

데이터 품질 향상을 위한 데이터 관리 프로세스 개선 사례 연구: 데이터 표준과 요구사항 관리 중심으로

A Case Study on Improvement of Data Management Process for Enhancing Data Quality: Focus on Data Standards and Requirement Management

허 희 정 (Hee Joung Heh)

(주)코이드 DB관리부

김 종 우 (Jong Woo Kim)

한양대학교 경영대학 경영학부 교수, 교신저자

요 약

최근 기업의 각 업무가 정보화되면서 부문별, 업무별 정보시스템의 데이터 간에 심각한 중복성과 불일치성의 문제가 대두되면서 데이터 품질관리에 관심이 집중되고 있다. 본 연구는 실제로 데이터 품질 관리 프로세스 개선을 통해 데이터 품질이 향상된 기업의 사례를 통하여 프로젝트 수행 과정에서의 주요 이슈와 위험요인을 살펴보고 그 해결방안을 제시함으로써 데이터 품질 향상을 위해 노력하는 타 기업들에게 도움을 주고자 하였다. 또한, 개선된 데이터 품질 관리 프로세스에 대한 다차원적인 평가로서 데이터 품질, 생산성, 고객만족도, 조직 및 문화의 측면에서 정성적이고 정량적인 지표를 통한 개선효과를 살펴보고 평가함으로써 제안된 프로세스에 의해 품질수준이 향상되었음을 검증하였고 평가 분석을 통한 시사점을 도출하였다.

키워드 : 데이터 품질, 데이터 관리 프로세스, 데이터 표준 관리, 요구사항 관리

I. 서 론

최근 기업의 각 업무가 정보화되면서 부문별, 업무별 정보시스템의 데이터 간에 심각한 중복성과 불일치성의 문제가 대두되면서 데이터 품질관리에 관심이 집중되고 있다(DPC, 2006c, 이춘열 외, 2004, Strong et al., 1997). 또한 정보 시스템 사용자의 업무와 의사결정을 효과적으

로 지원하기 위해 전사 차원에서 데이터를 통합한 전사 데이터 뱅크 및 데이터웨어하우스를 구축해 운영하는 기관이 증가하였으나, 구축한 데이터웨어하우스의 데이터 품질로 인해 시스템 운영에 많은 문제가 발생하고 있다(Ballon and Tayi, 1999). 뿐만 아니라 인터넷이 발전됨에 따라 기관이나 기업 내부의 시스템 통합뿐만 아니라 외부 시스템과의 연계가 필요하게 되면서 연계대상이 되는 외부 기관과 기업 간의 데이터 표준 및 데이터 일관성이 중요한 이슈가 되고 있다. 즉, 기업에서 운영하는 정보시스템의 범위

† 이 논문은 한양대학교 일반연구비 지원으로 연구되었음(HY-2007-G).

가 확대되고 복잡도가 증가함에 따라 정확성, 일관성, 적시성 등과 같은 데이터 품질이 중요한 이슈가 되기 시작하면서 이제 데이터 품질관리는 정보시스템 운영 및 관리의 필수 불가결한 요구사항이 되고 있는 실정이다.

낮은 품질의 데이터가 조직에 미치는 영향은 고객 불만족, 운영비의 증가, 비효과적인 의사결정, 그리고 전략의 수립 및 집행 능력의 저하 등을 들 수 있는데 더욱 민감한 문제는 구성원 사기저하, 조직의 불신증대, 그리고 조직의 단합을 어렵게 만들 수 있다는 것이다(Redman, 1998). 조직에 미치는 다른 영향을 보면 전략적인 이슈로는 일부 핵심 데이터만이 의사결정에 활용되고, 부정확한 데이터로 인해서 의사결정의 신속성이 낮아진다는 것이다. 운영적 이슈로는 부정확한 데이터를 개선하거나 조정하기 위한 현업과 IT담당자 간의 반복적인 데이터 확인 및 요청으로 시간의 소요와 비용이 발생하게 되는 것을 들 수 있다.

초기 데이터 품질에 대한 연구들은 주로 데이터 값의 품질(정확성, 최신성, 포괄성 등), 데이터 서비스의 품질(검색성, 편의성 등) 등과 같은 데이터의 현상적 품질에 대한 연구가 주종을 이루어졌다. 이후 데이터의 현상적 품질들을 결과적인 품질로서 인식하고, 원인적 요인에 해당하는 데이터 설계 구조와 같은 구조적 품질에 관심을 가지게 되었다. 최근에는 기존의 데이터 값, 데이터 서비스, 데이터의 구조뿐만 아니라 데이터 관리 프로세스 개념을 모두 포함하는 통합적인 데이터 품질관리에 대한 연구로 발전하고 있다. 데이터 품질에 대한 기존의 선행 연구들은 주로 데이터 품질 평가에 대한 연구들로 집중되고 있으나, 통합적 데이터 품질관리를 위해서는 평가뿐만 아니라 지속적인 품질수준의 유지와 개선의 관점이 필요하다(김찬수 외, 2003).

본 연구에서는 실제로 데이터 품질 관리 프로세스 개선을 통해 데이터 품질이 향상된 기업의

사례를 통하여 프로젝트 수행 과정에서의 주요 이슈와 위험요인을 살펴보고 그 해결방안을 제시함으로써 데이터 품질을 향상시키기 위해 노력하는 타 기업들에게 도움을 주고자 한다. 또한 개선된 데이터 품질 관리 프로세스에 대해 데이터 품질, 생산성, 고객만족도, 조직 및 문화의 측면에서 정성적이고 정량적인 지표를 통한 측정 및 평가를 수행함으로써 데이터 품질 향상의 유효성을 검증해 보도록 한다.

본 논문의 구성은 다음과 같다. 제 II장에서는 데이터 품질과 데이터 관리 프로세스에 관한 선행연구를 분석하였다. 제 III장에서는 관련연구를 기반으로 데이터 표준 관리와 요구사항 관리에 중점을 둔 개선된 데이터 관리 프로세스를 적용사례를 통해 제시하고 사례 프로젝트 수행 과정에서의 주요 이슈 및 위험요인 분석 및 해결방안을 제시하였다. 제 IV장에서는 개선된 프로세스에 대해 데이터 품질, 생산성, 고객만족도, 조직 및 문화의 측면에서 정성적, 정량적 측정 지표에 근거한 개선도를 평가하여 본 연구의 개선안에 의해 성능이 향상됨을 보이고 평가 분석을 통한 시사점을 도출하였다. 제 V장은 본 연구의 결론부문에 데이터 품질 관리 방안에 관한 발전방향을 기술하였다.

II. 관련연구

2.1 데이터 품질

그 동안 연구된 데이터 품질의 정의를 살펴보면, Wang *et al.*(1996)은 데이터 사용자에게 의해 사용하기 적합한(*fit for use*) 데이터로 정의하였고, Miller(1996)는 데이터 사용자에게 의해 데이터가 인식되고 사용되는 방법이라고 정의하였다. Kelly(1997)는 데이터 품질은 사용자에게 의해 정의되는 것으로서 비즈니스 목적에 대한 적합성의 정도로서 측정되는 것이라 하였고, Redman(1998)은 데이터 품질은 데이터 사용자의 요

구사항을 충족하는 정도라고 하였다. Orr(1998)는 데이터 품질은 정보 시스템과 실제 시스템 간 얼마나 일치하는가를 측정하는 것이라고 하였으며, Brackett(2002)는 데이터 품질은 기업의 현재와 미래의 비즈니스 정보 요구에 어떻게 지원하는가를 측정하는 것이라고 정의 내리고 있다.

데이터 품질의 정의를 종합해 보면, 데이터간의 일치성 유지가 중요하고, 모든 데이터가 완전무결할 필요는 없으며, 소비자의 요구에 부합하는 수준이면 된다고 할 수 있다. 또, 데이터가 자신의 업무를 수행하는데 얼마나 도움이 되는지를 근거로, 정보의 최종 소비자가 데이터 품질이 만족스러운지 아닌지를 평가한다. 또한, 데이터 품질의 실제 목표는 사용자와 이해관계자의 만족을 증대 시키는 것이며 최종 사용자의 만족도로 측정되어야 한다고 할 수 있다.

Wang *et al.*(1996)의 연구에서는 정확성만 보는 좁은 관점으로는 실제 사용에서 문제가 되는 형태나 구조 등 다른 영향을 포괄하지 못한다고 보고, 데이터 사용자(consumer)의 관점에서 데이터 품질에 관한 탐색적 연구를 통해 실제 사용자가 중요하게 생각하는 데이터 품질요인에 관한 연구를 수행하였다. 분류작업을 거쳐 15개의 품질요소를 본질적(intrinsic), 업무환경적(contextual), 표현적(representational), 접근성(accessibility)이라는 4개의 범주로 구분하여 데이터 품질요소의 개념적 프레임워크를 개발하였다. 제품의 품질 분야에서 소비자에 의해 강조되는 품질의 차원이 제조업자에 의해 강조되는 품질의 차원보다 더 폭 넓은 것과 마찬가지로 데이터 사용자들은 정보시스템 개발자보다 데이터 품질에 대해 더 넓은 개념을 가지고 있다. 즉 정보시스템 개발자들이 중요시 여기는 정확성, 객관성뿐만 아니라 신뢰성과 평판우수성 등도 포함되어야 한다고 언급하고 있다. 업무환경적(contextual) 속성이란 데이터 품질이 업무환경 내에서 고려되어야만 하는 요구사항을 의미하

지만 이것은 시간과 데이터 사용자에 따라 다르므로 높은 업무환경적 데이터 품질 수준을 달성하는 것은 하나의 도전(Strong, 1997)이라고 할 수 있다.

2.2 데이터 표준화

데이터 품질을 유지하고 개선하기 위해 수행하는 모든 활동을 데이터 품질관리 활동이라고 하는데 품질관리 활동의 중요한 요소 중의 하나가 데이터 표준화이다(Winkler, 2004). 정보통신 시스템의 통합과, 정보공유의 가장 큰 걸림돌은 시스템적인 통합 메커니즘의 개발뿐만 아니라, 데이터의 의미(semantic), 구문(syntax), 표현(representation)의 불일치를 해결해야 한다는 점이다. 이러한 정보공유를 위해서는 데이터의 표준화가 필연적이다. 데이터 표준화는 데이터 모델 Repository와 연계하여 모델의 용어 및 속성을 검증, 관리하며 용어의 표준, Naming Rule 적용관리를 위한 첫 단계가 되는 것이다. 이러한 정보의 공유가 가능하도록 표준화된 의미와 형태를 가진 정보의 요소를 표준 데이터 요소라 하며, 표준 데이터 요소는 자동화된 정보처리 시스템에서 사용될 수 있다. 데이터 표준화의 효과는 (1) 표준화 절차에 맞춘 업무 프로세스의 개선, (2) 전사데이터의 일관성, 통일성 확보, (3) 시스템 개발 생산성 향상, (4) 데이터 모델의 체계적인 관리 등이 있다(윤성호, 2003).

2.3 데이터 품질관리 프로세스

데이터 품질관리 프로세스는 고품질 데이터의 운영, 관리를 위한 데이터 관리 정책 수립에서부터 데이터 표준 관리, 데이터 모델 관리, 데이터 흐름 관리, DB 관리, DB 보안관리, 데이터 활용 관리, 요구사항 관리에 이르는 일련의 데이터 관리 활동들을 체계적인 흐름으로 표현한 프로세스이다(DPC, 2006c). 즉, 고품질의 데

이더를 지속적이고 안정적으로 서비스하기 위해 각 기업의 특성에 맞게 정의하고 필요한 활동간의 연관 관계를 정의한 프로세스를 의미한다. 정의된 데이터 품질관리 프로세스의 관리 기준은 데이터 관리 원칙에 맞게 정의되어야 하고, 각 기업의 기존 프로세스에 대한 특성을 고려하여 정의된 프로세스는 데이터와 관련된 모든 요소가 빠짐없이 관리될 수 있도록 해야 하며, 기존의 타 프로세스(변화 관리, 프로젝트 관리 등)와 상호 연관 관계가 명확하게 정의되어 있어, 적용함에 문제가 없어야 한다. 박종모(2006)의 연구에서는 데이터 관리 프로세스 개선을 위해서는 신규 시스템 또는 기존 시스템 개선 시 변화되고 생성되는 데이터 항목들에 대하여 기존 시스템과 연관성을 검토해야 하고, 전사 차원의 데이터 표준의 관리가 필요하다고 하였다.

특히, 데이터 표준관리 프로세스는 데이터의 구조적 설계, 즉 메타 데이터의 대한 전사적인 표준화를 통해 데이터의 구조적 품질을 높이고, 결과적으로 데이터의 연계 및 통합, 재사용의 품질을 높일 수 있는 현실적인 방안이 될 수 있다. 나아가 표준화된 데이터의 전사적 공통사용을 통해 균일화된 데이터 품질을 유도해 나갈 수 있으며, 데이터의 개발과 사용, 버전업에 대한 관리 프로세스를 포함함으로써 데이터 품질의 지속적인 유지 및 개선을 가능하게 하는 실제적인 방안이라는데 의의가 있다. 개발자에 의해 잘못 설계되기 쉬운 데이터의 구조 및 내용을 기능적, 기술적인 관점에서의 Cross-Check를 통해 검토, 관리함으로써 기술적 정확성과 업무관점의 기능적 정확성을 동시에 고려하여 관리한다는 점 또한 데이터 품질관리 프로세스에서 주목할 사항이다.

2.4 데이터 품질 평가

데이터 품질 평가는 데이터 품질 향상을 위해

서 필수적이다. 데이터 품질 평가 시 고려해야 하는 데이터 품질 차원에 대한 연구들은 Wang *et al.*(1996)의 연구, Pipino *et al.*(2002), Kahn *et al.*(2002) 등 활발하다. 예를 들어, 2.1절에서 언급한 Wang *et al.*(1996)의 연구에서는 4개 범주, 15개 품질요소를 제시하였으며, Pipino *et al.*(2002)의 연구에서는 접근성, 데이터 양의 적절성, 완전성, 일관성 등 16개 차원을 제시하였다. 국내에서는 한국데이터베이스진흥센터가 1998년 데이터베이스 품질 평가 지침, 2002년 데이터베이스 품질 평가 모델을 개발하였으며, 2003년 이를 확장한 모델을 제시하였고, 2006년 이를 개편하였다(DPC 2006c, 이춘열 외 2004). 한국데이터베이스진흥센터가 제시한 품질평가 모형에서는 데이터베이스 품질을 데이터 품질, 시스템 품질뿐만 아니라, 데이터 관리 프로세스 품질을 포함하도록 확장하였다. 즉, 데이터 값, 데이터 구조의 품질을 포함하는 데이터 자체의 품질과 이들 데이터를 다루는 정보시스템의 품질뿐만 아니라, 이들 데이터를 관리하는 프로세스의 품질을 포괄적으로 다루고 있다.

데이터 품질 평가에 대한 연구들을 종합해 보면, 데이터 사용자의 관점에서 데이터 품질은 다양한 요소들로 구성되어 있으며, 여러 가지 프레임워크를 통해서 이러한 내재 요소들간의 상호작용은 환경에 따라서 다르게 나타난다는 것을 알 수 있다(Pipino *et al.*, 2002). 하지만 품질의 측정기준에 있어서 품질을 평가만 하던 기존의 방식에서 벗어나 품질 저하의 요인을 분석하고 개선할 수 있는 방안을 수립하는 품질 평가 모델로 발전되고 있다. 즉, 데이터베이스 품질 평가모델의 변화흐름은 평가시각 및 평가대상의 다양화라고 특징 지워질 수 있다. 현상분석중심 평가모델들은 최종사용자에게 제공되는 데이터의 품질 및 정보서비스 평가로서 일회성으로 현상을 평가하였지만, 그 후에 구조적 평가모델들은 최종사용자뿐만 아니라 데이터베이스 관리자 및 설계자로 시각을 넓혔고, 개선을

위한 통합모델은 데이터 품질에 관련한 조직의 정책 수립자 및 운영자 등 평가 시각을 더욱 더 넓혀 지속적인 품질 개선을 도모하기 위해 평가 대상으로 조직 및 프로세스 정책 등 관련된 모든 요소를 포함한다. 이에 따라 평가 시각의 차이, 데이터의 특성이 고려된 명확하고 정형화된 데이터 품질 측정기준의 수립이 데이터 품질에 관련된 다양한 연구와 노력을 위해 선행적으로 이루어져야 한다.

III. 데이터 관리 프로세스 사례 분석

본 연구의 목적을 달성하기 위하여 사례분석 방법과 정책대안적 연구방법이 사용되었다. 실제로 데이터 품질 관리 프로세스 개선을 통해 데이터 품질이 향상된 국내의 한 기업을 사례 대상으로 선정하여 사례 기업의 대용량 고객 데이터베이스에 대해 데이터 품질 관리 프로세스 개선 전후에 대한 비교 분석을 중점적으로 검토하고 프로젝트 수행 과정에서의 주요 이슈에 대한 해결 방안과 시사점을 제시함으로써 데이터 품질 향상을 위해 노력하는 타 기업의 프로젝트에서 활용할 수 있도록 실무적 가치를 부여한 정책대안적 연구방법을 병행하여 사용하게 되었다.

3.1 사례 프로젝트의 이해

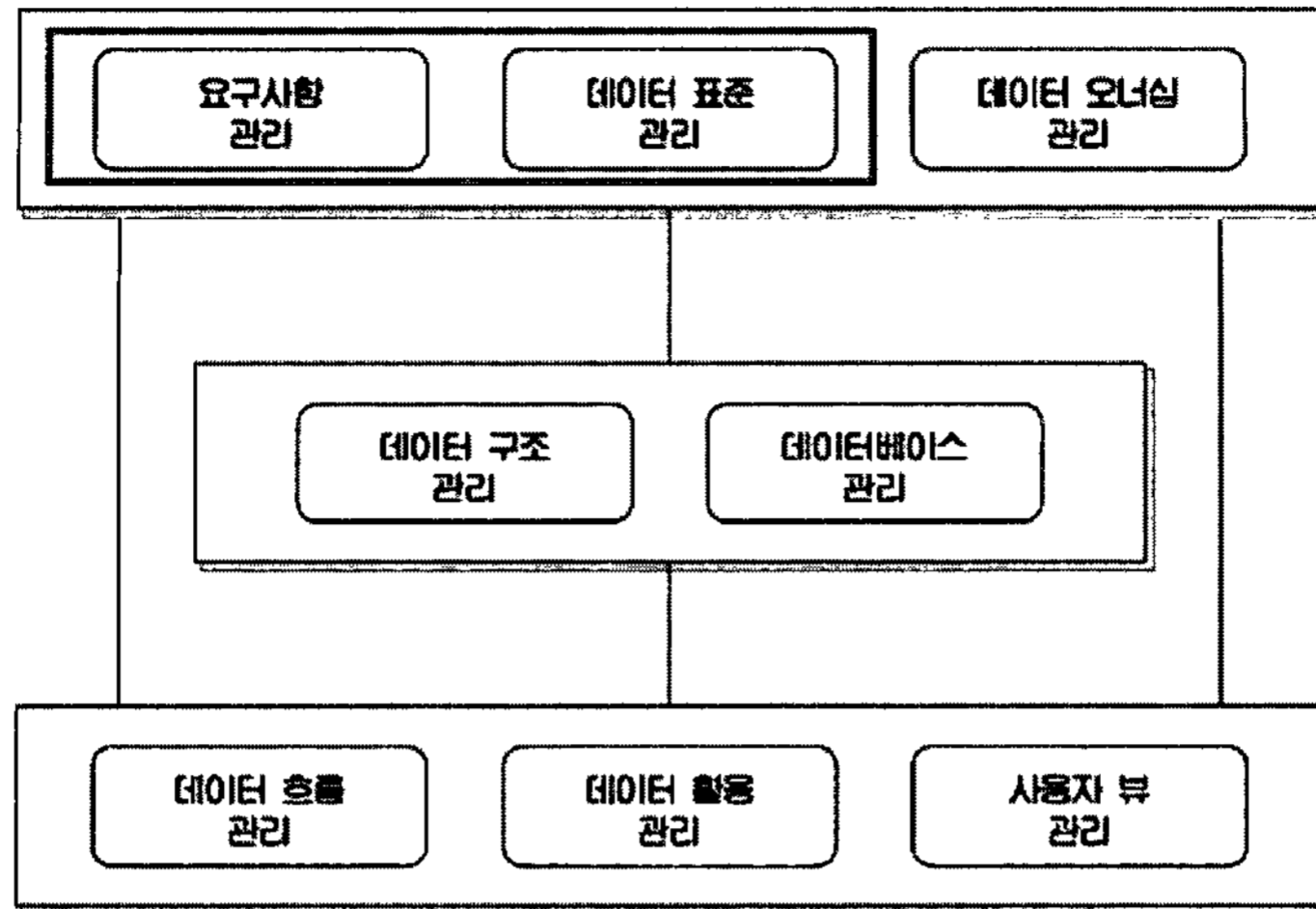
114종합정보 안내 서비스 기업인 (주)코이드(koid, 대표이사 박종수)는 KT에서 2001년 114번호 안내 서비스를 분사하여 설립된 텔레마케팅 전문 기업으로 기존 전화안내사업의 DB와 기술력을 바탕으로 114번호 안내 사업을 비롯하여 콜센터, 텔레매틱스, 인터넷 포털 서비스 등 다양한 사업을 통해 연평균 25% 이상의 고속 성장을 해온 기업이다. 코이드는 “한국서비스 품질 우수기업”과 “신 품질 혁신상”을 2회 연속 수상하는 등 대내외적으로도 공기업 구조조정의 성공적인 모델

사례로 꼽히고 있다. 70여년 동안 지속된 114번호 안내 서비스의 역량을 바탕으로 다양화된 생활정보 제공 사업을 추진하고 있으며 생활정보의 미래를 선도하는 글로벌 기업을 비전으로 삼고 있다.

기본서비스는 KT, 하나로, 데이콤 등 유선 통신 사업자의 가입자 전화번호를 대상으로 서비스 이용자가 일반전화 혹은 기타 통신수단을 통하여 전화번호를 문의시 해당 전화번호를 안내하는 서비스이다. 렛츠114 인터넷 포털서비스(www.lets114.co.kr)는 2,400만의 KT 인명, 상호 전화번호를 기반으로 전화번호 검색 및 지도 검색, 키워드 검색 등의 다양한 서비스로 인터넷 이용자의 생활편의를 도모하고 있다. 또한, 렛츠114 나비(www.lets114navi.co.kr) 텔레매틱스 서비스는 TDB(Telematics Database)를 이용하여 전화번호 목적지 검색, 시설물 검색, 상호 검색, 업종 검색을 통한 위치기반 서비스를 제공하고 있다.

사례 시스템인 NDMS(National Directory Management System)는 ICIS(Integrated Customer Information System, KT 통합고객정보시스템)에서 일 변동으로 제공 받은 고객 DB를 용도에 맞게 정제한 후 NDAS(전화번호 안내 DB), TDB(Telematics Database) 등의 수요처로 일 변동 자료를 중앙에서 배포하는 DB 서버이다. NDMS는 내부 업무 효율 향상과 고객 서비스 개선 및 기업 경쟁력 제고, 고객들의 정보 취득 및 전산화, 사업 패턴에 맞는 DB 관리 등의 목적이 있는 조직 내에서 핵심적인 역할을 수행하고 있는 시스템이다.

회사의 사업 다각화로 인하여 새로운 시스템이 증설되고 다양한 서비스가 제공되고 있으며, 고객 번호 DB에 대한 요구정보가 세분화되고 특화되는 추세로 새로운 성장 사업 발굴과 내부 역량 강화를 위한 데이터 품질 관리 프로세스 개선의 필요성이 대두되었다. 또한, 기존 NDMS 운용 S/W의 노후화에 따른 유지보수의 어려움



〈그림 1〉 데이터 품질 관리 프로세스

과 수작업 위주의 일 변동 처리로 인하여 데이터 정비 인력의 과다 소요, 데이터 처리의 부정확 사례 증가, 사업활용의 한계 등 제한적 환경을 극복하기 위하여 데이터 관리 프로세스 개선 프로젝트인 시스템 고도화를 추진하였다.

3.2 데이터 관리 프로세스 적용 사례

데이터 품질 문제의 해결을 위해서는 데이터 관리 프로세스의 개선이 필요하다고 판단하였다. 데이터 품질 향상의 주된 요소는 데이터 아키텍처가 품질을 결정하기 때문에 처음부터 정확한 아키텍처 확립에 주력해야 하고, 품질 관리의 대상인 산출물과 프로세스에 대해 명확한 관리를 해야 하며, 조직적인 측면에서 오류 예방과 품질 평가 그리고 개선활동을 수행해야 한다. 데이터 품질 관리 프로세스는 <그림 1>에서 보는 것과 같이 요구사항 관리, 데이터 표준 관리, 데이터 오너십 관리, 데이터 구조 관리, 데이터베이스 관리, 데이터 흐름 관리, 데이터 활용 관리, 사용자 뷰 관리의 8개 프로세스로 구분할 수 있다(DPC, 2006b). 본 논문에서는 이중 마스터 데이터베이스 관리에 필수적인 요소인 데이

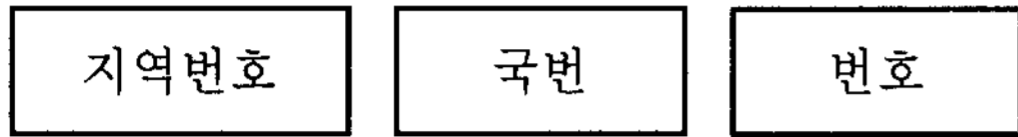
터 표준 관리와 데이터의 지속적인 품질 향상을 위한 요구사항 관리를 통한 데이터 품질관리 프로세스 개선 사례를 중심으로 연구를 진행하였다.

3.2.1 데이터 표준 관리

데이터 표준 관리는 데이터에 대한 공통된 시각을 유지할 수 있도록 데이터 표준화 원칙 정의, 표준화 원칙에 따른 표준 정의, 표준 변경, 표준 적용 통제 등을 수행하는 관리 프로세스를 의미한다(DPC, 2006c). 데이터 표준 관리와 관련성이 높은 데이터 품질 기준은 일관성 및 정확성이며, 데이터 관련자들 간의 공통된 용어와 기준을 사용하게 함으로써 데이터 품질을 향상시킬 수 있다. 여기서는 사례 데이터베이스의 데이터 표준화 항목 중에서 대표적으로 전화번호, 주소, 업종에 대한 표준화에 대해서만 살펴 보도록 하겠다.

사례 데이터베이스에서 전화번호는 모든 고객 정보에 대한 키 값으로 번호에 대한 표준화를 통해 오류 없이 체계적으로 관리되어야 하는 항목이다. 표준화를 통해 원시 데이터에서 잘못 들어오는 구 지역번호 및 구 국번을 수정해 주

며, 외형적인 오류(일률적 번호, 자릿수 오류 등) 번호를 찾아서 정제해 준다.



〈그림 2〉 전화번호 표준화 레이아웃

<그림 2>와 같이 전화번호를 지역번호, 국번, 번호로 파싱한 후 각 필드를 수정해 준다. 원시 데이터에서는 4byte를 기준으로 그 외 공백은 0으로 채워져 오기 때문에 실제 번호가 아닌 0을 제거해 준다. 예를 들어, “0042”은 “042”로 변경하고 “0412”는 구 지역번호이기 때문에 “041”로 변경 처리한다(<그림 3> 참조). 또한, 국번 마스터 DB를 기준으로 잘못 입력된 국번을 정제해 주는데 “052”지역번호에서 “2219”는 없는 국번이므로 유사국번인 “219”로 변경되어 처리되며, 지능망, 평생번호인 “0502”, “080”등의 특수번호도 관리해 준다.

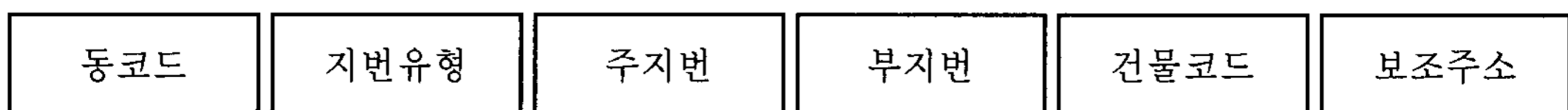
주소 표준화는 고객이 혼용하여 사용하는 주소체계 및 건물 명칭을 하나의 기준으로 통일시켜 주는 개념이다. 주소 체계는 행정동/법정동을 별도로 표준화시키며, 건물 명칭은 건물명 표준

화 지침에 의해 관리된다. 주소 체계의 표준화 방안은 행정동/법정동을 동시에 관리하는 것이 두 가지 정보의 손실 없이 각각 활용이 가능하며 지역별 분석이 용이하다. 주소 표준화의 기본 개념은 고객 주소 프로그램의 별도 변경 없이 기존 고객 주소 및 신규 입력 주소에 대하여 고객 주소를 분석하고 정비할 수 있도록 구성 요소 별로 자동적으로 세분화하는 작업이다. 파싱한 후 각 세분화된 자료를 숫자로 코드화 하는데 코드는 간결성, 식별성, 질서성, 친밀성, 공용성, 확장성 등을 갖추어야 한다. 행정동 코드와 법정동 코드는 각 10자리로 관리하고 건물 코드는 9자리, 지번유형 코드는 1자리로 관리한다.

<그림 4>는 주소 표준화 레이아웃을 나타내고 있으며, <그림 5>는 주소 표준화 결과를 예시로 나타낸 것으로 원시주소에 대해 주소 요소 별로 파싱을 통해 표준화 주소로 변형된 결과를 보여주고 있다. 예를 들어, 원시 주소에 주지번이 누락되어 있는 “충청북도 단양군 단양읍 상진리”의 경우에는 주소 기준 정보에 의해 “528”번지를 추가해주고, “충청남도 천안시 쌍용동”의 법정동 주소를 “쌍용 2동”이라는 행정동으로 변경해 주고, 원시주소의 건물명이 “형곡주공 4

원시 Data		전화번호 Parsing 결과		
전화번호		DDD	국번	번호
0042	0220 1114	042	220	1114
0412	0330 1004	041	330	1004
0052	2219 0507	052	219	0507
0502	0200 9017	0502	200	9017
0080	0790 2482	080	790	2482

〈그림 3〉 전화번호 표준화 결과 예



〈그림 4〉 주소 표준화 레이아웃

원시 주소		표준화 후 주소	
부산광역시 부산진구 부전동	503-15롯데백화점 5층휠라수영목	부산-부산진-부전-503-15(롯데백화점5층휠라수영목)	
충청북도 단양군 단양읍 상진리	공간아파트 103동 901호	충북-단양-단양-상진-528(공간아파트-103-901)	
부산광역시 사상구 모라동	728 제일빌딩 302호	부산-사상-모라-728-10(제일빌딩-302)	
충청남도 천안시 쌍용동	652일성3차능수아파트 307동 503호	충남-천안-쌍용2-652(일성3차능수아파트-307-503)	
대구광역시 달서구 월성동	895-7 대구화물터미널 414호	대구-달서-월성-895-7(대구화물터미널-414)	
경상북도 구미시 형곡동	169 형곡주공4차아파트 405동 502호	경북-구미-형곡-169(형곡4차주공아파트-405-502)	

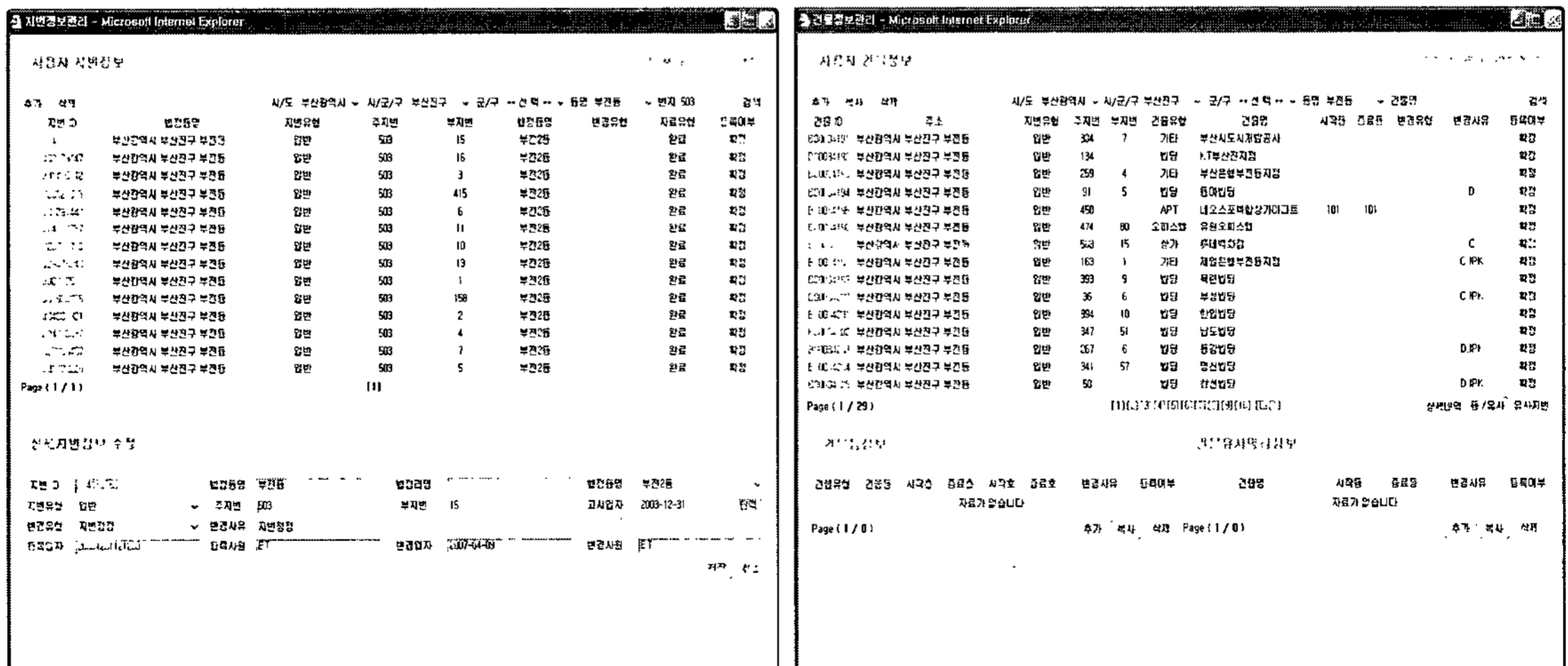
주소요소별 Parsing 결과									
건물유형	행정동코드	법정동코드	지번유형	주지번	부지번	건물코드	건물명	시작동	종료동
05	2623051000	2623010300	1	503	15	B00034197	롯데백화점		
02	4380025000	4380025027	1	955		B00068370	공간아파트	101	103
03	2653059100	2653010200	1	728	10	B00037971	제일빌딩		
02	4413057900	4413012300	1	652		B00181789	일성3차아파트	301	307
10	4713060500	4514013300	1	895	7	B00112826	대구화물터미널		
02	4719058200	4719010900	1	169	11	B00075101	형곡4차주공아파트	401	414

<그림 5> 주소 표준화 결과 예

차아파트” 등으로 입력된 경우에는 표준 건물명인 “형곡4차주공아파트”로 변경하여 처리해 준다. <그림 6>은 주소/건물 표준화를 관리하는 주소/건물 기준정보관리 시스템의 화면으로 지번정보를 관리하는 화면과 건물정보를 관리하는 화면을 보여주고 있다. 불완전한 고객 주소 정보에 대해 마스터 DB의 기준 정보를 통해 정확한 주소의 지번과 건물명을 입력해 주는 역할

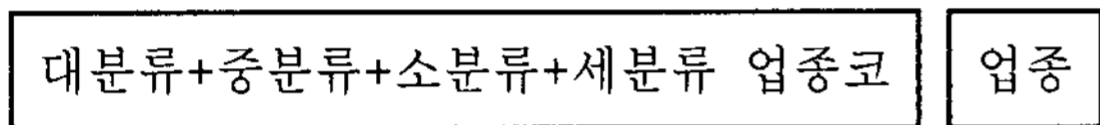
을 하고 있다. 또한, 건물명에 대해서는 건물유형, 시작동, 종료동 등의 정보를 함께 관리하고 있다.

업종은 인명을 제외한 상호인 경우에 반드시 입력해야 하는 필수항목으로 상세 업종에 대한 종류가 매우 많고 원시 데이터를 통해 들어오는 유사 업종에 대한 기준이 명확하지 않아 사례 데이터베이스에서는 업종코드를 표준화하여 사



<그림 6> 주소/건물 기준 정보 관리 화면

용한다. 업종 표준화는 고객이 다양하게 혼용하여 사용하는 업종체계를 동일 기관명을 기준으로 가장 많이 사용하는 업종코드(Most Frequency)로 통일 시켜주는 방법으로 처리한다. 업종코드는 대분류(2자리), 중분류(1자리), 소분류(1자리), 세분류(2자리)로 구성되는 6자리를 하나로 묶어서 관리한다. 예를 들어 “76”은 금융업이고 “761”은 일반금융업이고 “7619”는 기타 일반금융업이고 “761904”는 상호저축은행이라는 상세업종명이 되는 것이다(그림 7 참조).



<그림 7> 업종 표준화 레이아웃

<그림 8>에서 업종 표준화의 예를 들면, 동일 업종으로 분류되어야 하는 기관인 “어린이집”에 대해 원시 데이터에서 다른 업종코드가 입력되어 들어오는 경우 업종 마스터 DB를 기준으로 해당 기관에서 가장 많이 사용되는 업종코드로 변경하여 동일기관에 대한 상세업종코드를 통일시켜 준다.

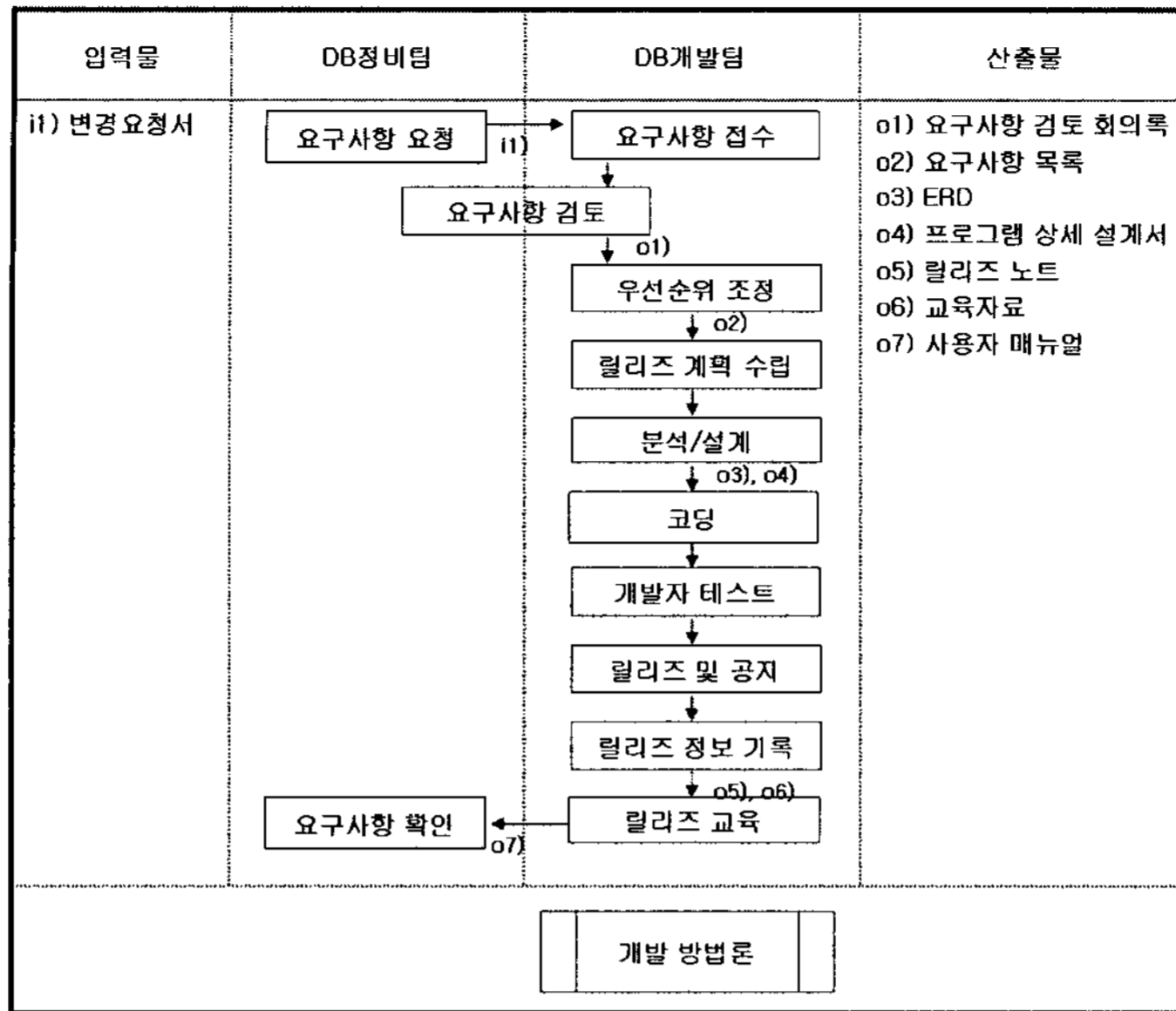
3.2.2 요구사항 관리

요구사항 관리는 데이터를 비롯하여 관련 애플리케이션 및 시스템 전반에 걸친 사용자의 요

구를 수집하고 분류하여 반영하는 작업을 의미한다. 사용자의 정보 요구사항을 종합적으로 검토, 확인하여 요건에 적합하도록 시스템을 개선, 반영함으로써 사용자의 만족도를 높이고 고품질의 서비스를 제공할 수 있다. 과거 사례 기업에서는 요구사항 관리 프로세스가 존재하지 않았다. 특별한 체계 없이 그때 그때의 필요한 부분에 대해 주먹구구식으로 요청하고 처리하는 방식이라 제대로 된 피드백을 주지 못하는 상황이었다. 현재 사례 기업에서는 요구사항의 변경을 지원할 수 있는 방안으로 요구사항 관리 프로세스를 수립하여 각종 데이터 분석 활동을 지원하고 있으며, 사용자 요구사항을 단계별 프로세스를 통해 체계적으로 관리하여 데이터 품질 개선에 반영함으로써 데이터 결함율을 감소시키고 수집된 데이터를 분석하여 그 결과를 다시 피드백함으로써 체계적인 데이터 품질 관리를 하고 있다. 일 변동 원시 자료를 NDMS 서버에서 배치 프로그램을 통해 데이터를 정제하고 DB 편집자들이 이에 대한 검증 및 정비 작업을 하여 보완된 데이터는 수집 분배 서버의 라우터와 전용선을 통해 내부 및 외부 수요처로 일 변동 자료가 전송되게 된다. 이때 일변동으로 발생하는 데이터 변경사항에 대하여 배치 로직 및 지침 처리가 잘못 되어 있는 경우 데이터를 정비하는 DB 정비팀에서 변경 요청 사항

원시 Data		표준화 후 업종	
업종코드	기관명	업종코드	업종명
980911	(사단)대한노인회	980911	노인회
8611	건양대학교병원	861105	병원
000000	경주문화재연구소	963202	문화재청
NULL	포항고등학교	871302	고등학교
871100	혜원어린이집	871102	어린이집
871102	중앙어린이집	871102	어린이집

<그림 8> 업종 표준화 결과 예



〈그림 9〉 요구사항 처리 흐름

을 요구사항 관리 프로세스를 통하여 등록하고 처리된 내용을 확인함으로써 데이터 품질 향상에 기여하고 있다. <그림 9>는 요구사항 처리 흐름을 요청부서와 처리부서로 구분하여 나타내고 있으며 각 항목별 입력물과 산출물을 표기하고 있다.

사례 기업에서는 앞서 제시한 요구사항 관리 프로세스에 대하여 요구사항의 등록 및 관리를 지원하는 요구사항 관리 도구를 자체 개발로 구축하여 실무에 활용하고 있다. 기존의 대부분의 요구사항 관리 도구들이 요구사항 분석 방법론에 의거한 분석지원 도구들로 다양한 요구사항

〈표 1〉 요구사항 관리 도구의 주요 기능

기능명	기능 설명
MyToDoList	자신이 처리해야 하는 업무 목록을 보여주고 각 역할에 따라 요청 처리 화면이 나타난다.
MyDoneList	자신이 등록했거나 처리한 모든 요구사항 정보를 보여주며 현재 진행상태를 확인 가능하다.
등록 및 처리	요구사항 변경 요청서를 등록하고 처리하는 기능으로 해당 담당자가 상태정보와 처리내역을 역할별로 입력 또는 처리할 수 있다.
검색	등록/처리중인 요구사항에 대한 다양한 검색 기능을 제공한다.
통계	요구사항 등록현황, 진행현황, 요청구분현황, 평균처리일수현황 등 주요통계 기능을 제공한다.
시스템 관리	개발구분, 요청구분, 우선순위, 버전, 기능 등 시스템에서 사용하는 각종 코드 정보를 등록하고 관리한다.
릴리즈 정보	릴리즈 정보를 통해 변경기록을 제공한다.

<표 2> 요구사항 관리 도구의 주요 항목별 코드

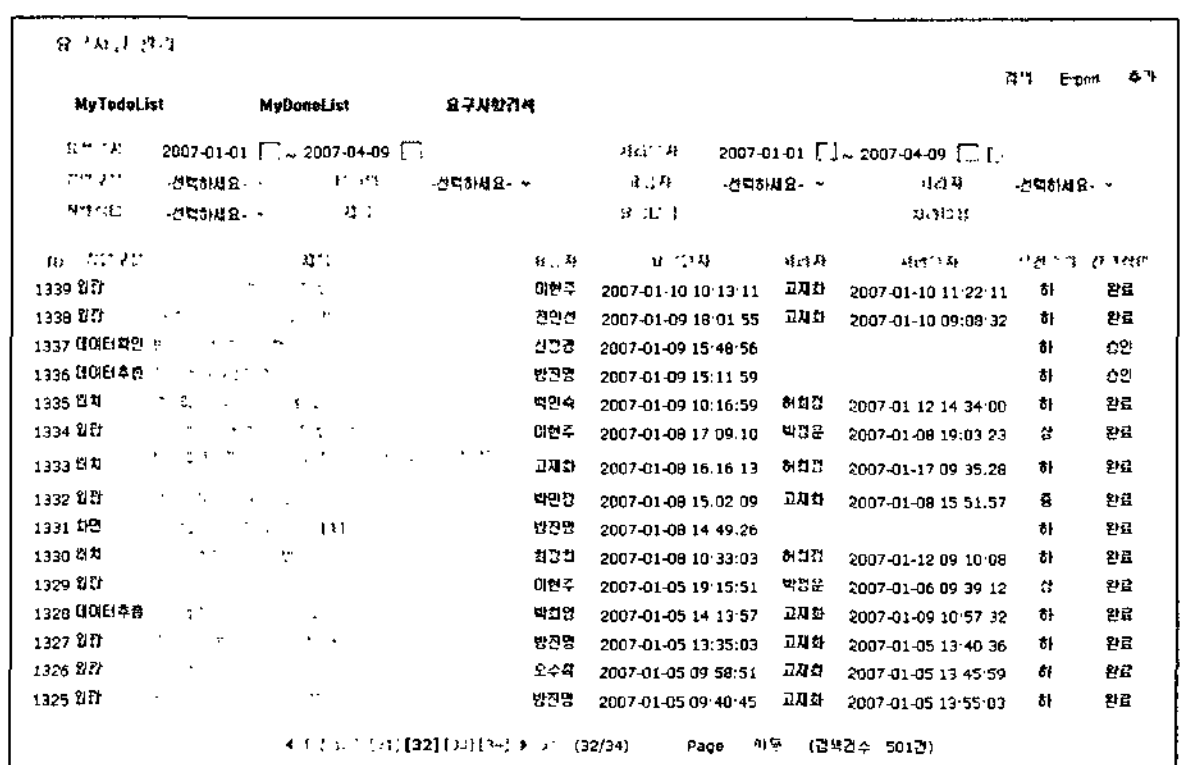
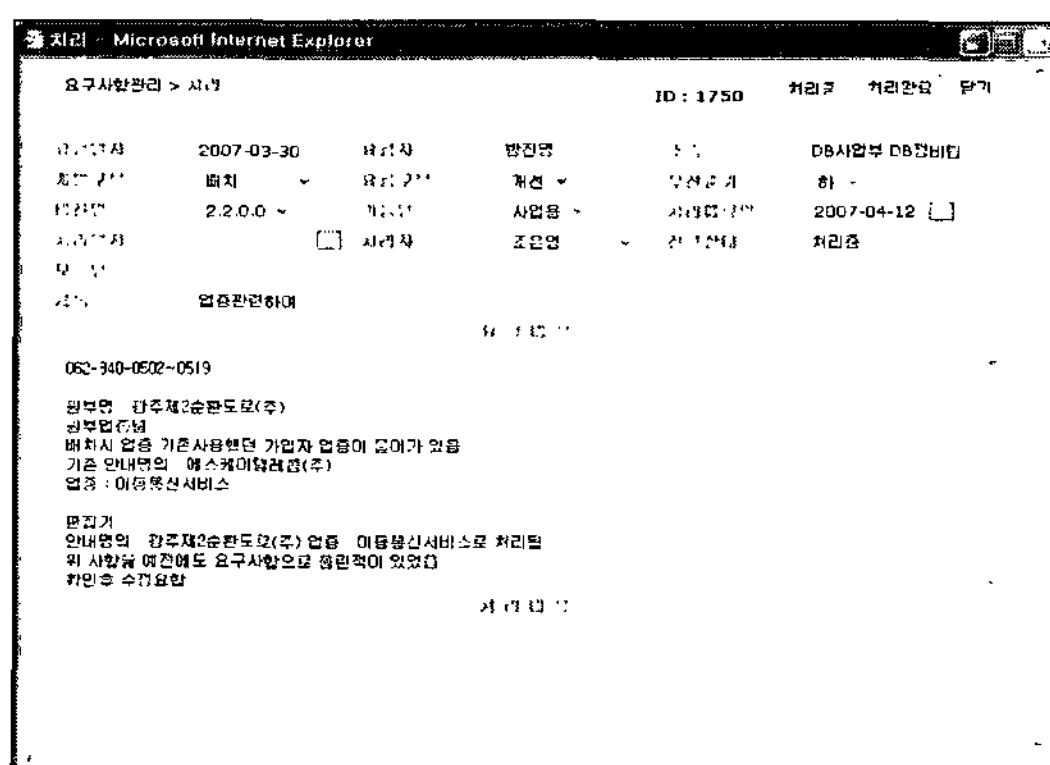
구분	내용
개발구분	<ul style="list-style-type: none"> 지침: 데이터 적용 지침에 대한 추가 또는 변경을 해야 하는 사항 배치: 전반적으로 배치 로직을 변경해야 하는 사항 화면: 편집기 화면 프로그램에 대한 요청 사항 일괄: 다수의 데이터를 추가/변경/삭제해야 하는 사항 데이터확인: 특정 case에 대한 확인이 필요한 사항 데이터추출: 특정 데이터 추출에 대한 요구사항
요청구분	<ul style="list-style-type: none"> 개선: 새로운 기능에 대한 요구사항, 기존 기능에 대한 변경 요구사항 결함: 일반적으로 사용하는 시스템 또는 데이터에 대한 오류나 결함 지원: 시스템 또는 데이터에 대한 모든 지원 질문: 시스템 기능 또는 데이터에 대한 질문 취소: 사용자의 잘못이나 이해 부족으로 인한 실수 중복: 중복된 요구사항(original report ID 명시)
우선순위	<ul style="list-style-type: none"> 상: 중요한 기능이 다르게 동작하거나 중요한 데이터가 다르게 입력되어 있는 것 중: 기능 또는 데이터의 개념이 빗나가고 있는 것 하: 기능에 영향을 받지 않지만 바람직하지 않은 것

들을 효과적으로 지원하지 못하는 단점을 극복하기 위하여 사례 기업 특성에 맞는 요구사항 관리 프로세스의 단계를 이용하여 요구사항들이 효과적으로 반영되고 관리되도록 설계하고 구현하였다. 이에 사례 기업의 요구사항 관리 도구의 주요 기능은 아래의 <표 1>과 같이 정리할 수 있다.

요구사항 관리 시스템은 요청이나 결함의 등록 및 처리에 있어서 각 담당자의 역할(요청자,

개발자, 관리자)별로 필요한 검토/승인 과정을 지원하고 있으며 검색과 통계 정보에 필요한 각종 상태정보를 각 담당자의 역할별로 효과적으로 입력, 관리할 수 있도록 지원해 준다. 요구사항 등록과 처리시에 입력되는 항목과 주요 항목별 코드는 <표 2>와 같다.

<그림 10>은 요구사항 관리 시스템의 처리 및 검색 화면이다. 요청자가 등록 화면을 통해 입력한 개발구분, 요청구분, 우선순위, 버전, 기



<그림 10> 요구사항 처리 및 검색 화면

능명 등의 항목을 처리자가 처리 화면에서 변경할 수 있으며, 처리예정일과 처리일자, 모듈명, 처리 내용 등을 입력하여 요구사항을 처리할 수 있다. 또한, 요구사항 검색 기능을 통해 요청일자, 처리일자, 개발구분, 버전, 요청자, 처리자, 진행상태, 제목, 요청내용, 처리내용 등의 다양한 검색조건으로 해당 요구사항을 조회할 수 있으며 요청서 제목을 클릭하여 상세 내용을 확인할 수 있다.

3.3 주요 이슈 및 해결 방안

어떠한 시스템이든지 시스템의 구축은 조직 혹은 시스템의 내부적인 구성의 단위에서 다양한 문제점이 제기된다. 시스템을 구축함에 있어서는 많은 문제점이 발생할 수 있지만 크게 조직단위의 문제와 시스템 구축 자체의 문제로 구분하여 볼 수 있다. 이 절에서는 사례기업의 데이터 품질 관리 프로세스 개선 프로젝트 수행 과정에서 발생한 주요 이슈 및 위험요인에 대하여 현상과 해결 방안을 정리해 보고자 한다.

3.3.1 조직의 변경

사례 프로젝트를 수행하기 위하여 처음에는 특별 전담반을 구성하여 진행 하였으나, 그 책임이 불분명하고 업무 수행상에 요구되는 모든 활동에 대한 권한을 부여하기가 힘들었으며, 특정 과제를 수행하기 위한 전담반의 상이한 팀 구성원들의 이질감 때문에 팀웍을 높이는데 어려움을 겪게 되었다. 데이터 관리 프로세스는 데이터 및 데이터 구조의 품질을 안정적으로 유지, 개선하기 위한 활동으로 절차, 조직, 인력 등을 포함한다. 데이터 품질 관리 조직은 데이터 표준 개발 및 형상 관리, 검증, 표준화 절차를 수립하고 운영한다. 또한, 지속적으로 데이터 요구사항 및 이슈 사항을 조정하고 통합하며 데이터 표준을 확인하고 적용하는 역할을 수행해야 한다. 이러한 데이터 품질 관리 전담 조직의 필요성이 제

기되어 프로젝트 진행 중반인 2005년 2월에 조직개편이 단행되었다. 기존에는 전사적 DB 관리 및 발전전략 수립을 위한 조직인 DB 관리부의 하위조직이 3개의 팀, 즉, 시스템운영팀, 품질관리팀, 정비팀으로 구성되어 있었으나 시스템 운영팀과 품질관리팀이 합쳐져 DB 개발팀이 신설되었다. 때문에 DB 관리부는 NDMS 데이터 품질 관리 프로세스 개선과 DB 구축 및 관리를 위한 DB 개발팀과 일변동 자료에 대한 정비 업무를 지원하는 DB 정비팀으로 편제되었다.

프로세스의 개선은 조직의 변화를 수반한다. 특별 전담 팀을 구성하여 프로젝트를 진행하였지만 해당 업무에 대한 책임감과 소속감을 증대시키고 이 프로젝트에만 전담할 수 있도록 하기 위해 별도의 조직으로 구성함으로써 업무 생산성 향상과 인력 효율화에 기여할 수 있기 때문이다.

3.3.2 조직간의 내부적 갈등 문제

내부적으로 DB 정비팀과 DB 개발팀 간의 갈등이 많았는데 기본적인 DB 정비팀의 입장은 전산화를 하더라도 대상 자료의 정확성을 믿을 수 없고 반드시 정비팀원이 확인작업을 해야 한다는 주장이었다. 사실상 시스템의 구축은 현업 사용자의 요구사항을 기반으로 구축되어야 하지만 이 사례의 경우 거의 대부분이 DB 개발팀의 원시 데이터 및 업무 프로세스 분석 작업에 의해 진행되었기 때문에 문제가 발생하게 되었다. 데이터 품질 관리 프로세스 개선 프로젝트를 진행하기 위해서는 해당 시스템의 주요 사용자인 DB 정비팀의 요구사항 수렴과 기존 업무 프로세스에 대한 정의 등이 필요했지만 바쁜 현업 스케줄에 의해 많은 지원을 받을 수 없었기 때문이다. 당시 프로젝트 팀원은 다음과 같이 말하고 있다.

“총괄 프로젝트 매니저였던 K팀장님은 예전 시스템에 대해 많이 알고 계신 분이셨고 저희한테도 무조건 옛날 시스템부터 분석하라고 하셨

어요. 그래서 당시 메인 개발자였던 E팀장님과 마찰이 심했죠. K팀장님은 옛날 것도 알아야지 지금 것도 안다고 주장하셨고 E팀장님은 어차피 시스템이 다 바뀌었기 때문에 옛날 건 불필요가 없다 그러셨어요. 그래서 두 분의 의견이 너무 안 맞아서 2005년 초에 조직개편이 되면서 K팀장님이 DB정비팀장으로 가시게 된 거죠. DB개발팀은 E팀장님 주축으로 다시 뭉치게 됐고요. 그 후에 또다시 분쟁이 생겼던 이유가 E팀장님이 전산화가 90% 이상 된다고 자신하시니까 K팀장님은 무슨 말이냐, 90% 될 수가 없다, 된다면 정비팀원들은 다 잘라야 한다, 이런 것들 때문에 또 많이 싸웠죠. 근데 정말 전산화가 90% 되면 정비팀원들은 할 일이 없는 거죠. 그래서 후에 인력을 다른 곳으로 돌린 것이고, 당시에 K팀장님은 전산 처리를 하더라도 확인 작업은 반드시 해야 한다, 나는 도저히 전산 못 믿는다, 사람이 더 정확하다 이러셨고 E팀장님은 그렇지 않다, 어떻게 사람이 더 정확하냐, 이렇게 된 거죠.”

전통적인 관점에서 갈등은 개인적인 차이 또는 리더십의 부재로 발생하여 조직의 효율을 떨어뜨리고 피해야만 하는 것이며 상위관리자의 중재나 갈등 원인의 실질적인 제거를 통해서 해결할 수 있다고 생각하였다. 그러나 현대에 들어와 갈등은 조직의 상호작용에 의한 갈등은 피할 수 없고 이로울 수도 있으며 담당자와 상사가 동시에 참여하여 원인과 문제해결 방안을 식별함으로써 해결할 수 있다고 생각하는 추세이다(PMI, 2004). 프로젝트 수행 중 가장 빈번하게 발생하는 7대 갈등 원인은 일정(Schedule), 프로젝트 우선순위(Project priorities), 자원(Resource), 기술적 옵션(Technical opinions), 관리 절차(Administrative procedure), 원가(Cost), 대인관계(Personality)라고 알려져 있다. 프로젝트 수행 도중 발생하는 이러한 갈등은 상황에 따라 갈등 해결 기법이 다른데, 개인적인 갈등(intra-individual), 개인간의 갈등(inter-personal), 그룹

간의 갈등(inter-group) 등 종류에 따라 약간씩 차이가 있으며 어떤 면에 주안점을 두느냐에 따라 문제 해결(Problem Solving), 양쪽 의견 절충(Compromising), 보류(Withdrawal), 상대 의견 수용(Smoothing), 자기 의견 관철(Forcing) 등의 전략을 다양하게 적용할 수 있다.

갈등의 주된 원인은 자기부서를 우선적으로 생각하는 부서이기주의와 인력 감축으로 인한 피해의식, 전산 시스템 및 자료 처리 지침에 대한 신뢰 부족 등이었다. 현업부서는 적극성을 가지고 업무를 지원하려는 의욕이 부족하였고 소극적이고 의존적이었다. 일반적으로 프로세스 개선에서 가장 중요한 것은 최고경영자의 강력한 후원이며 다음으로 전사적인 공감대 형성, 그리고 기업상황에 적합한 기법선택과 지속적인 변화관리라고 할 수 있는데 본 사례에서도 마찬가지로 최고경영자의 의지가 가장 중요하였다. 부서간의 이해가 상충되는 등의 장애돌파는 최고경영자를 만나서 진행과정과 문제를 보고하고 지시를 받아서 진행하는 등, 최고경영자의 지속적인 관심이 장애를 돌파하는 가장 중요한 요소였다.

3.3.3 개발자의 전문성

사례 기업은 주로 외주개발과 아웃소싱으로 시스템 개발을 추진하는 조직이었기 때문에 사내에 전산 관련 전문인력이 부족한 실정이었다. 정보시스템 개발 프로젝트 성공에 관한 연구에서나 실패를 야기시키는 요소들로 기술적 능력과 채택된 기술의 변경, 기술적으로 우수한 인력 확보 등을 들고 있다. 프로젝트에 사용된 기술이 검증된 기술이며 프로젝트 내부에 충분한 전문가가 투입되어 있어야 한다. Pinto와 Slevin (1987)의 연구에서는 프로젝트에서 요구되는 충분한 기술적 역량을 가진 인력과 도구의 완비는 필수 조건이라고 하였다.

기술적으로 교육된 전문인력이 부족했기 때문에 프로젝트 초반에는 관련 전문 교육을 수행

하여 기존 인력의 전문성을 향상시켰으며, 중급 개발자와 초급 개발자를 같은 업무에 배정함으로써 위험요인을 줄일 수 있었다. 또한, 현업 요구부서의 명확하지 않은 요구사항에 대해서는 단계별 검토회의 등을 진행하는 등 사용자 검토를 증가시켰고 요구사항 정의 방법 등의 적절한 지식을 전수함으로써 해결하였다.

3.3.4 성능 저하에 따른 튜닝

시스템 시험 운용 단계에서 성능 문제가 발생하였는데 사례 기업과 같이 대용량의 데이터를 관리하는 데이터베이스에서의 성능은 매우 중요한 이슈이다. 데이터베이스 관리(성능 관리) 프로세스는 데이터 처리 작업을 모니터링하고 성능의 최적화를 위한 튜닝 등의 작업을 수행하는 프로세스이다. 데이터베이스의 처리 상황에 대한 지속적인 모니터링과 성능의 최적화를 위한 튜닝 활동, 그리고 적정 성능 유지를 위해 필요한 전산 자원을 예측하고 자원 조달에 대한 계획, 배분, 관리 활동 등이 포함된다. 적시성을 향상시키기 위해서는 지속적으로 개선 요소를 발굴하고 튜닝 및 자원을 재분배하여 성능을 최적화하는 한편 데이터 흐름에 있어서는 불필요한 흐름의 원인을 제거하며 성능 향상을 위한 기술 동향을 파악하고 적용해야 한다(DPC, 2006b).

문제의 원인은 자체적인 쿼리(Query) 튜닝(Tuning)을 수행하였지만 대부분 개발서버 환경에 맞게 되어 있어 실제 운용서버에서 실행했을 때에는 시스템 및 데이터베이스 환경이 달라 문

제가 발생했다. 또한, 프로그램 개발자의 데이터베이스 시스템에 적용되어 있는 인덱스(Index)를 고려하지 않고 작성한 프로그램에서도 발생했다. 때문에 별도의 튜닝(Tuning) 작업을 진행하였는데 튜닝은 데이터베이스 튜닝, 프로그램 튜닝까지 전면적으로 이루어졌다. 개발자들에 대한 자체적인 교육과 전체 프로그램에 대한 튜닝 툴(Tool)을 도입하여 발생하는 Query에 대하여 실시간 분석 및 지속적인 튜닝 작업으로 인해 빠른 시일 내에 안정화 시킬 수 있었다.

IV. 개선된 데이터 관리 프로세스의 측정 및 평가

제시된 데이터 관리 프로세스에 대하여 데이터 품질 향상에 대한 개선효과를 확인해 보기 위해 정성적이고 정량적인 측정 지표를 기준으로 분석 및 평가한 결과를 통해 품질 개선에 대한 유효성을 검증하였다. 사례 프로젝트의 전체 추진기간은 2004년 9월부터 2005년 7월까지 약 10개월이 소요되었으며, 개선 전의 측정 시점은 2004년 7월이고 개선 후의 측정 시점은 2005년 9월이다.

4.1 데이터 품질에 대한 평가

사례기업의 데이터베이스에 대한 물리적인 분석을 통하여 각 데이터 품질 항목별로 품질수준을 평가하였다. 본 연구에서 관련 연구(박종모,

〈표 3〉 데이터 품질수준의 측정지표

영역	데이터 측정 지표
데이터	<ul style="list-style-type: none"> • 데이터 값의 누락여부(NULL 데이터 포함) • 데이터 정확성에 대한 신뢰도 • 데이터 표현형식의 적절성(예: 날짜유형) • 데이터의 최신성 • 데이터의 유일성, 유효범위(제약조건)
데이터 관리	<ul style="list-style-type: none"> • 데이터 표준 관리(용어사전, 명명규칙, 코드표준)

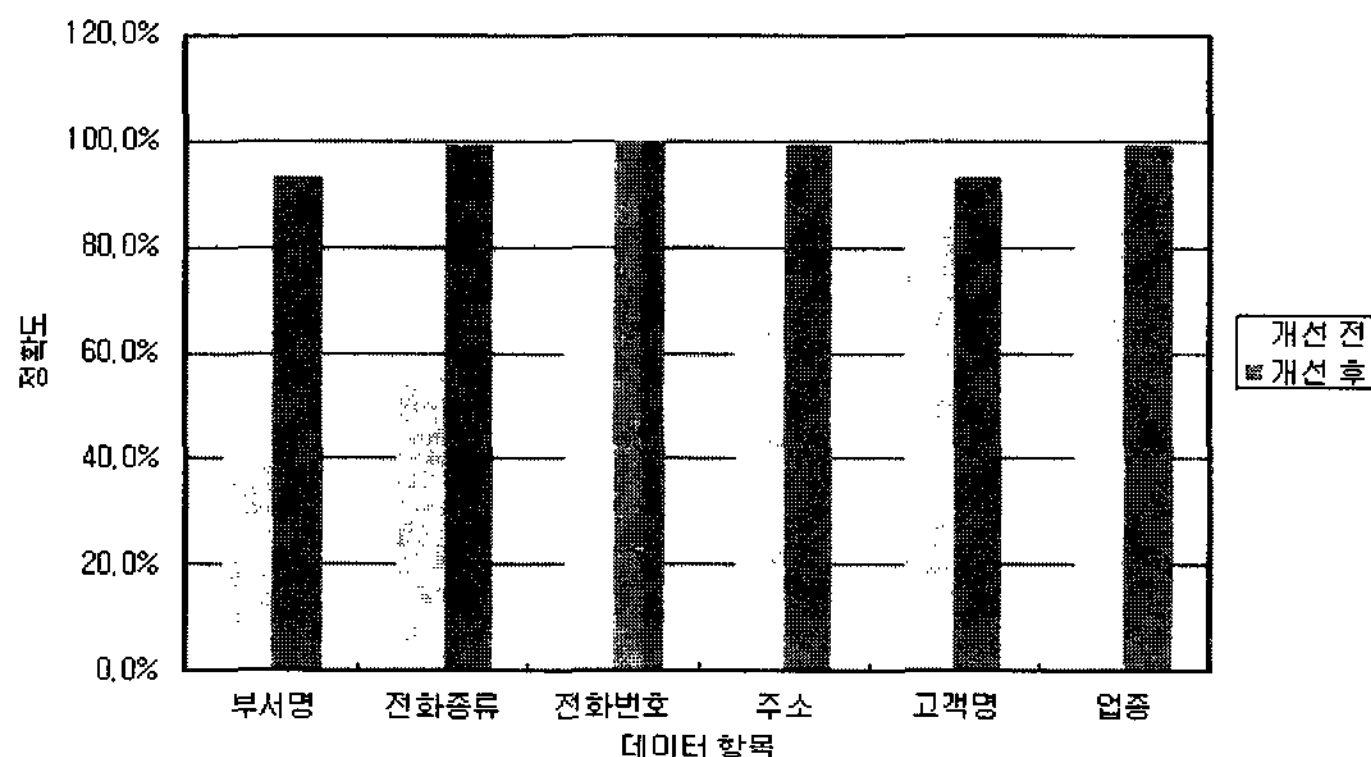
<표 4> 고객 정보의 항목별 데이터 정확도 비교

데이터 항목	개선 전	개선 후	결함의 주요 원인
부서명	41.8%	93%	값의 누락
전화종류	55.4%	99%	데이터 표준관리
전화번호	68.7%	100%	데이터 표현형식의 적절성
주소	69.4%	99%	데이터의 제약조건
고객명	84.5%	93%	데이터의 최신성
업종	87.3%	99%	정확성에 대한 신뢰도

2006)를 참조하여 사례기업의 고객 데이터베이스의 데이터 품질수준을 비교하기 위하여 <표 3>과 같은 측정지표를 제시하였다. 항목별 데이터 정확도의 측정은 <표 3>의 측정지표를 모두 사용하였으며 표본조사를 통해 누적된 전체 마스터 DB에 대해 지역별로 각 1000개를 샘플링하여 각 항목별로 개선 전과 개선 후에 검증작업을 통해 각각의 측정지표의 기준을 적용하여 평가하였다.

<표 4>의 고객 정보의 항목별 데이터 정확도 비교에서 보면 부서명(기관인 경우 존재)과 전화종류 항목이 개선 전의 정확도가 낮은 것을 볼 수 있는데 결함의 주요 원인을 분석한 결과, 데이터의 표준관리와 값의 누락이 가장 문제인 것으로 나타났다. 이는 데이터 자체는 관리하지만 표준화된 업무 규정에 의한 데이터 관리가

부족하기 때문이다. 부서명과 같은 경우에는 상대적으로 중요도가 떨어져서 값의 누락이 많았으며, 전화종류의 경우에는 데이터의 표준이 관리되지 않아 잘못된 값으로 표현되었다. 또한 주소의 경우에는 제약조건이 존재하지 않아 잘못된 형식의 값이 입력되었다. 이와 같이 실제 데이터 품질 관리 프로세스 개선 프로젝트 수행 과정에서 코드의 불일치, 테이블의 정의 불분명, 전체 테이블 파악 미비, 불확실한 매핑 등의 기존에 관리되었던 데이터에 대한 문제점이 발생했다. 또한 기존 데이터에 대한 가장 최신의 데이터 현황이 관리되지 않고, 과거 데이터 형태의 불일치로 인한 문제가 발생했다. 예를 들어, 숫자가 입력되어야 할 필드에 문자가 입력되는 경우나, 코드 값이 불일치한 경우도 있었다. 이로 인해 정확한 데이터의 유형을 분석하지 못하



<그림 11> 항목별 데이터 정확도 비교

<표 5> 작업대상(일평균) 건수 비교

구분	기존	개선	건수차이	개선율(%)
확인대상	21,515	16,238	5,277	24.5%
수정건수	3,679	640	3,039	82.6%
삭제건수	1,166	0	1,166	100%
합계	26,360	16,878	9,482	64%

고, 데이터의 정확도가 떨어지게 되는 것이다.

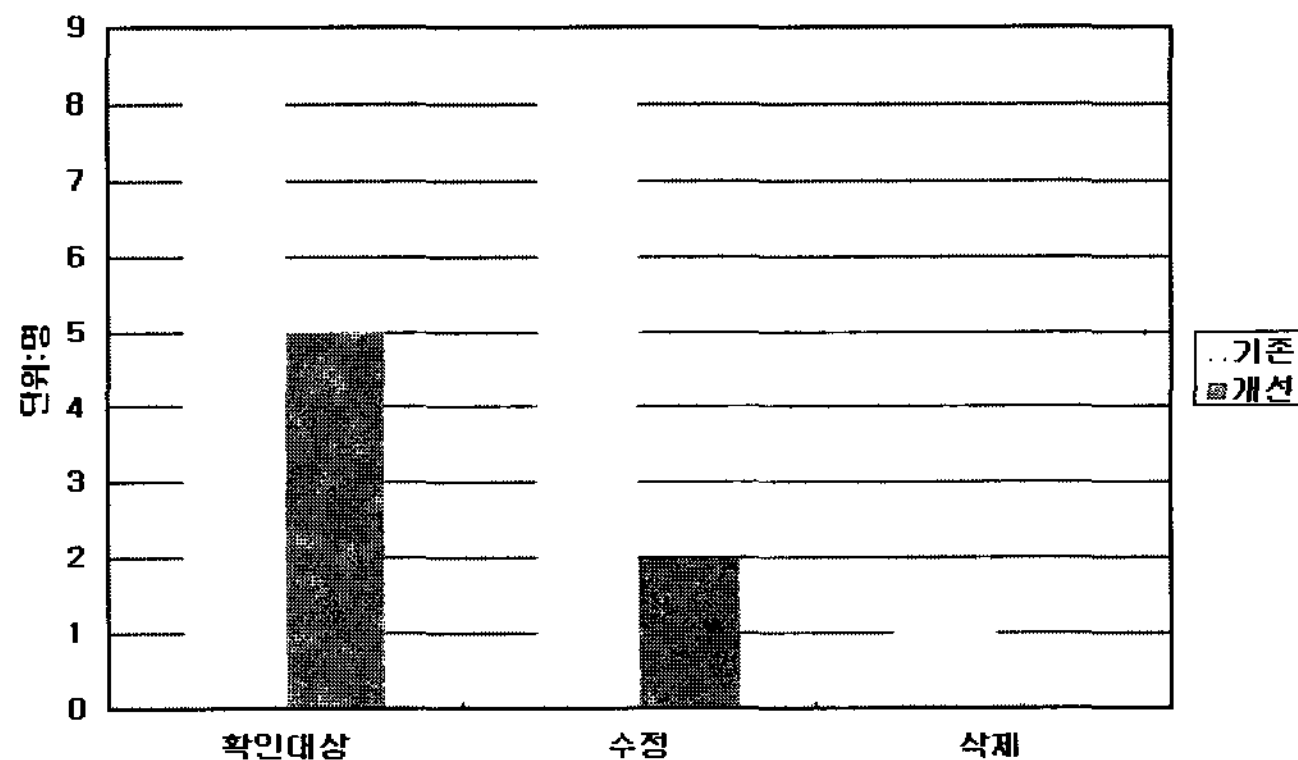
<그림 11>의 항목별 데이터 정확도 비교 분석 결과에서 알 수 있듯이 데이터의 정확한 파악이 필요하고 이에 따라 결함의 근본적인 원인을 제거함으로써 데이터의 품질을 목표비율까지 향상시킬 수 있었다. 그리고 데이터베이스를 구축한 이후에 검증 작업을 거쳐 필요 없는 데이터에 대해 필터링과 클린징을 수행함으로써 변경에 의해 파생되는 문제점을 확인할 수 있었다. 즉 데이터의 정제가 이루어졌기 때문에 유효한 데이터의 건수가 증가하여 데이터의 품질이 향상된 것이다.

4.2 생산성에 대한 평가

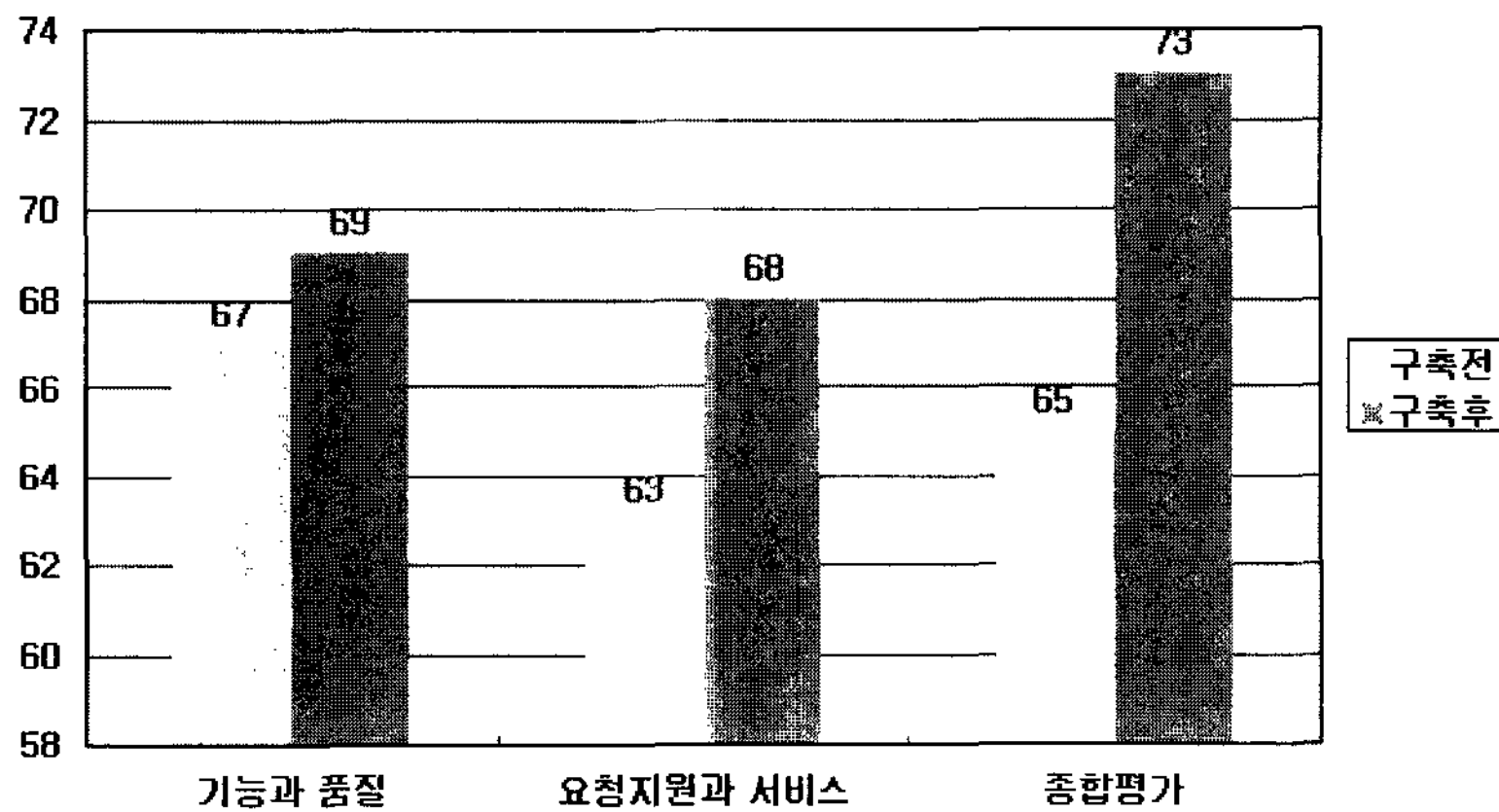
사례 기업의 내부 자료를 분석하고 통합하여 사례 시스템을 대상으로 개선된 데이터 관리 프

로세스를 측정 및 평가하였다. 일 변동 원시 자료를 배치 프로그램을 통해 데이터를 정제하고 DB편집자들이 이에 대한 검증 및 정비 작업을 하게 되는데 확인대상은 전체 검증 대상 건수를 의미하고 수정은 이중 편집한 데이터를 의미한다. 수작업 위주의 자료 처리를 전산 자동화 기능으로 프로세스를 개선함으로써 자동화에 따른 소요 인력을 절감하여 생산성을 높일 수 있었다. <표 5>의 작업대상(일 평균) 건수의 비교에서 유형별 수작업 건수 감소로 인한 개선 비율은 확인대상 24.5%, 수정 82.6%, 삭제 100%로 감소 원인은 일 변동 처리 전산화와 자료 정비 지침에 대한 데이터 표준화 적용 결과이다.

사례기업의 데이터 품질 관리 프로세스 개선 프로젝트 수행 이전에는 DB정비 작업에 18명의 정비 인력이 투입되었다. 수작업성 업무 자동화 및 데이터 관리 프로세스 개선에 따른 건



<그림 12> 투입인력의 변화



〈그림 13〉 시스템 만족도 조사 결과 비교

수 대비 인력 감소 효과는 확인대상 3명, 수정 6명, 삭제 2명으로 데이터 품질 관리 프로세스 개선 프로젝트 수행 이후의 전체 감소 인력은 약 11명이다. <그림 12>에서 보는 것과 같이 개선 비율에 따라 인건비가 절감되는 효과를 볼 수 있었다.

4.3 고객만족도에 대한 평가

데이터 품질 관리 프로세스를 개선하기 전인 2005년도에 시스템 사용자인 DB정비팀과 지역본부 DB담당자들을 대상으로 시스템 기능과 데이터 품질의 전반적인 만족도에 대한 설문조사를 실시하였으며, 프로세스 개선 후인 2006년도에 같은 항목으로 만족도에 대한 설문조사를 재 실시하였다. 이는 사용자들의 시스템 만족에 대한 변화 추이를 관찰하여 데이터 품질 관리 프로세스 개선이 사용자 만족에 어떠한 영향을 끼치는지를 파악하고자 한 것이다. 시스템 사용자 설문조사를 통하여 나타난 데이터 품질 관리 프로세스 개선 프로젝트 수행 전후의 만족도의 변화결과는 <그림 13>과 같다. 데이터 품질 관리 프로세스 개선 전인 2005년도의 시스템 만족도는 기능과 품질 측면에서는 67점, 요청지원과

서비스 측면에서는 63점이었으며, 프로세스 개선 후인 2006년도에는 기능과 품질 측면에서는 69점, 요청지원과 서비스 측면에서는 68점으로 만족도가 증가 하였으며, 종합평가 점수는 데이터 관리 프로세스 개선 전 65점에서 개선 후 73점으로 8점 증가한 것으로 나타났다. 이는 기능과 품질 측면에서는 데이터 표준화 결과가 반영된 것이고, 요청지원과 서비스 측면에서는 요구 사항 관리 프로세스 수립에 따른 결과가 반영된 것으로 볼 수 있으며, 기능과 품질 측면은 이전에 관리되고 있는 것으로 증가율이 낮게 나타난 반면 요청지원과 서비스 측면은 데이터 품질 관리 프로세스 개선을 통한 체계적인 피드백으로 인해 상대적으로 변동폭이 높게 나타났다.

4.4 조직 및 문화에 대한 평가

과거 사례 기업에서는 일 변동 데이터를 정비하는 편집기 프로그램을 외부에서 제공되는 패키지 S/W를 사용하였기 때문에 조직에 맞게 시스템을 변경하는 것이 쉽지 않았다. 때문에 현업의 실무자들은 업무 변화에도 불구하고 기존 프로세스를 그대로 사용하면서 불편을 감수 할 수 밖에 없는 실정이었다. 그러나 데이터 품질

관리 프로세스 개선 프로젝트 수행 이후에는 자체 개발로 구축된 시스템이기에 유지보수를 통해 프로세스를 업무 변경 사항에 맞게 개선할 수 있게 되어 변화의 대응하는 직원들의 마인드가 업무 개선과 혁신의 관점으로 변하게 되는 계기가 되었다. 또한, 유지보수 업무를 효율화하는 것은 각종 비용절감 및 경영개선에 큰 효과를 미치게 된다. 결과적으로 요구사항 관리 프로세스 체계를 확립함으로써 데이터 품질 확보 및 유지보수 업무의 효율화를 달성할 수 있었다.

그리고 시스템화를 통한 데이터 관리 노하우에 대한 지식을 보유함으로써 향후 다른 유사 프로젝트에서 프로세스를 재사용할 수 있는 지식 경영의 의미도 갖게 되었다. 더불어 또 하나의 큰 기대효과는 고객의 요구에 신속하게 대응할 수 있는 체계를 구축한 것이다. 모든 고객의 요구를 신속히 대응할 수 있도록 사업지원 데이터베이스를 구축하였고 고객에 맞게 특화 정비를 함으로써 고객가치에 따른 대응력을 강화하였다. 데이터 표준화 및 요구사항 관리 프로세스를 통해 신속, 정확, 표준화된 데이터베이스를 제공함으로써 고객만족도를 제고할 수 있게 된 것이다.

4.5 평가 분석을 통한 시사점

본 연구에서 제시한 데이터 표준화를 비롯한 데이터 품질 관리 프로세스의 개선을 통해서 얻을 수 있었던 효과를 다시 한번 정리하면 다음과 같다.

첫째, 데이터의 정확한 분석을 통해 결함을 제거함으로써 데이터 품질을 향상시킬 수 있었다.

둘째, 수작업 위주의 자료 처리를 데이터 표준화와 전산화 기능으로 프로세스를 개선함으로써 자동화에 따른 소요 인력을 절감하여 생산성을 높일 수 있었다.

셋째, 데이터 품질 관리 프로세스 개선 후에

사용자 만족도는 기능과 품질 측면, 요청지원과 서비스 측면에서 모두 증가하여 사용자 만족도가 증대한 것으로 나타났다.

넷째, 유지보수에 대한 데이터 요구사항 관리 프로세스를 통해 프로세스를 업무 변경 사항에 맞게 개선할 수 있게 되어 변화의 대응하는 직원들의 마인드가 업무 개선과 혁신의 관점으로 변하게 되는 계기가 되었다.

본 연구에서는 사례기업을 대상으로 데이터 품질 관리 프로세스에 대하여 데이터 표준 관리와 요구사항 관리를 이용한 구현 사례를 연구하였다. 특히 현업 실무자들의 고객만족도 평가를 통하여 사용의 편리성, 안정성, 신뢰성, 운영관리의 용이성에 대하여 사용자의 공감을 얻을 수 있었다. 이것은 데이터 품질 관리 프로세스의 개선을 통해 업무 효율을 높였다는데 그 의의가 있다.

V. 결 론

본 연구는 실제로 데이터 품질 관리 프로세스 개선을 통해 데이터 품질이 향상된 기업의 사례를 통하여 프로젝트 수행 과정에서의 주요 이슈와 위험요인을 살펴보고 그 해결방안을 제시함으로써 데이터 품질 향상을 위해 노력하는 타 기업들에게 도움을 주고자 하였다. 또한, 개선된 데이터 품질 관리 프로세스에 대한 다차원적인 평가로서 데이터 품질, 생산성, 고객만족도, 조직 및 문화의 측면에서 정성적이고 정량적인 지표를 통한 개선효과를 살펴보고 평가함으로써 제안된 프로세스에 의해 품질수준이 향상되었음을 검증하였고 평가 분석을 통한 시사점을 도출하였다.

본 연구 사례에서는 데이터 표준 관리와 데이터 요구사항 관리를 통해 데이터 관리 프로세스를 개선하였다. 데이터 표준 관리 프로세스를 통해 데이터 공유 및 통합, 프로세스 개선 활동을 지원하고 있으며 공통된 표준을 사용하여 데

데이터의 일관성과 정확성을 유지함으로써 데이터 품질을 향상시킬 수 있었다. 또한, 요구사항 관리 프로세스를 수립하여 각종 데이터 분석 활동을 지원하고 있으며 사용자 요구사항을 단계별 프로세스를 통해 체계적으로 관리하여 데이터 품질 개선에 반영함으로써 데이터 결함율을 감소시키고 수집된 데이터를 분석하여 그 결과를 다시 피드백함으로써 체계적으로 데이터 품질 관리를 할 수 있었다.

사례 프로젝트 수행 과정에서의 주요 이슈에 대한 해결방안으로는 첫째, 부서간의 이해가 상충되는 등의 내부적 갈등 문제는 최고경영자를 만나서 진행과정과 문제를 보고하고 지시를 받아서 진행하는 등, 최고경영자의 지속적인 관심이 장애를 돌파하는 가장 중요한 요소였다. 둘째, 기술적인 전문인력이 부족했기 때문에 발생한 문제는 프로젝트 초반에는 관련 전문 교육을 수행하여 기존 인력의 전문성을 향상시켰으며, 중급 개발자와 초급 개발자를 같은 업무에 배정함으로써 위험요인을 줄일 수 있었다. 셋째, 성능 저하에 대한 문제는 튜닝 툴을 도입하여 Query에 대한 실시간 분석 및 지속적인 튜닝 작업과 개발자들에 대한 자체적인 교육을 통해 빠른 시일 내에 안정화 시킬 수 있었다.

데이터 품질의 문제점을 해결하기 위해서는 현재의 품질 상태를 정확하게 인식해야 한다. 품질 수준이 파악 되어야만 이로 인한 문제점 및 원인을 정확하게 분석할 수 있고, 대응 방안도 마련할 수 있다. 일반적인 접근은 데이터 품질을 측정하고 품질 불량인 대상에 대한 개선을 수행하는 것이지만 이러한 방식은 데이터 관리 기능의 미흡이라는 근본적인 원인을 해결하지 않는 것이므로 품질을 특정 수준 이상으로 끌어 올리는 힘들다. 또한, 데이터 품질관리는 지속적으로 수행되어야 효과를 얻을 수 있으며 일시적인 조치를 통해서 확보된 품질은 오래 지속되기 어렵기 때문에 데이터 품질과 관련이 있는 핵심 프로세스를 정비하고 유지하는 것이 중요

하다. 궁극적으로 데이터 품질 수준을 측정하는 것도 중요하지만, 데이터 품질을 관리하는 프로세스의 수준을 평가하고 이를 향상시키는 것이 근본적인 해결책이라고 할 수 있다.

참 고 문 헌

- DPC, 2006 데이터베이스 백서, 한국데이터베이스진흥센터, 2006a.
- DPC, 데이터 품질관리 성숙모형, 한국데이터베이스진흥센터, 제1.0호, 2006b.
- DPC, 데이터 품질관리 지침, 한국데이터베이스진흥센터, 제2.1호, 2006c.
- 김문영, 체계적인 데이터 품질 관리를 위한 대안을 찾아서, 마이크로소프트웨어, 제255호, 2005, pp. 208-216.
- 김은영, “데이터 품질 기준의 중요도에 관한 연구: 평가자 역할과 시스템특성을 중심으로”, 한국외국어대 경영정보대학원 석사학위논문, 2004.
- 김진섭, “데이터표준화 사례를 통한 데이터 품질 향상에 대한 연구”, 한국정보과학회 가을 학술발표논문집, Vol.33, No.2(C), 2006, pp. 210-213.
- 김찬수, 박주석, “데이터 품질관리 성숙도모델에 대한 연구”, 정보관리학회지, 제20권, 제4호, 2003, pp. 249-275.
- 박종모, “데이터 품질 향상을 위한 데이터웨어하우스 개발 프로세스의 개선”, 단국대학교 대학원 박사학위논문, 2006.
- 원종근, “외부 데이터 보강을 통한 고객 데이터 품질 제고에 관한 연구”, 서울시립대 대학원 석사학위논문, 2005.
- 이승원, “데이터 품질관리 적용 사례”, 데이터베이스 그랜드 컨퍼런스, 2004.
- 이재범, 한만호, “데이터 품질관리 방안 연구”, 서강경영논총, 제15권, 제1호, 2004, pp. 113-122.

- 이정우, “정보품질에 영향을 미치는 요인들”, 데이터품질관리포럼 세미나, 2005.
- 이춘열, 박현지, “데이터베이스 품질 평가에 관한 사례 연구”, *Journal of Information Technology Applications and Management*, 제11권, 제4호, 2004, pp. 209-225.
- 이화식, 데이터 아키텍처: 전사적 아키텍처와 데이터 아키텍처의 개념, 컴퓨터월드, 제241호, 2003, pp. 214-234.
- 정경수, 김병곤, 장상도, “데이터 웨어하우스의 데이터 품질 향상을 위한 개념적 프레임워크의 개발”, 경영교육논총, 제19권, 1999, pp. 191-201.
- 홍기선, “전사 데이터품질관리시스템을 통한 데이터 품질향상 사례”, 데이터품질관리포럼 세미나, 2005.
- 홍일유, “DB 품질분석을 위한 프레임워크의 개발에 관한 연구: 데이터 수명주기 관점”, 경영학논집, 제30권, 제2호, 2004, pp. 165-192.
- 홍현진, “웹 기반 데이터베이스의 품질평가 기준 개발에 관한 연구”, 한국문헌정보학회지, 제39권, 제2호, 2005, pp. 211-235.
- Ballou, D. P. and Tayi, G. K., “Enhancing Data Quality in Data Warehouse Environments”, *Communication of the ACM*, Vol.42, No.1, 2002, pp. 73-78.
- Brackett, Michael, “The Concept of Data Resource Data”, *Data Resource Management*, Vol.4, 2002, pp. 149-150.
- Kahn, Beverly L., Strong, Diane M. and Wang, Richard Y., “Information Quality Benchmarks: Product and Service Performance”, *Communication of the ACM*, Vol.45, 2002, pp. 184-192.
- Kelly, Sean, *Data Warehousing in Action*, Wiley, 1997.
- Miller, H., “The Multiple Dimensions of Information Quality”, *Information Systems Management*, Spring 1996, pp. 79-82.
- Orr, Ken, “Data Quality and Systems Theory”, *Communication of the ACM*, Vol.41, No.2, 1998, pp. 66-71.
- Pinto, J. K. and Slevin, D. P., “Critical Factors in Successful Project Implementation”, *IEEE Transactions on Engineering Management*, Vol.EM-34, No.1, 1987, pp. 22-27.
- Pipino, L.L, Lee, Y.W., Wang, R.Y., “Data Quality Assessment,” *Communication of the ACM*, Vol.45, No.4, 2004, pp. 211-218.
- PMI, A Guide to the Project Management Body of Knowledge(PMBOK Guide), Project Management Institute, 2004.
- Redman, Thomas C., “Impact of Poor Data Quality on the Typical Enterprise”, *Communication of the ACM*, Vol.41, No.2, 1998, pp. 79-82.
- Strong, Diane M., Lee, Yang W. and Wang, Richard Y., “Data Quality in Context”, *Communication of the ACM*, Vol.40, No.5, 1997, pp. 103-110.
- Wang, Richard Y. and Strong, Diane M., “Beyond Accuracy: What Data Quality Means to Data Consumers”, *Journal of Management Information System*, Vol.12, No.4, 1996, pp. 5-34.
- Wang, Richard Y., “A Product Perspective on Total Data Quality Management”, *Communication of the ACM*, Vol.41, No.2, 1998, pp. 58-65.
- Wang, Richard Y., Lee, Yang W., Pipino, Leo L. and Strong, Diane M., “Manage Your Information as a Product”, *Sloan Management Review*, Vol.39, No.4, 1998, pp. 95-105.
- Wang, Richard Y., Storey, Veda C. and Firth,

Christopher P., "A Framework for Analysis of Data Quality research", *IEEE Transactions on Knowledge and Data Engineering*, Vol.7, No.4, 1995, pp. 623-640.

Winkler, W. E., "Methods for Evaluating and

Creating Data Quality", *Information Systems*, Vol.29, 2004, pp. 531-550.

<http://dbgguide.net/>, 데이터베이스 구축 운영 종합정보.

<http://dbq.dpc.or.kr/>, 데이터 품질관리 인증센터.

A Case Study on Improvement of Data Management Process for Enhancing Data Quality: Focus on Data Standards and Requirement Management

Hee Joung Heh* · Jong Woo Kim**

Abstract

Recently, as most functional business activities in an enterprise are supported by computerized information systems, data duplication and inconsistency among functional information systems become serious problems. It brings people to have many interests on data quality management. This paper presents a case study in which a company had improved their data quality by enhancing their data quality management processes. Though the case study, we describe main issues and risk factors in the process of data quality improvement projects as well as solutions to resolve the issues, which can be referred by other companies who pursue data quality improvement. Also, the improvement effects are evaluated by multidimensional perspectives which include quantitative and qualitative measures on data quality, productivity, customer satisfaction, organization, and culture.

Keywords: Data Quality, Data Management Process, Data Standards Management, Requirement Management

* Assistant Manager, DB Management Division, koid Corp.

** Corresponding Author, Professor, School of Business, Hanyang University

◎ 저 자 소 개 ◎



허 희 정 (hhj0112@paran.com)

한양대학교 경영대학원 경영정보학과를 졸업하였다. 하나은행 전산정보부, (주) 핸디소프트 품질경영실에서 책임연구위원으로 사내 MIS 시스템 구축을 주도하였으며, 현재 (주)코이드 DB관리부에 재직 중이다. 주요 관심분야는 데이터 아키텍처, 데이터 품질관리 프로세스 개선, 데이터 통합, 대용량 데이터베이스 설계 및 구현 등이다.



김 종 우 (kju@hanyang.ac.kr)

한양대학교 경영대학 경영학부 교수로 재직 중이다. 서울대학교 수학과에서 학사(1989), 한국과학기술원 경영과학과에서 석사(1991), 한국과학기술원 산업경영학과(1995)에서 박사학위를 취득하였다. 충남대학교 통계학과 부교수, University of Illinois at Urbana-Champaign 방문연구원 등을 역임하였다. 주요 관심연구분야는 상품 추천 시스템, 데이터 마이닝 응용, 지능정보시스템, 의사결정 지원시스템, 비즈니스 프로세스 모델링 및 통합, 데이터 품질 등이다.

논문접수일 : 2007년 05월 30일
1차 수정일 : 2007년 09월 10일

게재확정일 : 2007년 10월 18일