다.

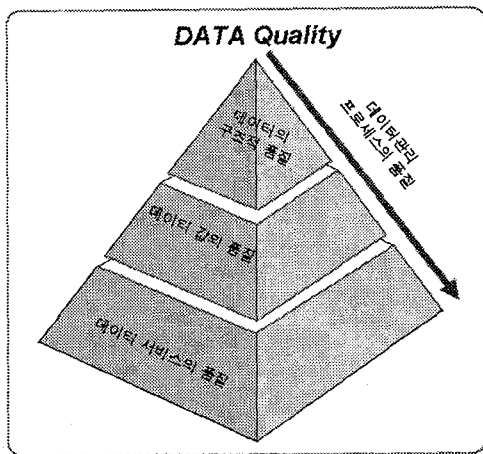
2. 데이터품질 아키텍처의 확장정의

2.1 데이터품질 영역의 정의

데이터품질에 대한 초기의 연구들은 주로 데이터값의 품질(정확성, 최신성, 포괄성 등)과 데이터 서비스의 품질(접

색성, 편의성 등)과 같은 데이터의 현상적 품질에 대해 이루어졌다. 이후 이러한 데이터의 현상적 품질들은 결과적으로 나타나는 품질로써, 원인적 품질 개념인 데이터의 구조적 품질이 제대로 관리될 때 현상적인 품질들도 결과적으로 관리될 수 있다는 관점을 바탕으로, 데이터의 구조적 품질에 대한 연구로 발전되었다. 데이터의 구조적 품질은 데이터 설계구조의 품질을 의미한다.

최근에는 기존의 데이터값, 데이터서비스, 데이터의 구조 뿐만 아니라 데이터의 관리프로세스 개념을 모두 포함하는 통합적인 데이터품질관리에 대한 연구로 발전하고 있다. 기존의 연구들은 주로 평가의 관점에서 이루어져 왔으나, 통합적 데이터품질관리에 대한 연구에서는 평가뿐만 아니라 지속적인 품질수준의 유지 및 개선의 관점을 반영하여 연구되고 있다.

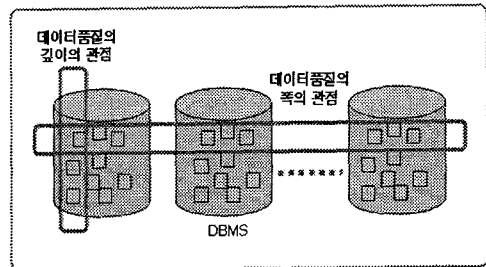


〈그림 1〉 데이터품질 구성영역

지금까지 연구되어온 데이터의 품질요소를 종합하면 크게 세가지 영역으로 분류하여 볼 수 있다. 즉, 데이터값과 데이터서비스라는 현상적 품질요소, 데이터의 구조적 품질요소, 데이터 관리 프로세스의 품질요소로 분류하여 볼 수 있다. 〈그림 1〉은 이러한 데이터품질의 구성영역을 나타낸 그림이다.

2.2 수평적관점의 데이터품질 정의

데이터품질의 평가와 관리는 깊이와 폭의 관점을 가지고 수행되어야 한다. 데이터품질에 있어 깊이의 관점(수직적 품질관점)은 데이터 그자체의 독립적인 측면에서의 품질을 의미한다. 이는 데이터의 정확성, 완전성, 최신성, 검색성 등과 같은 품질 구성요소를 의미한다.



〈그림 2〉 폭과 깊이 관점의 데이터품질

이에 반해 데이터품질에 있어 폭의 관점(수평적 품질관점)은 분산된 정보 시스템, 분산된 데이터베이스환경에서 데이터와 데이터간의 전체적인 연계성과 통합성을 고려한 품질을 의미한다.

〈표 1〉 계층적 구조 관점의 메타데이터 분류

유 형	정 의	예
표준데이터정보	전사적 관점에서 정의된 논리적인 수준의 공통 데이터요소(비즈니스 객체)에 대한 정보	전사적 데이터모델 차원에서 도출되는 공통적인 데이터에 대한 엔티티정보, 속성정보, 도메인정보 등
물리데이터베이스 정보	각 단위시스템에서 생성된 물리적인 데이터베이스시스템의 시스템 카탈로그정보	데이터베이스명, 테이블명, 뷰명, 컬럼명 등의 물리적인 DBMS의 관리 정보
표준데이터와 물리데이터베이스의 매핑정보	표준데이터와 물리적인 데이터베이스간의 연결관계를 관리하는 정보	소스 및 타겟매핑정보, 변환규칙 등의 중개자역할에 관한 정보

이는 전사적인 통합과 일관성의 관점에서 데이터의 단일성, 일치성과 같은 품질 구성요소를 의미한다. 〈그림 2〉는 이러한 폭과 깊이 관점의 데이터품질을 표현한 것이다.

기존 연구는 데이터의 품질을 데이터의 깊이라는 측면에서 주로 연구해 왔다. 그러나 오늘날과 같이 분산되고 복잡화된 정보시스템환경에서는 데이터의 폭에 대한 관점을 고려하지 않고서는 데이터 품질평가의 객관성에 의미를 부여하기 어렵다.

예를들면 독립적인 관점에서 A부서의 제품이라는 데이터가 정확하고 완전하게 입력되어 있다고 평가되었다 하더라도, B부서에서는 동일한 대상에 대해 생산품이라 명명하고 다른 코드체계, 다른 도메인으로 관리하고 있다면, 데이터의 품질이 제대로 관리되고 있다고 할 수 없기 때문이다. 즉 데이터품질은 폭과 깊이의 개념을 모두 고려하여 평가·관리 되어야 한다.

특히 폭의 데이터품질을 높이기 위해서는 전사적인 연계와 통합 관점에서의 데이터품질관리가 이루어져야 한다. 이러한 전사적 데이터품질관리를 위해서는 전사적으로 사용되는 데이터에 대한 표준화와 데이터아키텍처의 관리를 통해 데이터의 구조적 품질이 필수적으로 관리되어야 한다. 이에 대해서는 본 연구에서 제시되는 데이터품질관리를 위한 성숙도모델에서 보다 자세히 기술하기로 한다.

2.3 메타데이터 아키텍처의 정의

정보시스템구조의 관점에서 데이터의 구조적품질관리를 위해 관리되어야 할 메타데이터 카테고리를 〈표 1〉과 같이 개념화하여 분류해 보았다.

표준데이터정보에서는 데이터의 구조 및 도메인에 대한 전사적인 공통표준을 설정한 후, 이러한 표준에 의해 데이터를 개발·사용함으로써 데이터의 구조

적 품질수준을 향상시키고, 결과적으로 데이터의 전사적 품질수준을 유지·발전시키려는 것이 목적이다. 이를 위해 데이터모델링의 논리적 모델링의 단계에서 분석되는 논리적 개념의 데이터객체를 우선적으로 관리하고자 한다. 이러한 논리적 개념의 데이터객체 중 전사적으로 표준화하여 사용해야할 데이터요소를 선별하여 메타데이터정보를 관리하게 된다. 또한 이러한 논리단위의 데이터가 물리적 데이터베이스의 테이블로 변환되어 사용되기 위한 물리적 메타데이터정보를 2차적으로 관리하게 된다. 마지막으로 논리단위와 물리단위의 메타데이터정보를 가지는 데이터객체가 실제 분산되어 있는 각 단위의 정보시스템과 어떻게 연결되는가에 대한 정보를 관리하게 된다. 이러한 매핑정보의 관리는 두가지 관점에서 이루어지게 된다. 첫째는 새롭게 개발되는 신규 정보시스템과의 연결정보이고, 둘째는

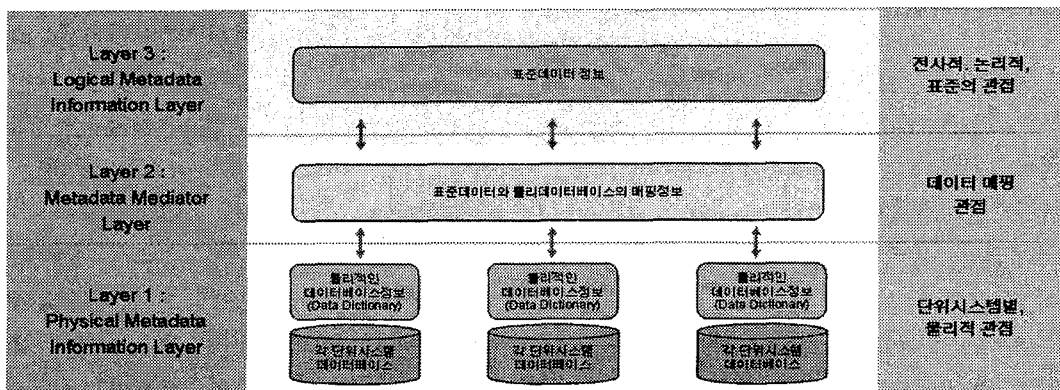
기존 레거시 시스템과의 변환·매칭정보이다.

위의 세가지 메타데이터의 역할관점을 기반으로하여 정보시스템구조 관점에서 관리의 대상이 논리적인 정보인가, 물리적인 정보인가, 논리적인 데이터와 물리적인 데이터를 연결하는 정보인가에 따라 논리단위의 메타데이터정보, 물리단위의 메타데이터정보, 매핑 메타데이터정보로 분류하였다.

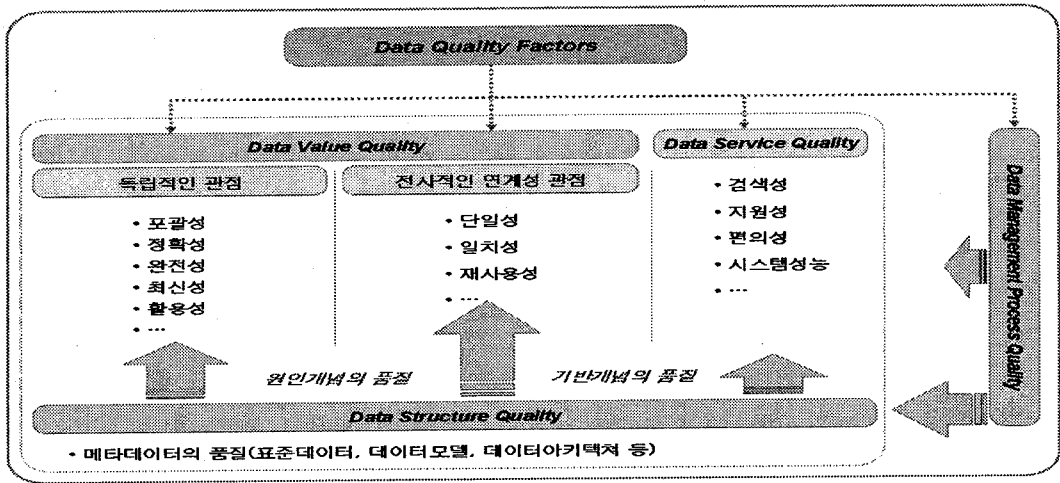
논리단위의 메타데이터정보에서는 데이터요소(Data Element: Attribute), 주단어(Primary Word: Entity), 데이터모델(Data Model) 자체와 이들에 대한 Naming Rule, 도메인, 소속 주단어, 데이터요소간·주단어간 관계 등의 논리적 속성정보가 관리되어야 한다.

물리단위의 메타데이터정보에서는 테이블명, 칼럼명, 스키마 등의 물리적 속성정보가 관리되어야 한다.

매핑 메타데이터정보에서는 사용되고



〈그림 3〉 메타데이터 3Tier 아키텍처



〈그림 4〉 확장된 데이터품질 아키텍처 모형

있는 데이터베이스명, 테이블명, 변환 규칙 등의 매핑정보가 관리되어야 한다.

이러한 세가지 메타데이터를 통해 데이터의 구조적 품질은 전사적인 차원에서 유기적으로 연동되어 관리되어질 수 있다. 〈그림 3〉은 정보시스템계층의 관점에서 메타데이터 아키텍처를 도시한 것이다.

2.4 데이터품질 아키텍처의 확장정의

앞절에서 데이터품질의 구성영역, 수평적관점의 데이터품질, 데이터의 구조적품질관리를 위한 메타데이터 아키텍처를 정의하였다. 이를 종합하여 데이터품질의 아키텍처를 도시하면 〈그림 4〉와 같다.

데이터의 품질은 기존에 연구된 독립적인 관점에서의 품질과 전사적 연계성

의 관점에서의 품질로 확장하여 정의되어야 한다. 이는 앞 절에서 설명한 데이터품질의 폭과 깊이의 관점을 반영한 것이다. 또한 데이터값과 서비스의 품질과 같은 현상적인 품질은 데이터의 구조적 품질에 의한 결과요인이라 할 수 있다. 그러므로 데이터의 구조적 품질이 선행되어 평가·관리 되어질 때 현상적 품질 또한 유지·발전 되어질 수 있다는 품질관리의 개념으로 확장 정의되어야 한다. 특히 데이터의 구조적 품질은 전사적인 폭의 관점을 가져갈 수 있도록 평가·관리 되어져야 한다. 이를 위해서는 단위정보시스템의 데이터의 구조적 품질 뿐만 아니라 전사적 차원의 데이터의 구조적 품질을 평가관리할 수 있는 개념으로 확장되어야 한다. 데이터의 구조적 품질관리를 위해서는 메타데이터의 품질이 제대로

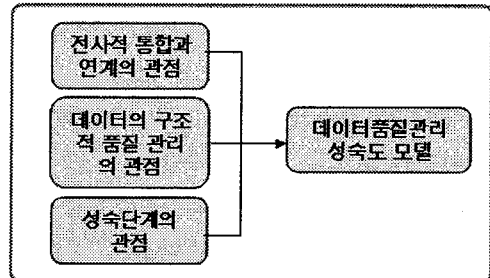
관리되어야 하며, 특히 전사적 연계성 관점에서 데이터표준, 데이터아키텍처가 제대로 관리되어야 한다. 이러한 전사적 데이터의 구조적 품질관리에 따라 전사적인 연계성 관점의 품질과 독립적인 관점의 품질모두가 향상될 수 있다. 데이터관리프로세스의 품질은 1차적으로 데이터를 설계하게 되는 데이터의 구조적 품질을 관리하는 품질이며, 나아가 데이터의 구조적품질을 유지하고 발전시키는 관리프로세스를 통해 데이터의 현상적 품질인 데이터값과 서비스의 품질수준까지 상승시키는 효과를 가져 오는 역할을 하게 된다.

즉 데이터의 총체적 품질향상을 위해서는 데이터의 구조적 품질수준이 높아져야 하며, 이러한 데이터의 구조적 품질수준을 높일 수 있도록 데이터관리 프로세스의 품질이 발전되어야 한다. 그러므로 데이터의 구조적 품질관리를 위한 관리요소를 성숙단계별로 정의하고 관리할 수 있어야 한다. 이러한 성숙단계의 정의에 의해 기업의 데이터품질관리의 현 수준을 파악해내고, 보다 높은 데이터품질수준으로 성숙될 수 있는 가이드라인이 마련될 수 있을 것이다. 또한 성숙단계의 정의는 현상적인 데이터 품질의 평가의 한계점을 극복할 수 있는 방안이기도 하다.

3 데이터품질관리를 위한 성숙도모델

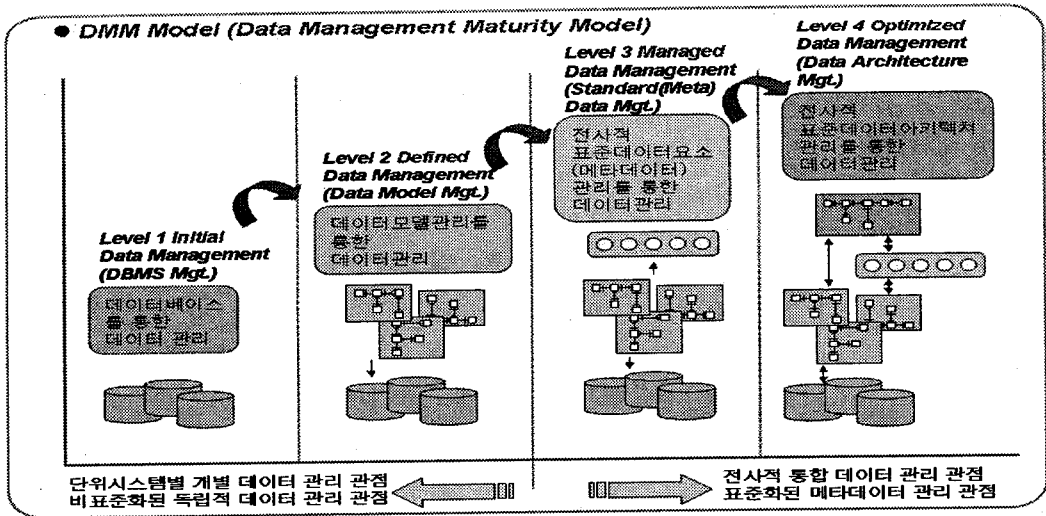
3.1 데이터품질관리 성숙도모델의 제시

데이터품질의 평가와 관리를 위해서는 앞서 언급된 바와 같이 전사적인 통합과 연계의 관점, 데이터의 구조적 품질 관리관점, 관리의 성숙단계가 정의 되어야 한다. 본 연구에서 제시하는 데이터품질관리를 위한 성숙도 모델은 <그림 5>와 같이 위의 세가지 관점을 반영하여 제시된 모델이다.



<그림 5> 성숙도모델에 반영된 관점

데이터의 전사적 연계성과 통합의 관점을 가져가기 위해서는 전사적인 관점에서 표준화된 데이터에 대한 메타데이터의 관리와 데이터아키텍처에 대한 관리가 이루어져야 한다. 이러한 개념을 기반으로 소프트웨어프로세스 평가를 위한 성숙도모델(Capability Maturity Model)의 개념을 데이터관리의 성숙도 모델로 적용하여 제시하면 <그림 6>과



〈그림 6〉 데이터품질관리 성숙도모델

같다.

Level1 Initial Data Management의 단계는 데이터베이스시스템을 통한 데이터 관리의 단계이다. 이 단계에서는 데이터베이스의 시스템카탈로그에 정의된 규칙을 통해 데이터의 구조적품질이 관리되는 초기적인 수준이라고 할 수 있다. Level2 Defined Data Management의 단계는 데이터모델의 관리를 통한 데이터 관리의 단계이다. 이 단계에서는 데이터베이스의 설계를 위해 논리적인 데이터모델과 물리적인 데이터모델의 설계를 통해 데이터베이스를 설계하고 데이터를 관리하는 단계이다. 데이터의 구조적 변경이나 추가시에는 데이터모델을 참조하여 이루어지게 되며, 이러한 변화는 1차적으로 데이터모델에 반영되어 다시 데이터베이스에 피드백되어 반영되게 된다. Level3 Mana-

ged Data Management의 단계는 데이터표준화를 통한 데이터 관리의 단계이다. 이 단계에서부터는 앞의 Level1, Level2와는 달리 전사적인 통합과 연계의 관점 하에 데이터표준화가 이루어지게 된다. 전사적으로 사용되어지는 데이터를 선별하고 선별된 데이터에 대해 각종 속성 및 스키마, 도메인, 데이터모델에 대한 표준화를 가져가게 되는 메타데이터 관리의 단계이다. 메타데이터 표준화를 통해 표준화된 데이터의 공유와 재사용, 단위정보시스템간 연계 및 통합을 이루고자 하는 단계이다. Level4 Optimized Data Management의 단계는 데이터아키텍처관리를 통한 데이터 관리의 단계이다. 이 단계에서는 전사적인 표준데이터 아키텍처 모델을 정의하고, 이를 기반으로 데이터, 데이터모델, 데이터간 관계에 대한

DMM Level	Data Service Quality	
	Data Value Quality	
	독립적인 관점	전사적인 연계성 관점
Level 1 Initial Data Management (DBMS Mgt.)	<ul style="list-style-type: none"> 포괄성 정확성 완전성 최신성 활용성 	<ul style="list-style-type: none"> 중복성 불일치성 재사용성
Level 2 Defined Data Management (Data Model Mgt.)	→	
Level 3 Managed Data Management (Standard(Meta) Data Mgt.)	→	→
Level 4 Optimized Data Management (Data Architecture Mgt.)	→	→

〈그림 7〉 성숙단계별 데이터품질 성취영역

표준을 관리해 나가는 가장 최적화된 데이터관리의 단계이다.

제시된 데이터품질관리 성숙도에 따라 성취될 수 있는 데이터품질의 영역들을 도시하면 〈그림 7〉과 같다.

Level1과 Level2의 단계에서는 데이터품질의 독립적인 관점, 즉 데이터 품질의 깊이의 관점에서의 품질수준을 만족시킬 수 있다. 이는 Level1과 Level 2가 각각의 독립적인 정보시스템관점에서 데이터품질관리가 이루어지는 단계이기 때문이다. 단순히 DBMS의 물리적 시스템카탈로그에 의해 데이터가 관리되는 Level1 보다는 데이터의 논리적모델이 관리되는 Level2가 데이터품질관리의 수준이 높다고 할 수 있다. Level1에서는 물리적 구현과 퍼포먼스를 위해 Denormalized되어 있어 본질적인 데이터의 형태를 파악하기

어렵다. 이에 비해 Level2에서는 논리적인 단위로 데이터를 규정해둔 논리적 데이터모델을 기반으로 하여 데이터베이스를 관리하기 때문에 데이터의 추가, 변화시에 이를 참조하여 변화관리를 할 수 있어 데이터의 기형화를 막을 수 있다.

Level3와 Level4의 단계에서는 데이터의 독립적인 관점에서의 품질수준 뿐만 아니라 전사적인 연계성 관점에서의 품질수준까지 만족시킬 수 있다. 이는 전사적인 데이터의 표준화를 통해 데이터를 공유하고 재사용할 수 있는 기반이 마련되기 때문이다. 데이터표준화의 단계인 Level3보다는 데이터아키텍처까지 관리되는 Level4에서 보다 전사적인 연계성을 가진 데이터품질관리가 가능해진다. 이는 데이터표준화의 단계에서는 데이터의 아키텍처에 대한

관점의 부족으로 고립된 데이터의 표준화현상이 발생되기 때문이다. 이러한 고립된 데이터의 표준화현상에 대해서는 뒷절의 성숙단계별 이슈에서 자세히 언급하기로 한다.

3.2 성숙단계별 이슈사항 및 해결방안의 제시

본 절에서는 본 연구에서 제시한 데이터품질관리 성숙도모델의 각 단계별 데이터품질관리 이슈사항을 설명하고, 해결방안을 제시함으로써 성숙단계의 타당성을 제시하고자 한다.

3.2.1 Level1 DBMS를 통한 데이터 품질 관리의 이슈사항 및 해결 방안

Level1에서는 데이터베이스관리시스템에 의해 데이터를 관리하는 단계이다. 이 단계에서는 데이터베이스의 시스템카탈로그를 통해 데이터를 정의하게 되며, 테이블이라는 물리적인 수준에서 데이터를 관리하게 된다. 즉 테이블의 물리적 속성정의와 테이블간의 참조관계(Referential Integrity)에 의해 데이터가 관리되는 단계이다. 그러나 테이블 수준에서는 데이터가 비정규화 되어 있을 가능성이 높다. 즉 물리적인 데이터베이스시스템의 다양한 종류에 따라, 그리고 퍼포먼스나 프로그램 구축의 관점에서 데이터가 변형되고

통합되어 있을 가능성이 높다. 데이터 자체는 물리적인 시스템과는 별도의 기업비즈니스를 설명하는 객체이다. 그러나 이러한 개념적 비즈니스 객체인 데이터가 물리적으로 시스템화 되면서 본래의 모습이 통합되거나 변형되는 경우가 발생하게 된다. 시스템 구축의 초기에는 이러한 통합과 변형의 정도가 약하고 또한 통합되어 있다고 하더라도 어떻게 통합되어졌는지를 개발자나 관리자가 인지하고 있으나, 시간이 경과되면서 본래의 개념적 데이터의 모습을 점점 더 상실하게 된다. 왜냐하면 데이터의 본래의 개념적 모델이 관리가 되지 않고 통합되고 변형된 물리적인 테이블단위의 관리가 이루어져왔기 때문에, 데이터가 추가되거나 변경되면서 변형된 모습을 기준으로 지속적인 변형이 발생되어 데이터의 모습은 점점더 기형적으로 변하게 되기 때문이다.

또한 물리적인 시스템을 근거로 데이터를 설계하고 관리하기 때문에 논리적인 기업의 비즈니스규칙이 제대로 관리되지 않을 가능성이 높으며, 이 또한 시간이 경과되면서 이러한 비즈니스규칙을 점점 더 관리하기 어려워지게 된다.

이러한 이슈사항을 해결하기 위해서는 논리적데이터모델과 물리적데이터모델을 통해 데이터베이스를 생성하고 관리하여야 한다. 즉 기업의 비즈니스를 시스템과는 무관하게 개념적으로 설명하는 논리적데이터모델과 이를 시스템

화하기 위한 물리적데이터모델의 연결관리를 통해 원래의 개념적이고 논리적인 데이터의 모습을 관리할 수 있어야 한다. 데이터가 추가되거나 통합되거나 수정되어지더라도 논리적인 데이터모델을 참조하여 데이터의 기형적인 변형을 방지할 수 있으며, 데이터의 품질수준을 유지해나갈 수 있다.

3.2.2 Level2 데이터모델을 통한 데이터품질관리의 이슈사항 및 해결 방안

Level2에서는 논리적데이터모델과 물리적데이터모델을 통해 데이터를 관리하는 단계이다. 이 단계에서는 기업의 비즈니스 객체를 논리적이고 개념적인 논리데이터모델을 통해 표현하게 된다. 이후 구축하고자 하는 물리적인 시스템, 데이터베이스, 프로그래밍언어에 따라 물리적인 데이터모델로 변경한 후 이를 기반으로 데이터베이스를 설계하고 구축하게 된다. 그리고 논리적모델과 물리적모델의 연관관계를 관리해 줌으로써 비정규화되고 통합된 물리적 테이블의 원래상태를 추적할 수 있다. 또한 새로운 데이터의 추가나 수정, 비즈니스 규칙의 변경시에도 논리적데이터모델을 참조하여 수정을 하게 되므로 기형적인 시스템으로의 변경을 막을 수 있다는 장점이 있다.

그러나 아직까지 이 단계에서는 단위 정보시스템의 독자적 관점으로 데이터

가 설계되고 구축되었기 때문에, 전사적 연계와 통합의 관점이 약하다. 즉 타조직, 타부서, 타정보시스템과의 데이터의 연관관계를 고려하지 않고 해당 부문의 관점에서만 데이터를 생성하여 관리하기 때문에 타조직, 타부서, 타정보시스템과의 데이터연계와 통합에 많은 어려움을 가지고 있다. 이로 인해 전사적인 관점에서 봤을 때, 어떤 데이터가 어느 시스템에서 어떻게 관리되고 있는지 알 수 없을 뿐만 아니라, 전사적으로 표준화된 데이터 명명법(Naming Convention) 부재로 인해 각 부문간 명칭부여 규칙의 일관성이 결여되어 있다. 또한 조직/부서별, 업무별, 시스템별로 동일 데이터에 대한 명칭이 상이하게 되는 문제점이 발생하게 된다. 이러한 데이터의 상이성은 동일 데이터에 대한 정의, 값영역(Value Domain), 데이터타입, 표현방식 등에서도 나타나게 된다.

즉 전사적인 통합과 연계의 관점에서의 데이터의 획득, 저장, 검색의 메커니즘이 결여되어 있어 데이터의 재사용 및 공유가 어렵고, 시스템설계의 비효율성이 내재되어 중복설계비용이 지속적으로 발생하는 문제점을 가지고 있다. 이러한 문제점은 Level1에서는 더 심하다고 할 수 있을 것이다. 이러한 이슈사항을 해결하기 위해서는 전사적으로 표준화하여 사용해야 할 데이터를 선정하여 데이터명명법, 데이터에 대한

정의, 값영역, 데이터타입, 표현방식 등의 데이터의 각 속성을 표준화하여 관리하여야 한다. 즉 조직/부서별, 업무별, 시스템별로 연계되어 사용되어야 할 핵심적인 공통데이터에 대해 전사적 차원에서 중앙집중적으로 데이터를 표준화하여, 이렇게 표준화된 데이터를 각 부문이 공통적으로 사용함으로써, 데이터의 통합과 연계성을 높여 나갈 수 있도록 해야 한다.

3.2.3 Level3 메타데이터관리(데이터 표준화)를 통한 데이터품질관리의 이슈사항 및 해결방안

Level3에서는 전사적인 데이터표준화를 통해 데이터를 관리해나가는 단계이다. 이 단계에서는 전사적으로 연계성을 가지고 사용해야 할 데이터와 각 부문의 핵심데이터에 대해 표준화를 수행하게 된다.

이렇게 선정된 데이터의 메타데이터에 대해 표준을 정의하고 이를 중앙집중적으로 관리하게 된다. 각 부문은 신규시스템개발시에 표준화된 데이터를 사용하여 시스템을 개발하게 되고, 기존 레거시 시스템은 표준화된 데이터로 마이그레이션하거나 매핑하는 작업을 통해 표준화된 데이터를 사용하게 된다.

이러한 전사적 차원의 데이터의 표준화를 통해 각부문간 데이터의 통합과 연계, 데이터의 재사용 등이 가능해지게 된다. 또한 기능적(업무의 관점), 기술

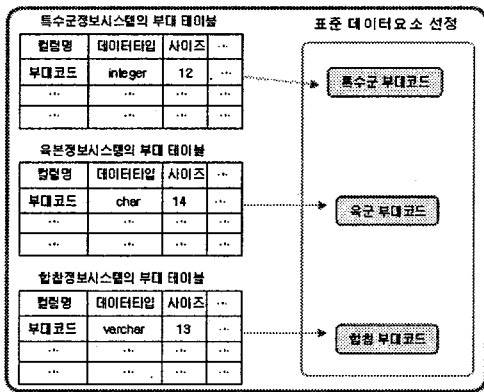
적(시스템의 관점)인 양방향의 표준화 절차에 의해 데이터품질이 유지되고 향상되어질 수 있다. 그러나 이 단계에서도 아래에서 설명하는 문제점들이 내재되어 있으며, 이는 다음 성숙단계인 Level4 데이터아키텍처의 관리단계에서 해소되어질 수 있다.

3.2.3.1 고립된 표준화의 섬

대부분의 데이터표준화는 테이블명세서나 데이터베이스시스템의 시스템카탈로그, 또는 간혹 데이터모델에서 표준 데이터요소를 추출하여 선정하는 방식을 취하고 있다. 즉 물리적 스키마에서 표준데이터요소를 도출해내는 Bottom-Up 방식을 이용하고 있다.

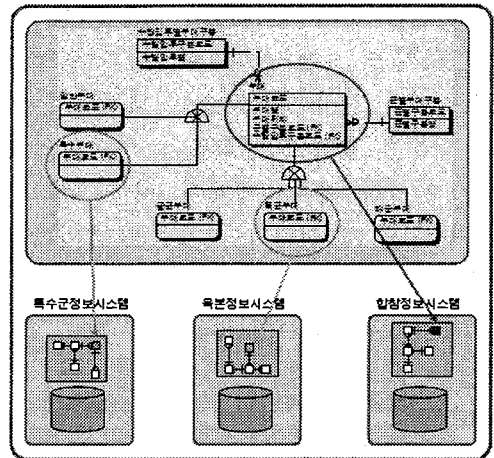
이러한 데이터표준화의 방식은 전사적 관점에서의 논리적, 구조적인 데이터체계에 대한 정의를 기반으로 하지 않고, 각 단위정보시스템의 물리적인 스키마(데이터베이스, 테이블명세서)에서 표준데이터요소를 선정함으로써, 전사적이고 통합적인 관점의 데이터표준화가 아닌, 부분적인 관점의 고립된 데이터표준화만 이루어지게 되는 결과를 낳고 있다. 이 수준의 데이터표준화에서는 전사적인 단위정보시스템별로는 데이터표준화가 이루어졌다고 볼 수 있지만, 전사적인 데이터공유와 통합관점의 데이터표준화가 완전히 이루어졌다고 보기는 어렵다.

〈그림 8〉은 전사적인 데이터체계를 고려하지 않고, 각 단위시스템만의 관점으로 표준화할 데이터를 선정하는 예를 도시한 그림이다. 부대코드라는 데이터요소를 각 정보시스템별로 특수군 부대코드, 육군부대코드, 합참부대코드라는 각 정보시스템의 독자적인 표준화 관점으로 표준화한 것을 보여주는 그림이다.



〈그림 8〉 단위시스템 관점의 표준화

그러나 실제로 부대코드는 〈그림 9〉와 같은 데이터체계를 가진 데이터요소이다. 즉 〈그림 9〉와 같은 전사적인 관점에서의 데이터체계에 대한 분석없이, 각각의 단위시스템에서 사용되고 있는 데이터의 관점으로 데이터를 표준화하게 되어, 각각의 표준화된 데이터가 연계성과 통합성을 상실한 고립된 표준화 데이터가 되는 현상을 가져올 수 있다.



〈그림 9〉 부대코드 데이터의 데이터체계

위의 예에서처럼 부대라는 Parent Entity가 각각의 부대유형에 따라 다양한 부대형태로 나누어 지고, 부대코드라는 PK가 상속된 것이 이 조직의 데이터체계이다. 그러나 이러한 전사적인 데이터체계에 대한 고려 없이 각각의 정보시스템 관점에서 사용되고 있는 부대코드, 즉 특수군부대코드, 육군부대코드, 합참부대코드라는 독자적인 개개의 관점으로 코드체계, 스키마, 값도메인 등이 표준화된다면 부대코드의 관점으로 전사적인 데이터연계는 불가능하게 될 것이다. 이러한 현상이 나타나는 이유는 Top-Down의 데이터체계분석을 수반하지 않고, Bottom-Up의 단위시스템에서의 표준화대상을 선정하는 방식으로 인해, 전사적인 체계의 관점이 결여되었기 때문이다.

이러한 고립된 데이터표준화의 문제를 해결하기 위해서는 전사적 관점에서

의 논리적, 구조적인 데이터체계에 대한 분석을 기반으로 하여, 각 단위정보시스템과 전사 데이터 아키텍처와의 연결성을 파악하고 관리해야 하며, 이러한 전사적 데이터체계에 대한 관리가 가능해질 때 전사적인 차원의 데이터공유와 통합이 가능한 진정한 의미의 데이터표준화가 가능하다.

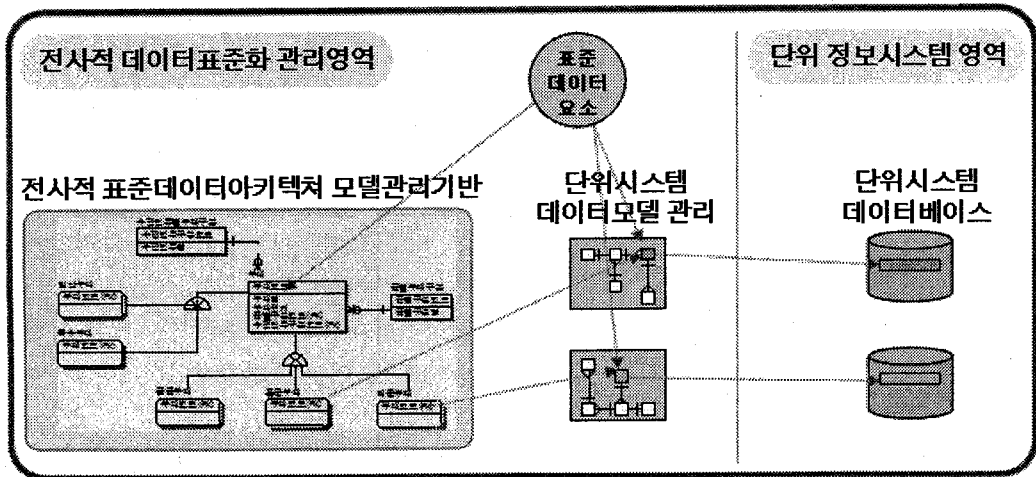
3.2.3.2 표준데이터의 모호한 정체성

데이터표준화 단계의 또 하나의 이슈사항은 표준데이터요소의 명확한 소속 엔티티를 규정하고 이를 기반으로 물리적인 단위정보시스템과의 연계성 관리를 하는 것에 대한 문제점이다.

즉 "표준데이터요소가 데이터모델의 어떤 Entity에 속하는가?", "표준데이터요소가 여러 Entity에 속한다면 이 표준데이터요소가 발생한 근본 Entity

는 무엇인가?", "표준데이터요소가 각 단위정보시스템에 어떻게 매핑되어 있는가?"라는 이슈에 적절히 대응하기 힘들다. 이는 물리적 스키마에서 Bottom Up 방식으로 선정된 표준데이터요소 관리방식으로는 표준데이터요소와 데이터모델과의 연결관리에 대한 문제점을 해결하는데 어려움이 있기 때문이다. 이러한 표준데이터요소의 명확한 정체성을 규정하기 위해서는 전사적 표준데이터아키텍처 분석이 수반되어야 한다. 즉 전사적 표준데이터아키텍처 모델을 통해 표준데이터요소와 데이터모델, 단위정보시스템간의 연결성을 명확히 규정할 수 있다.

<그림 10>에서 보여지듯이 전사적 표준데이터아키텍처를 기반으로 각 단위시스템의 논리데이터모델, 물리데이터모델과의 연결관계, 표준데이터요소와



<그림 10> 표준데이터아키텍처 기반의 데이터요소 연결관계 규정

데이터모델과의 연결관계를 명확히 규정할 수 있다. 데이터표준화의 이러한 각 요소들 간의 연결관계의 규정은 데이터의 변화관리를 적절히 수행해 나갈 수 있다는 점, 표준데이터요소가 어느 시스템에서 어떻게 사용되고 있는지를 명확히 파악할 수 있다는 점에서 매우 중요하다고 할 수 있다. 그러나 현실적으로 현재 이루어지고 있는 Bottom-Up 방식의 데이터표준화 접근방식으로는 이러한 연결관계의 규정에 많은 어려움을 가지고 있다.

3.2.3.3 표준데이터 관리주체의 모호성

데이터표준화 단계의 또 하나의 이슈 사항은 각각의 데이터의 관리주체에 대한 문제이다. 즉 데이터표준화 단계에서 해당 데이터에 대해 표준화를 위한 의사결정을 내릴 주체, 해당 데이터에 대해 유지관리를 수행할 주체에 대한 판단이 어렵다는 점이다. 데이터를 공유의 성격에 따라 분류하여 보면 <표

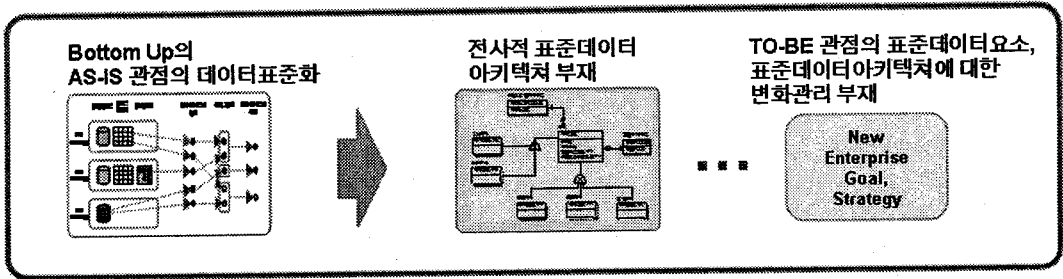
2>와 같이 나타낼 수 있다.

즉 <표 2>에서 보여지듯이 공유데이터와 공통데이터는 여러 조직에 걸쳐 나타나게 된다. 이 때 해당데이터의 생산/관리주체에 대한 명확한 파악이 없이는 데이터를 표준화하여 관리하기 힘들어진다. 왜냐하면 기본적으로 데이터의 표준화는 데이터 생성주체의 관점에서 데이터표준화안이 만들어져야 하기 때문이다. 즉 앞서의 부대코드의 예를 들면, 부대코드를 표준화할 때 어느 조직의 부대코드 표준화안을 따라야 할 것인가?, 아니면 각각의 표준화안을 모두 수용해서 각각의 표준부대코드를 만들어야 할 것인가라는 문제점에 부딪히게 된다. 이러한 데이터생성 및 관리주체의 모호성을 해결하기 위해서는 두가지 관점에서의 접근이 필요하다.

첫번째는 앞서 제시한 전사적 표준데이터아키텍처를 기반으로 하여 구조적으로 데이터요소를 파악해야 한다는 것이다. 앞서 설명하였듯이 데이터의 구

<표 2> 데이터 공유성격에 따른 분류

구 분	설 명
공통으로 나타나는 데이터	공유(Shared) 데이터 : 특정 조직에서 생산되었지만 여러 조직에서 공유하여 사용되며 여러 조직에서 관리되는 데이터
	공통(Universal) 데이터 : 전 조직에 걸쳐 참조 대상이 되는 데이터로 데이터 생산주체가 불명확한 일반성의 성격을 가진 데이터
한조직에서만 나타나는 데이터	고유(Unique) 데이터 : 한 조직에서만 사용되고 관리되는 조직 내의 고유한 데이터



〈그림 11〉 표준데이터 변화관리의 어려움

조를 체계적으로 분석해 넘으로써 해당 데이터의 원 소속지를 명확히 파악할 수 있고, 현재 사용되고 있는 데이터 요소가 어떠한 관점으로 분류되었는지를 확인할 수 있기 때문이다.

두번째는 데이터의 기능관점의 분석을 통해 데이터의 생산과 관리의 주체를 파악해야 한다는 것이다. 예를 들면 DFD를 통해 어떠한 External Agent를 통해 데이터가 생산되고 유지되는지를 파악해내야 한다.

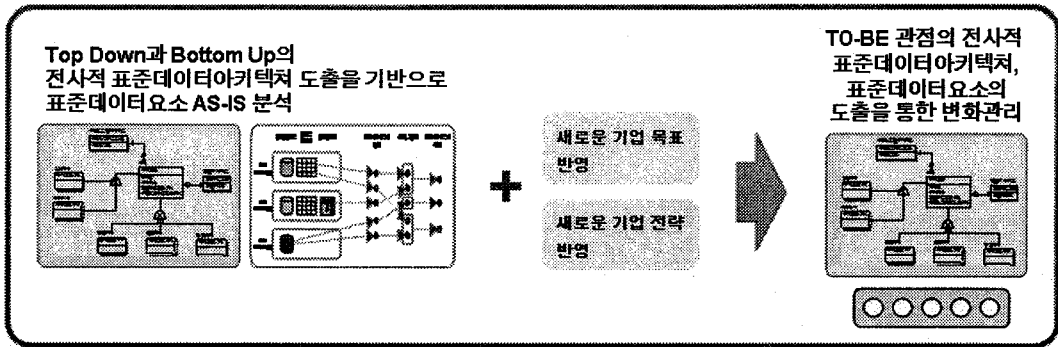
즉, 데이터생성 및 관리주체의 명확한 파악을 통해 데이터표준화 주체를 밝혀내기 위해서는 구조적 관점의 데이터체계에 대한 분석과 기능적 관점의 데이터생성과 흐름에 대한 분석이 동시에 수반되어야 한다.

3.2.3.4 표준데이터 변화관리의 어려움

기업은 항상 변화하게 된다. 변화하는 기업환경에서 새로운 기업목표와 전략에 맞는 새로운 비즈니스 규칙을 만족하는 데이터관리 전략이 필수적으로 필요하게 된다. 그러나 현재 이루어지

고 있는 데이터표준화의 단계에서는 이러한 변화하는 TO-BE 관점의 표준데이터에 대한 변화관리에 어려움을 가지고 있다. 이러한 문제점은 근본적으로 표준데이터아키텍처 관리의 부재에 기인한다. 즉 〈그림 11〉에서 보여지듯이 Bottom-Up의 현상적 데이터표준화를 수행하였기 때문에, 데이터의 전사적체계 관점의 이미지가 관리되지 않아 변화하는 데이터환경에 맞는 변화관리가 어렵기 때문이다.

이러한 문제점을 해결하기 위해서는 〈그림 12〉에서 보여지듯이 전사적 표준데이터아키텍처를 기반으로 하여 데이터표준화가 수행되어야 한다. 즉 전사적 표준데이터아키텍처를 기반으로 기업의 새로운 비즈니스 전략을 수용할 수 있도록 데이터를 변화관리 하여야 한다. 이러한 전사적 표준데이터아키텍처를 관리함으로써 데이터의 변화가 발생하더라도 전사적 체계의 관점에서 적절히 변경사항을 반영해 나갈 수 있다.



〈그림 12〉 전사적 표준데이터아키텍처를 기반으로한 데이터 변화관리

3.2.4 Level4로의 이행전략

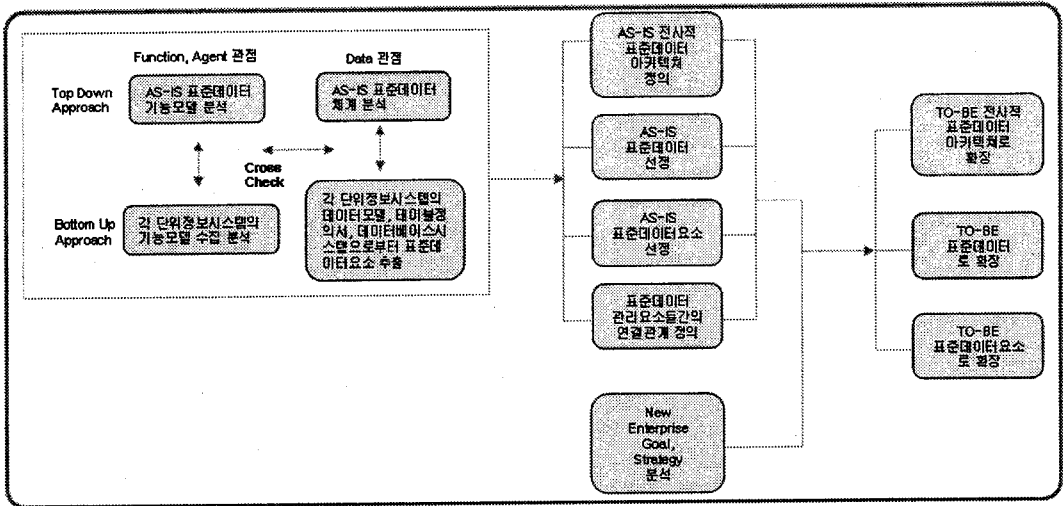
진정한 의미의 전사적, 통합적, 공유 가능한 표준데이터요소를 관리를 통한 데이터의 구조적 품질수준 향상은 전사적인 표준데이터아키텍처 모델을 기반으로 할 때 가능하며, 전사 표준데이터아키텍처 모델, 표준데이터요소, 단위 정보시스템의 데이터모델에 대한 정보 관리와 연결관계가 관리되어질 때 가능하다. 즉 Level 4 표준데이터아키텍처의 관리 단계에서는 위에서 제시한 데이터의 구조적 요소들이 유기적으로 관리되는 단계로써 앞서 제시된 Level 3 데이터표준화의 이슈사항에 대한 해결 요소들이 반영된 단계이다.

전사적인 데이터아키텍처 수립을 위해서는 기존의 데이터표준화 방법인 Bottom Up의 Reverse Engineering의 접근과 Top Down의 전사적인 표준데이터체계 분석이 동시에 수행되어야 한다.

또한 데이터요소를 생산하고 관리하는 주체를 정확히 파악하기 위해서는 기능적 관점에서의 데이터분석이 같이 수행되어야 한다. 그리고 기업의 새로운 목표, 전략의 관점의 분석을 통해 새로운 TO-BE 표준데이터아키텍처로 확장해나가야 한다. 이러한 관점을 반영한 Level 4의 데이터관리를 위한 접근방법론을 도출하면 〈그림 13〉과 같이 나타낼 수 있다.

〈그림 13〉에서 보여지듯이 데이터표준화와 전사적 표준데이터아키텍처 수립을 통한 데이터품질관리를 이루어내기 위해서는 상하향식의 접근법이 동시에 고려되어야 한다. 즉 Bottom-Up의 접근방법을 통해 현정보시스템에서 표준화대상이 되는 데이터요소를 수집하는 것과 동시에 Top-Down의 전사적 데이터구조에 대한 분석도 서로 보완적인 관계에서 수행되어야 한다.

또한 이러한 양방향적 접근은 데이터에 대한 구조적 분석뿐만 아니라 기능



〈그림 13〉 표준데이터아키텍처 수립을 위한 상하향식 접근법

적 분석이 같이 수행되어야 한다. 앞서 언급하였듯이 데이터를 생산하고 관리해나갈 주체에 대한 명확한 규정 없이는 데이터표준화와 이의 관리가 어렵기 때문이다.

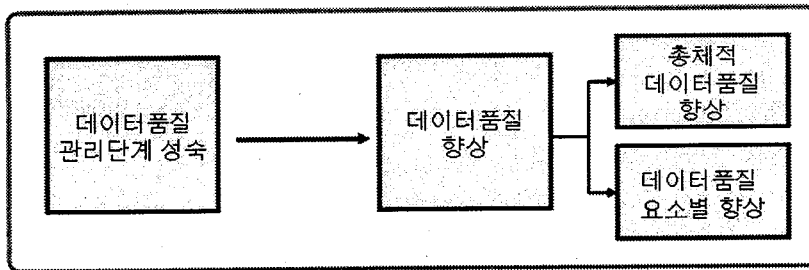
이러한 상하향식, 구조적·기능적 관점을 통해 현 기업의 표준데이터체계 및 데이터요소를 선정해내고, 이를 기반으로 새로운 기업환경에 맞는 TO-BE의 데이터관리로 변화할 수 있도록 확장해 나가야 한다.

4 데이터품질관리

성숙도모델의 실증적 검증

4.1 검증모형 및 검증방법

앞 장에서는 데이터품질관리 성숙도 모델을 제시하고 상위 성숙단계로의 이행을 통해 하위성숙단계의 데이터관리 이슈 해결방안을 사례 및 이론을 통해 제시하였다. 이를 통해 본 연구에서 제



〈그림 14〉 데이터품질관리 성숙도모델의 실증적 검증모형

〈표 3〉 설문지구성 요약

업 종	그룹A : 성숙도단계2인 기업 (성숙도단계1에서 단계2로 발전한 기업)	그룹B : 성숙도단계3인 기업 (성숙도단계2에서 단계3으로 발전한 기업)
설문대상기업수	3개사	3개사
설문대상직원수	46명	73명
설문방법	데이터모델 관리를 통한 데이터관리를 구분기준으로 사전과 사후로 분류하여 데이터품질수준을 설문	메타데이터 관리를 통한 데이터관리를 구분기준으로 사전과 사후로 분류하여 데이터품질수준을 설문
척도형태	리커트5점 척도	
설문문항수	16개 문항에 대해 사전과 사후로 나누어 응답	

시한 데이터관리성숙도 모델을 사례와 이론적 관점에서 검증해 보았다. 본 장에서는 데이터관리 성숙도 모델을 실증적으로 검증해보고자 하며, 〈그림 14〉와 같은 실증적 연구 검증모형을 수립하여 성숙도모델에 대한 검증을 수행하였다.

즉, 연구에서 제시한 데이터관리의 각 단계에 속한 기업을 대상으로 데이터품질의 변화정도를 측정하여, 데이터관리의 단계가 높아질수록 데이터품질이 높아지는지를 실증적으로 검증하였다. 이와 같은 실증적 검증을 위해 두단계의 설문을 통해 기업의 데이터관리수준과 데이터품질의 변화를 측정하였다.

1차적으로 기업의 IT책임자와의 인터뷰설문을 통해 해당기업이 본 연구에서 제시한 데이터관리 성숙도 수준의 어느 단계에 속하는지를 파악하였다. 인터뷰설문의 내용은 각 성숙도단계의 핵심이행요소의 이행유무를 파악하는

내용으로 구성되었다.

설문대상기업은 금융 및 공공분야의 6개 기업을 대상으로 하였으며, 1차 인터뷰설문의 결과 Level1에서 현재의 Level2로 발전한 기업 3개사(그룹A)와 Level2에서 현재의 Level3으로 발전한 기업 3개사(그룹B)인 두 개의 그룹으로 분류할 수 있었다.

현재 국내에서는 Level3를 일부 대기업을 중심으로 시작하려는 단계이고, Level2 또는 Level1인 기업이 대부분인 상황이다. 그러므로 Level4인 기업은 현재 국내에 거의 전무한 실정으로 Level3에서 Level4로 발전한 기업은 본 연구에서 포함시킬 수 없었다.

이렇게 분류된 기업들의 정보화관련 부서 직원을 대상으로 성숙단계의 변화에 따른 데이터품질의 변화를 묻는 설문을 2차적으로 실시하였다.

예를들면 Level1에서 현재의 Level2로 발전한 기업에 대해서는 데이터모델

관리 이전의 데이터품질과 데이터모델을 관리하는 현재의 데이터 품질을 설문하였다. 또한 Level2에서 현재의 Level3으로 발전한 기업에 대해서는 메타데이터(표준데이터)관리 이전의 데이터 품질과 메타데이터(표준데이터)를 관리하는 현재의 데이터 품질을 설문하였다. 설문항목의 구성은 각 기업의 성숙단계에 관계없이 데이터 품질을 묻는 16개 항목을 동일하게 설문하였다. 설문의 대상은 1차 인터뷰설문에서 분류된 두 개의 성숙단계 그룹별로 그룹 A(Level1에서 Level2로 발전한 3개사) 46명, 그룹B(Level2에서 Level3으로 발전한 3개사) 73명, 총 119명을 대상으로 하였다. 이러한 설문과정을 통해 수집된 데이터를 이용해 Reliability Test, Factor Analysis와 Paired Sample Test를 수행하였다.

첫 번째로 Reliability Test를 수행하여 설문척도의 신뢰성을 확인하였다.

두 번째로 데이터 품질을 묻는 설문항목을 Factor Analysis를 통해 요인별로 분류해 봄으로써, 데이터 품질을 묻는 설문항목 구성타당도를 검증하고 데이터 품질을 요인별로 분류하였다.

세 번째로 Paired Sample T-Test를 통해 각 성숙단계별로 데이터 품질에 차이가 있는지를 검증하였다. 데이터 품질 평균의 비교를 위해 Paired Sample Test를 사용한 이유는 동일한 모집단에 대해 사전과 사후의 데이터 품질을 비교

하기 위해서이다. 즉 본 연구의 설문은 동일기업에 대해 현재의 데이터관리단계에서의 데이터 품질과 이전의 데이터관리단계에서의 데이터 품질을 비교하는 내용으로 구성되었기 때문이다.

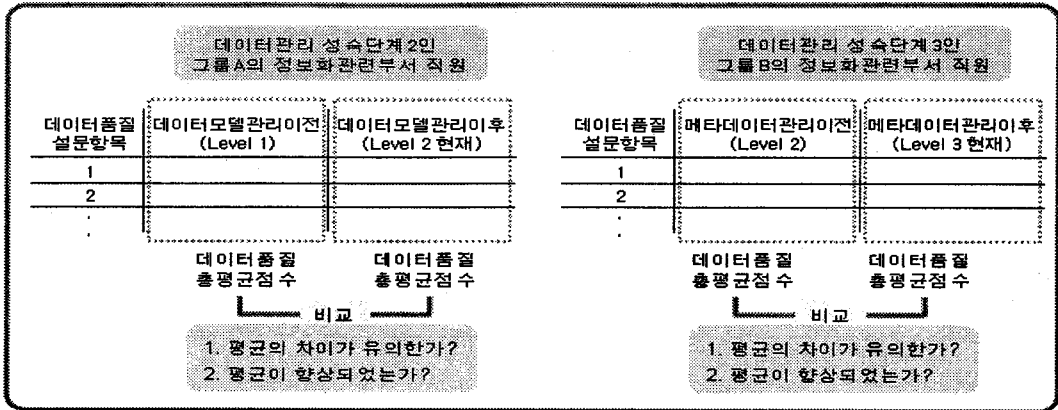
데이터관리 단계별 데이터 품질 평균의 비교는 크게 세가지 관점으로 이루어졌다.

첫 번째 관점은 데이터 품질을 묻는 16개의 설문항목 전체의 사전과 사후의 데이터 품질의 총 평균을 비교하였다.

즉 그룹A(Level1→Level2)에서 데이터 품질의 전체적인 향상이 있는지, 그룹B(Level2→Level3)에서 데이터 품질의 전체적인 향상이 있는지를 검증해 봄으로써, 본 연구에서 제시한 데이터 품질관리 성숙도모델의 타당성을 검증해 보았다.

두 번째 관점은 Factor Analysis를 통해 그룹화 된 데이터 품질 요인별로 사전과 사후의 데이터 품질 평균을 비교하였다. 즉 그룹A(Level1→Level2)에서 데이터 품질의 각 요인별로 향상이 있는지, 그룹B(Level2→Level3)에서 데이터 품질의 요인별로 향상이 있는지를 검증해 봄으로써, 본 연구에서 제시한 데이터관리단계의 향상이 데이터 품질의 모든 요인에 향상효과를 가져오는지를 검증해 보았다.

세 번째 관점은 그룹A와 그룹B의 전사적 연계와 통합관점의 데이터 품질 수준 향상의 차이를 분석하였다.



〈그림 15〉 성숙단계별 데이터품질의 향상효과 비교

앞 장의 데이터관리 성숙도모델의 설명에서 언급하였듯이, Level3부터는 전사적인 연계와 통합관점의 데이터품질향상의 효과가 증점적으로 나타나게 된다. 그러므로 그룹A(Level1에서 Level2로 성숙된 기업)와 그룹B(Level2에서 Level3으로 이행된 기업)의 전사적 연계와 통합성 관점의 데이터품질요인이 향상된 비중을 비교해 보았다. 이를 통해 데이터관리 성숙도모델에서 제시한 Level3부터의 전사적 연계와 통합관점의 데이터품질향상의 증점적인 확산효과를 검증해 보았다.

4.2 검증결과

4.2.1 설문구성의 신뢰도분석 결과

회수된 6개 기업의 총 119부의 설문 응답을 대상으로 신뢰도분석을 먼저 수행하였다. 사전과 사후의 데이터 품질을 복수로 응답하는 본 설문 특성상 데

이터품질에 대한 사전 응답과 사후 응답에 대해 각각 신뢰도 분석을 수행하였다. 신뢰도분석결과 사전응답과 사후응답의 Chronbach's Alpha계수가 각각 .918, .873으로 데이터품질에 대한 설문척도가 신뢰할만한 수준인 것으로 나타났다.

4.2.2 설문항목의 요인분석 결과

요인분석은 다음 단계에서 진행할 데이터품질에 대한 사전/사후의 차이비교를 위해 16개의 데이터품질측정항목을 요약하기 위한 목적으로 수행하였다. 요인추출 방법으로는 主成分分析(PCA: Principle Component Analysis) 방식을 이용하여 아이겐 값(Eigen value)이 1 이상이 되는 요인만 추출하였다.

추출된 요인의 패턴을 명확히 확인하기 위하여 직각회전 방식인 Equamax 방식을 이용하여 요인들을 구조화 하였다.

〈표 4〉 요인분석표

설문항목 요약	요인 1	요인 2	요인 3	요인 4
	전사적 연계 및 통합성품질	정확성품질	표현성품질	서비스품질
정보시스템간 데이터형식의 일관성	.843			
정보시스템간 코드값 도메인의 일관성	.739			
정보시스템간 데이터명칭의 일관성	.695			
정보시스템간 명명규칙의 일관성	.628			
정보시스템간 데이터의 비중복성	.598			
데이터값의 누락여부		.787		
데이터항목의 누락여부		.716		
데이터의 정확성에 대한 신뢰		.555		
데이터의 오탈자여부		.549		
데이터값의 객관성		.511		
데이터표현형식의 적절성			.819	
데이터표현형식의 동일성			.744	
데이터요약표현의 적절성			.502	
데이터 권한 및 보안의 적절성				.787
데이터의 최신성				.677
데이터조회 편의성				.570

〈표 4〉는 회전된 요인구조표이다. 표에서 보여지듯이 추출된 4가지 요인들을 구성하는 각 설문항목의 내용과 기존 데이터품질에 대한 선행연구들에서 제시된 데이터품질요인의 공통적 개념

을 이용하여 아래와 같이 요인의 명칭을 부여하였다. 요인1은 본 연구에서 제시하는 전사적 연계와 통합관점의 데이터품질로 요약해 볼 수 있으며, 요인2는 데이터의 정확성에 대한 품질, 요

〈표 5〉 연구가설 1

[가설 1] 데이터품질관리 성숙단계가 높아질수록 데이터품질수준은 향상될 것이다.
- 세부가설
[가설 1-1] 데이터관리 성숙단계가 높아질수록 전체적인 데이터품질 수준은 향상될 것이다.
[가설 1-2] 데이터관리 성숙단계가 높아질수록 데이터의 전사적 연계 및 통합의 품질수준은 향상될 것이다.
[가설 1-3] 데이터관리 성숙단계가 높아질수록 데이터정확성의 품질수준은 향상될 것이다.
[가설 1-4] 데이터관리 성숙단계가 높아질수록 데이터표현성의 품질수준은 향상될 것이다.
[가설 1-5] 데이터관리 성숙단계가 높아질수록 데이터서비스의 품질수준은 향상될 것이다.

인3은 데이터의 표현성에 대한 품질, 요인4는 데이터의 서비스에 대한 품질로 요약할 수 있다.

4.2.3 성숙단계별 데이터품질의 차이 비교 분석 결과

Paired Sample Test는 본 연구에서 제시한 데이터관리 성숙도모델을 실증적으로 검증하기 위해 수행되었다. 즉 본 연구에서 제시한 데이터관리의 단계가 성숙될수록 데이터품질수준이 향상되는가를 검증하고자 하였다. 이를 위해 본 연구에서 제시한 데이터관리의 각 단계에 속한 기업들의 정보화관련부서 직원들에 대해 이전단계였을 때의 데이터품질과 현재단계의 데이터품질에 대해 각각 품질수준을 평가하게 하였다. 설문응답자에게 이전과 현재단계를 구분하게 하는 기준으로써 Level1에서 현재의 Level2로 발전한 기업에 대해서는 Level2의 핵심이행요소인 데이터모델을 통한 데이터관리를 구분기준으로 제시하였고, Level2에서 현재의 Level3으로 발전한 기업에 대해서는

Level3의 핵심이행요소인 메타데이터(표준데이터)관리를 통한 데이터관리를 구분기준으로 제시하였다.

이러한 설문응답을 바탕으로 성숙단계모델의 검증을 위해 수립된 〈표 5〉의 연구가설 1에 대해 Paired Sample Test를 수행하였다. 평균차이의 검증을 위해 Paired Sample Test를 사용한 이유는 앞서 설명하였듯이 동일모집단에 대해 사전과 사후의 데이터품질의 차이가 유의한지를 검증하기 위해서이다.

세부가설 1-1의 전체적인 데이터품질이란 16개의 데이터품질설문항목의 전체평균값을 말하며, 세부가설 1-2부터 1-5에 사용된 각각의 데이터품질항목은 앞서 16개의 설문항목을 요인 분석한 결과 그룹화되어 제시된 4개의 요인들이다. 〈표 6〉에서 보여지듯이, Level1과 Level2의 전체적인 데이터품질의 평균차이에 대한 P-value는 .000으로 나타나 평균의 차이가 없다는 귀무가설은 $\alpha = .05$ 에서 기각된다. 즉, Level1과 Level2의 전체적인 데이터품질은 차이가 있으며, 〈표 7〉에서 보여지듯이

〈표 6〉 Paired Sample Test 결과표

		평균	t	자유도	유의확률 (양쪽)
데이터관리 성숙단계2인 기업(데이터모델 관리를 통해 데이터를 관리하는 기업)	단계1: 전체데이터품질 -단계2: 전체데이터품질	-1.00	-12.0	45	.000
	단계1: 데이터의전사적 연계 및 통합의품질 -단계2: 데이터의전사적 연계 및 통합의품질	-1.11	-10.7	45	.000
	단계1: 데이터정확성품질 -단계2: 데이터정확성품질	-.90	-10.0	45	.000
	단계1: 데이터표현성품질 -단계2: 데이터표현성품질	-.97	-8.6	45	.000
	단계1: 데이터서비스품질 -단계2: 데이터서비스품질	-1.02	-10.2	45	.000
데이터관리 성숙단계3인 기업(메타데이터 관리를 통해 데이터를 관리하는 기업)	단계2: 전체데이터품질 -단계3: 전체데이터품질	-1.48	-19.0	72	.000
	단계2: 데이터의전사적 연계 및 통합의품질 -단계3: 데이터의전사적 연계 및 통합의품질	-1.70	-18.0	72	.000
	단계2: 데이터정확성품질 -단계3: 데이터정확성품질	-1.26	-13.5	72	.000
	단계2: 데이터표현성품질 -단계3: 데이터표현성품질	-1.56	-17.6	72	.000
	단계2: 데이터서비스품질 -단계3: 데이터서비스품질	-1.40	-14.2	72	.000

Level1의 전체 데이터품질평균값은 2.59, Level2의 전체 데이터품질평균값은 3.60으로 데이터관리 단계가 Level1에서 Level2로 성숙될 때 전체적인 데이터 품질이 향상되었다는 것을 알 수 있다.

마찬가지로 Level2와 Level3의 전체적인 데이터품질의 평균차이에 대한 P-value는 .000으로 나타나 평균의 차이가 없다는 귀무가설은 $\alpha = .05$ 에서 기각된다. 즉, Level2와 Level3의 전

체적인 데이터품질은 차이가 있으며, Level2의 전체 데이터품질평균값은 2.54, Level3의 전체 데이터품질평균값은 4.02로 데이터관리 단계가 Level2에서 Level3로 성숙될 때 전체적인 데이터 품질이 향상되었다는 것을 알 수 있다.

위의 결과를 종합해볼 때 데이터관리의 단계가 성숙될수록 전체적인 데이터 품질수준은 향상될 것이라는 연구가설 1-1이 채택되었다.

<표 7> Paired Sample 통계

		평균	표준 편차	N	표준오차 평균	
데이터관리 성숙단계2인 기업(데이터모델 관리를 통해 데이터를 관리하는 기업)	전체데이터품질	Level1	2.59	.46	46	.067
		Level2	3.60	.37	46	.054
	데이터의전사적 연계 및 통합의품질	Level1	2.38	.47	46	.069
		Level2	3.49	.55	46	.081
	데이터정확성품질	Level1	2.74	.58	46	.085
		Level2	3.65	.45	46	.066
	데이터표현성품질	Level1	2.62	.67	46	.099
		Level2	3.62	.46	46	.068
	데이터서비스품질	Level1	2.63	.59	46	.087
		Level2	3.66	.49	46	.072
데이터관리 성숙단계3인 기업(메타데이터 관리를 통해 데이터를 관리하는 기업)	전체 데이터품질	Level2	2.54	.60	73	.070
		Level3	4.02	.35	73	.040
	데이터의전사적 연계 및 통합의품질	Level2	2.32	.63	73	.073
		Level3	4.02	.51	73	.060
	데이터정확성품질	Level2	2.69	.76	73	.089
		Level3	3.95	.41	73	.048
	데이터표현성품질	Level2	2.51	.70	73	.082
		Level3	4.07	.46	73	.054
	데이터서비스품질	Level2	2.67	.69	73	.080
		Level3	4.07	.45	73	.052

<표 8> 연구가설 2

[가설 2] 전사적 연계와 통합성 관점의 데이터품질은 Level1에서 Level2로 성숙될 때보다 Level2에서 Level3으로 성숙될 때 더욱 향상될 것이다.

나머지 데이터의 전사적 연계와 통합의 품질요인, 데이터의 정확성 품질요인, 데이터의 표현성 품질요인, 데이터의 서비스 품질요인에 대해서도 Level1과 Level2, Level2와 Level3의 품질평균의 차이에 대한 P-value가 모두 .000

으로 나타나 데이터관리의 단계가 성숙될수록 각 데이터품질요소의 차이가 있는 것으로 판명되었다. 또한 <표 7>에서 나타났듯이 각 데이터품질요인의 평균이 Level1보다는 Level2가 Level2보다는 Level3이 높게 나타나 데이터

관리의 단계가 성숙될수록 각 데이터품질요소의 수준은 향상될 것이라는 연구가설 1-2, 1-3, 1-4, 1-5가 모두 채택되었다.

본 연구에서 제시한 데이터관리 성숙도모델에서 Level2와 Level3을 경계로 하여 데이터의 전사적 연계와 통합성이 보다 높아진다(연구가설 2)고 주장하였다.

앞서의 Paired Sample Test의 결과를 분석해 보면 Level1에서 Level2로 성숙될 때 전사적 연계와 통합성의 데이터품질의 향상정도는 1.11로 향상정도가 두 번째인 데이터서비스품질의 1.02과 0.09의 격차, 최소 데이터품질 향상요인인 데이터의 정확성품질의 0.90과 0.21의 격차를 보인다.

이에반해 Level2에서 Level3으로 성숙될 때 전사적 연계와 통합성의 데이터품질의 향상정도는 1.70으로 향상정도가 두 번째인 데이터표현성품질의 1.56과 0.14의 격차, 최소 데이터품질 향상요인인 데이터의 정확성품질의 1.26과 0.44의 격차를 보인다. 즉 Level1에서 Level2로 성숙될 때보다 Level2에서 Level3으로 성숙될 때 전사적 연계와 통합성의 데이터품질이 다른 품질요인보다 향상정도가 높다고 인지되고 있는 것을 확인할 수 있다.

이상과 같은 실증적 분석을 통해 본 연구에서 제시한 데이터관리 성숙도모델을 검증해보았으며, 실제로 유효하다

는 것을 확인할 수 있었다.

5 결 론

본 연구에서는 데이터 품질을 평가하고 관리해나가기 위한 데이터품질관리 성숙도모델을 제시하였다. 본 연구에서 제시된 데이터품질관리 성숙도모델의 의의는 다음과 같이 요약할 수 있다.

첫째, 기존 데이터품질관리 연구의 결과적 평가관점의 모델에서 데이터품질의 향상을 위한 단계적 성숙도모델을 제시함으로써, 기업의 현재 데이터품질관리의 수준에 대한 파악과 발전을 위한 이행요건을 제시하였다.

둘째, 데이터품질관리를 위한 구조적 품질관리의 단계를 실증적인 사례를 통해 제시함으로써, 기존의 현상적 품질평가방법인 이론적인 품질평가의 개념을 현실적인 평가와 관리의 개념으로 발전시킬 수 있는 기반을 마련하였다.

셋째, 현재 몇몇 대기업 및 공공기관에서 시도되고 있는 데이터표준화단계에서의 이슈사항을 제시하고 이를 해결해나갈 수 있는 방안을 제시함으로써 데이터표준화를 통한 데이터품질관리의 지침을 마련하였다.

넷째, 현재의 관점만이 반영되어 평가되던 데이터품질에 대한 관점을 미래의 관점을 반영하여 평가하고 관리할 수 있는 토대를 마련하였다.

데이터품질관리는 품질평가를 위한 세부적인 절차와 방법도 중요하지만, 기업에 있어서 현재 자신의 데이터품질 관리의 수준이 어느 위치에 있는지를 파악하고 다음단계로 나아갈 수 있는 단계적 관리지침의 제시도 중요하다. 본 연구에서는 이러한 데이터품질의 단계적 성숙모델을 제시함으로써, 거시적인 데이터품질관리의 관점을 제시하였는데 가장 큰 의의가 있다.

참 고 문 헌

- Antonis, C. Stylianou and Ram L. Kumar, 2000. "An Integrative Framework For IS Quality Management." *Communications of the ACM*, 43(9): 99-104.
- Beverly, K. Kahn, Diane M. Strong, and Richard Y. Wang, 2002. "Information Quality Benchmark." *Communications of the ACM*, 45(4): 184-192.
- Diane, M. Strong, Yang W. Lee, and Richard Y. Wang, 1997. "Data Quality in Context." *Communications of the ACM*, 40(5): 103-110.
- Giri, Kumar Tayi and Donald P. Ballou, 1999. "Enhancing Data Quality In Data Warehouse Environments." *Communications of the ACM*, 42(1): 73-78.
- Giri, Kumar Tayi and Donald P. Ballou, 1998. "Examining Data Quality." *Communications of the ACM*, 41(2): 54-57.
- Leo, L. Pipino, Yang W. Lee, and Richard Y. Wang, 2002. "Data Quality Assessment." *Communications of the ACM*, 45(4): 211-218.
- Richard, Y. Wang, 1998. "A Product Perspective on Total Data Quality Management." *Communications of the ACM*, 41(2): 58-65.
- Thomas, C. Redman, 1998. "The Impact Of Poor Data Quality On The Typical Enterprise." *Communications of the ACM*, 41(2): 79-82.