

품질코스트를 이용한 데이터 QC 활동의 자원할당 모형 연구

이상철^{1*} · 신완선²

¹한국국방연구원 / ²성균관대학교 시스템경영공학과

A Resource Allocation Model for Data QC Activities Using Cost of Quality

SangCheol Lee¹ · WanSeon Shin²

¹Korea Institute for Defense Analyses

²Systems Management Engineering, Sungkyunkwan University

This research proposes a resource allocation model of Data QC (Quality Control) activities using COQ (Cost of Quality). The model has been developed based on a series of research efforts such as COQ classifications, weight determination of Data QC activities, and an aggregation approach between COQ and Data QC activities. In the first stage of this research, COQ was divided into the four typical classifications (prevention costs, appraisal costs, internal failure costs and external failure costs) through the opinions from five professionals in Data QC. In the second stage, the weights of Data QC activities were elicited from the field professionals. An aggregation model between COQ and Data QC activities has been then proposed to help the practitioners make a resource allocation strategy. DEA (Data Envelopment Analysis) was utilized for locating efficient decision points. The proposed resource allocation model has been validated using the case of Korea national defense information system. This research is unique in that it applies the concept of COQ to the data management for the first time and that it demonstrates a possible contribution to a real world case for budget allocation of national defense information.

Keywords: resource allocation, data quality control, cost of quality

1. 서론

오늘날 기업은 제품의 설계, 개발, 생산뿐만 아니라 고객관리, 유통망관리 등에 이르기까지의 모든 업무를 정보시스템으로 구축하여 운영하고 있다. 과거에는 기업의 업무 단위별로 개별 시스템을 구축하였으나, 글로벌 경쟁 환경에 대응하고자 개별 시스템을 통합하는 방향으로 나아가고 있다. 이러한 정보시스템의 구축을 통해서 기업은 급변하는 비즈니스 환경에

신속히 대처하고, 정확한 데이터를 기반으로 합리적인 의사결정이 수행되기를 기대하고 있다. 그러나 정보시스템의 기초가 되는 데이터의 부정확성 및 신뢰성 저하 등으로 인한 데이터의 품질 문제가 기업의 효율적인 운영과 신속한 의사결정을 저해하는 중요한 요인으로 부각되고 있다. 이러한 문제를 해결하기 위해서는 데이터 품질관리를 지속적으로 수행해야 한다. 하지만, 데이터 품질관리를 위해 수행하는 활동을 코스트 측면에서 평가할 수 있는 경제적이고 합리적인 척도가 없기에

* 연락저자 : 이상철, 130-866 서울시 동대문구 회기로 39번지 한국국방연구원,

Tel : 010-5072-6145, E-mail : yeji_father@hanmail.net

투고일(2010년 10월 26일), 심사일(1차 : 2010년 12월 08일, 2차 : 2010년 12월 16일), 게재확정일(2010년 12월 28일)

기업의 최고경영자는 데이터 품질관리의 필요성을 인식한다 하더라도 쉽게 투자 결정을 할 수 없으며, 또한 적절하게 투입했는지에 대해서도 알 수 없는 상황이다.

본 연구의 목적은 품질코스트 기법을 이용하여 데이터 품질 관리 활동을 품질코스트로 분류하고, 데이터 품질관리를 위한 효율적인 자원할당 모형을 개발하는 것이다.

2. 이론적 배경

2.1 데이터 품질관리의 개념

2.1.1 데이터 품질관리의 정의

데이터 품질에 대한 정의는 학자마다 의견이 다른데, Jack E. Olson(2006)은 데이터 품질을 “사용 목적을 위해 요구되는 사항들을 충족시키는 적합한 데이터 특성”으로 정의하였으며 데이터 품질이 가져야 할 첫 번째 요소로 정확성을 주장하였다. 그는 데이터 품질 보증 프로그램의 필요성 및 수행 방안을 제시하고 있으며, 이를 위해 데이터 프로파일링 기술을 소개하고 있다.

박주석(2003)은 데이터 품질을 “목적 달성을 위해 데이터가 사용자의 만족도를 충족시킬 수 있는 수준”으로 정의하였으며, Larry P. English(1999)는 데이터 품질을 “일관성 있게 고객(지식근로자 및 최종사용자)의 데이터에 대한 기대치를 충족시키는 것”으로 정의하였다.

Alan F. Karr(2006)은 데이터 품질을 “의사결정을 평가하고 알려주기 위해 효과적으로, 경제적으로 신속하게 사용되는 데이터의 능력”으로 정의하였다.

이처럼 데이터 품질은 데이터의 적합성, 정확성, 완전성, 적절성 및 접근 가능성 등을 통틀어 이르는 말이며, 데이터 품질을 지속적으로 유지하고 관리하기 위한 종합적인 관리 활동을 데이터 품질관리라 할 수 있다.

2.1.2 데이터 품질관리 프레임워크 및 활동

한국DB진흥원에서는 2006년 데이터 품질관리의 대상이 되는 각 구성요소들 간의 관계를 정의한 데이터 품질관리 프레임워크를 제시하였다. 데이터 품질관리의 대상을 데이터 값, 데이터 구조, 데이터 품질관리 활동으로 구분하고, 관리조직은 CIO(Chief Information Officer)/EDA(Enterprise Data Administrator), DA(Data Administrator), Modeler, DBA(Database Administrator), User로 구분하여 데이터 품질관리의 대상이 되는 각 구성요소들 간의 관계를 정의하였다.

또한, 데이터 품질관리 활동을 8개의 대분류 활동, 22개의 중분류 활동, 94개의 소분류 활동으로 구성하여 제시하였다.

표 1. 데이터 품질관리 프레임워크

대상 조직	데이터 값	데이터 구조	데이터 품질관리 활동
CIO/EDA	데이터 관리 정책		
DA	표준 데이터	개념데이터 모델 데이터참조 모델	데이터 표준관리 요구사항관리
Modeler	모델 데이터	논리데이터 모델	데이터 모델관리 데이터흐름 관리
DBA	관리 데이터	물리데이터 모델 데이터베이스	DB관리 DB보안관리
User	업무 데이터	사용자View	데이터 활용관리

표 2. 데이터 품질관리 활동

대분류	중분류	소분류
데이터 관리정책	데이터 관리정책 수립	데이터 관리정책 수립
		데이터 관리정책 검토
		데이터 관리정책 공표
	데이터 관리정책 변경	데이터 관리정책 변경 요구사항 검토
		데이터 관리정책 추가 및 변경
		데이터 관리정책 변경 사항 공표

(이하 생략)

데이터 품질관리 활동의 결과는 데이터 품질관리 성숙수준으로 평가될 수 있다. 한국DB진흥원에서는 데이터 품질관리 성숙수준을 도입, 정형화, 통합화, 정량화, 최적화의 5단계로 구성하고, 정확성, 일관성, 유용성, 접근성, 적시성, 보안성의 6가지 데이터 품질 기준별로 측정할 수 있도록 하였다.

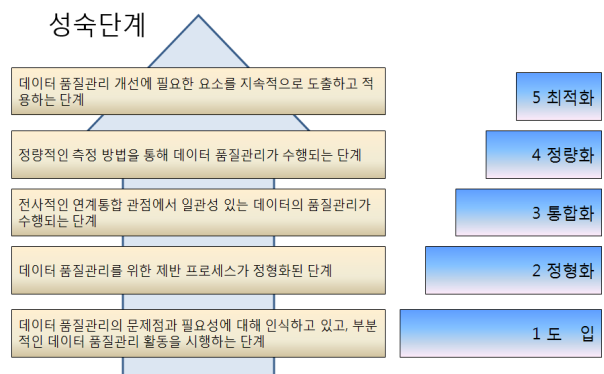


그림 1. 데이터 품질관리 성숙수준 단계

2.2 품질코스트 관련 연구

품질코스트와 관련된 연구는 주로 품질코스트가 최적이 되는 품질코스트 모델에 대한 연구 및 품질코스트간의 인과관계에 대한 연구가 주를 이루어 왔다. 그리고 국내에서는 산업별

품질코스트 항목을 도출하고 이를 실제로 적용한 사례 중심의 연구와 품질코스트 관리 시스템을 구축하는 연구로 수행되어 왔다.

Juran(1974)은 예방코스트와 평가코스트를 독립변수로 하고, 실패코스트를 종속변수로 하여 전체 품질코스트가 최저가 되는 최적의 품질수준 모형을 제시하였다. 또한, 전체 품질코스트 곡선을 개선영역, 중립영역, 완전영역으로 구분하여 영역별 품질코스트 개선책을 제시하였다.

Schneiderman(1986)은 100%의 적합 품질수준에서 품질코스트의 최적 배분이 이루어 질 수 있음을 제시하였다.

Harrington(1987)은 미국기업의 품질코스트 구성비율과 매출액 대비 품질코스트 비율에 대해서 조사, 분석하였으며, Crawsey(1976)는 총 품질코스트의 구성 비율에 있어서 기업성과 지수가 높을수록 예방코스트는 증가하고 실패코스트는 감소한다고 주장하며 품질코스트 간에는 적절한 구성 비율이 존재한다고 제시하였다.

Oppermann(2003)은 다이나믹 프로그래밍을 이용하여 전자산업분야에 적합한 품질코스트의 최적 모델에 관한 연구를 수행하였으며, 서주창(1996)은 제조 시스템에서의 품질코스트를 측정하는 방법에 대하여 제시하였다.

국내 연구를 살펴보면, 정영배(2004)는 전기·전자, 자동차, 기계, 제철, 화학·장치, 서비스, 건설과 같이 7개 업종별로 품질코스트 항목을 정의하고, 웹 기반의 품질코스트 관리 시스템을 설계 및 구축하였다.

강지호(1995)는 품질 코스트의 항목을 예방코스트, 평가코스트, 사내실패코스트, 사외실패코스트로 분류하고 항목별로 발생코스트별 산출방법을 구체적으로 제시하였으며, 이광재(2002)는 일본 K사의 숨겨진 코스트를 품질코스트에 중점을 두고 품질코스트 시스템 전개방법을 제시하였으며 이를 위하여 인과관계도 및 IDEFO 모형을 활용할 수 있다고 하였다.

품질코스트의 소프트웨어 응용 연구로 박동후(2007)는 국방 소프트웨어의 소프트웨어 개발 활동 내에서의 테스트 활동 및 제품의 품질에 대한 문제점을 품질코스트 모델을 통해서 분석하고 고품질의 제품을 개발하기 위한 테스트 프로세스 개선에 대한 동기 부여 및 개선방안을 제시하였으며, Herb Krasner(1998)은 제조 산업에서의 품질코스트가 소프트웨어 분야에서도 동일하게 적용됨을 CoSQ(Cost of Software Quality) 모델을 통해 제시하였다. 또한, Knox(1993)는 품질코스트 모델의 연구를 통한 소프트웨어 프로세스 개선 연구에서 CMM(Capability Maturity Model) 레벨에 따른 Total Cost of Software Quality(TCoSQ)를 제시하였다.

3. 품질코스트와 데이터 품질관리 활동 분석

3.1 전문가 그룹을 활용한 품질코스트와 데이터 품질관리 활동 분석

3.1.1 품질코스트별 데이터 품질관리 활동 정의

본 연구에서 적용한 데이터 품질관리 활동은 한국DB진흥원

에서 제시한 데이터 품질관리 활동으로 가정하며, 품질코스트 활동은 ASQ에서 제시한 품질코스트 분류를 사용한다. 연관관계 분석을 위해서 데이터 품질관리 활동을 세로축으로, 품질코스트 활동을 가로축으로 하여 강한 연관관계, 보통 연관관계, 약한 연관관계를 조사하였으며, 5명의 특급기술자 등급을 지닌 데이터 품질관리 전문가 그룹을 구성하여 델파이 기법에 따라 3라운드를 수행하였다.

총 94개의 데이터 품질관리 활동을 강한 연관관계를 중심으로 분석한 결과, 예방코스트(P) 활동 39개, 평가코스트(A) 활동 28개, 내부 실패코스트(IF) 활동 6개, 외부 실패코스트(EF) 활동 23개로 나타났으며 DB 성능개선과 데이터 보안 개선 활동은 예방코스트와 외부 실패코스트 모두에 강한 연관관계를 가지는 것으로 나타났다. 예방코스트와 외부 실패코스트의 활동으로 분류된 DB 성능개선과 데이터 보안 개선 활동은 사전 활동과 사후 활동으로 구분하여 각각 예방코스트와 외부 실패코스트로 분류할 수 있으나, 본 연구에서는 DB 성능 개선과 데이터 보안 개선 활동을 외부 실패코스트 활동으로 가정한다. 외부 실패코스트 활동으로 가정한 이유는, 2009년도 데이터 품질관리 성숙수준 보고서에 따르면 기관 및 기업의 데이터 품질관리 성숙수준이 평균 0.9로 도입 전 단계에 해당하기 때문에 모니터링 활동에 의한 개선보다는 사용자가 문제를 느끼고 개선책을 요구하는 형태로 DB 성능 개선과 데이터 보안 개선 활동이 수행될 것으로 예측되었기 때문이다.

표 3. 데이터 품질관리 활동과 품질코스트 간의 연관관계 분석 결과

분류	데이터 품질관리 활동
P (37)	데이터관리 정책수립, 데이터 관리정책공표, 표준화요구 사항수립, 표준화원칙정의, 표준단어사전정의, 표준도메인 사전정의, 표준코드정의, 표준용어사전정의, 데이터 관련 요소표준정의, 데이터 표준공표, 데이터 표준-데이터 모델 매핑, 개념데이터 모델정의, 데이터 참조 모델정의, 논리 데이터 모델정의, 물리 데이터 모델정의, 데이터 추출요건 정의, 소스 데이터 분석, 소스 데이터 추출설계, 소스 데이터 추출도출반영, 소스데이터 추출모니터링, 데이터 흐름점검 기준도출, 데이터 흐름점검지표생성, DB생성, 백업주기 및 스케줄 정의, DB백업수행, 데이터 보안대상선정, 데이터 보안 적용, 데이터보안 교육수행, 인증정책수립, 패스워드 정책, 인증관리, 접근수준 및 권한부여, 접근제어설정 갱신, 접근제어메커니즘, 암호화, 개선방안마련, 개선활동수행
A (28)	데이터 관리정책변경 요구사항도출, 데이터 표준검토, 데이터 표준준수 체크, 변경영향도분석, 데이터 표준미준수원인분석, 개념-논리모델 매핑, 논리-물리모델 매핑, 물리모델-DB매핑, 개념-논리모델 ALIGN 분석, 물리-DB ALIGN 분석, 소스데이터 추출테스트, 데이터정합성체크, 오류 데이터분석, 변경영향도분석, DB객체관리 효율성체크, 비효율원인분석, 변경영향도분석, 인증정책검토, 로깅, 감사, 핵심 데이터 수집, 데이터 활용도 측정 기준수립, 데이터 활용측정, 활용저하요인분석, 개선활동평가

분류	데이터 품질관리 활동
IF (6)	데이터 관리 정책변경 요구사항검토, 데이터 관리정책 추가 및 변경, 데이터 관리정책 변경사항공표, 데이터정제, 데이터흐름변경, 데이터 정제
EF (23)	변경요구사항검토, 표준추가 및 변경, 표준변경영향도 평가, 표준등록 및 공표, 개념 데이터 모델변경, 논리 데이터 모델변경, 물리데이터 모델변경, DB 성능개선, 데이터보안개선, DB복구, 테스트 DB변경, 운영DB이관, 변경계획수립, 사고접수, 사고등록 및 대응, 사고이력관리 및 보고, 보안사고 사례교육, 변경요청, 요구사항수립, 요구사항검토, 변경영향도분석, 공식화, 변경작업계획수립

3.1.2 데이터 품질관리 중요도 분석

데이터 품질관리 활동은 기업의 규모, 데이터 품질관리에 대한 인식 수준, 경영 환경, 데이터 품질관리 전담팀의 유무 등에 따라서 서로 다른 중요도를 지니게 된다. 이러한 사항은 특정 데이터 품질관리 활동이 다른 데이터 품질관리 활동에 비해서 상대적인 중요도를 가질 수 있으며 이는 각 활동별 가중치로 환산될 수 있음을 의미한다.

데이터 품질관리 활동별 중요도 또한 연관관계와 동일하게 전문가 그룹을 활용하여 조사하였다. 데이터 품질관리 활동의 중요도 조사 시 데이터 품질관리 소분류 활동은 데이터 품질관리를 수행하는 가장 작은 단위 활동으로, 데이터 품질관리 중분류 활동에 따른 필수적인 활동이며 선택적인 활동으로 구분할 수 없다는 전문가 그룹의 의견에 따라 중분류 활동까지만 중요도를 조사하였다. 즉, 소분류 활동에 대한 가중치는 중분류 활동의 가중치를 소분류 활동의 개수만큼 나눠서 산정하면 된다.

가중치 조사를 위한 세부적인 3단계 절차는 다음과 같다.

- 1단계 : 데이터 품질관리 대분류 활동별로 가중치를 조사한다.
- 2단계 : 데이터 품질관리 중분류 활동별로 가중치를 조사한다. 중분류 활동에 대한 가중치 조사는 대분류 활동별로 중분류 활동의 가중치의 합이 1이 되도록 한다.
- 3단계 : 데이터 품질관리 중분류 활동에 대해서 전체 중분류에 대한 상대적 가중치를 구한다. 계산 결과는 (1)에서 조사한 대분류 활동의 가중치 값(A)에 (2)에서 조사한 중분류 활동의 가중치 값(B)을 곱한다.

데이터 품질관리 활동별 가중치는 품질관리 활동에 대한 상대적인 중요도를 파악할 수 있는 자료이다. 또한, 데이터 품질관리 활동의 개선방안 수립 시 활용될 수 있으며, 예산의 한정으로 인해 데이터 품질관리 활동을 모두 수행하기가 곤란한 경우 데이터 품질관리 활동별 가중치에 따라 데이터 품질관리 활동의 우선순위를 부여하여 선택적으로 적용할 수 있다.

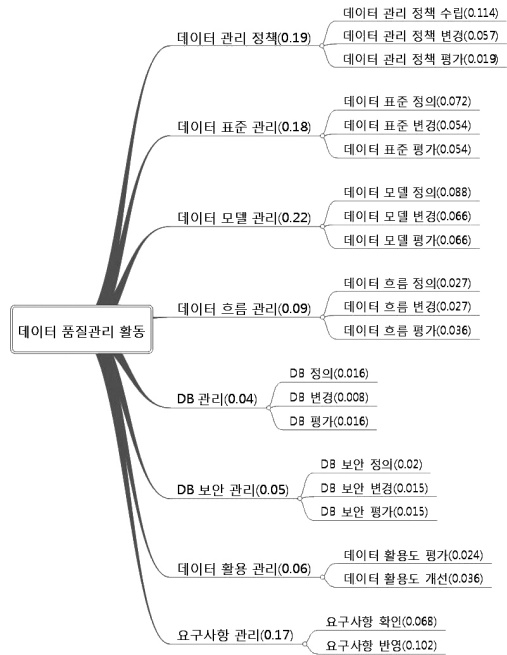


그림 2. 데이터 품질관리 활동의 가치 구조

3.2 국방정보화 분야의 데이터 품질관리 분석

3.2.1 설문조사 방법

설문조사는 국방정보화 분야의 데이터 품질관리 중요도 분석과 데이터 품질관리 수준 분석을 위하여 최소 5년 이상의 국방정보화 업무 경력을 지닌 K연구소 직원을 대상으로 수행하였다.

데이터 품질관리 중요도 분석을 위해서 각 활동별로 총 28개의 설문항목을 구성하였으며, 각 문항별로 좌측의 데이터 품질관리 활동과 우측의 데이터 품질관리 활동을 쌍대 비교하도록 설계하였다. 예를 들어, 응답자가 좌측 데이터 품질관리 활동의 숫자 5에 체크를 하였다면, 이는 좌측의 데이터 품질관리 활동이 우측의 데이터 품질관리 활동보다 5배 중요하다는 것을 의미한다.

표 4. 데이터 품질관리 중요도 분석 설문지

활동	중요도	활동
데이터 관리정책	98765432123456789	데이터 표준관리
데이터 관리정책	98765432123456789	데이터 모델관리
(중략)		
데이터 활용관리	98765432123456789	요구사항 관리

데이터 품질관리 수준 분석은 기대 수준과 인지 수준이 보통인 경우는 4점, 매우그렇다인 경우는 7점, 전혀아니다인 경우

는 1점을 체크하도록 Likert의 7단계 척도를 이용하여 설문을 수행하였다.

표 5. 데이터 품질관리 수준 분석 설문지

활동	기대수준 (원하는정도)	인지수준 (실제성과)
	1 ← 4 → 7	1 ← 4 → 7
데이터 관리정책	1 2 3 4 5 6 7	1 2 3 4 5 6 7
데이터 표준관리	1 2 3 4 5 6 7	1 2 3 4 5 6 7
(중략)		
요구사항 관리	1 2 3 4 5 6 7	1 2 3 4 5 6 7

3.2.2 국방정보화 분야의 데이터 품질관리 중요도 분석

데이터 품질관리 활동의 상대적인 중요도를 파악하는 것은 국방정보화 분야에 있어서 데이터 품질관리의 우선순위를 정하기 위한 기준이 된다.

설문 분석은 AHP 응용 프로그램인 Expert Choice 2000을 통해 분석하였다. 일관성 지수의 기준을 0.1로 할 때, 총 32명의 설문 응답자 중 28명은 일관성을 만족하였으며, 4명은 일관성이 부족한 것으로 나타났다. 따라서, 일관성을 만족하는 28명을 대상으로 AHP 분석을 수행하였으며, 종합 일관성 지수는 0.0047로 일관성을 만족하는 것으로 나타났다.

표 6. 국방정보화 분야의 데이터 품질관리 활동별 상대적 가중치

데이터 품질관리 활동	가중치
데이터 관리정책	0.2230
데이터 표준관리	0.2080
데이터 모델관리	0.1383
데이터 흐름관리	0.0723
DB 관리	0.0886
DB 보안관리	0.0788
데이터 활용관리	0.0805
요구사항 관리	0.1105
합계	1

분석 결과, 국방정보화 분야에서는 데이터 관리 정책과 데이터 표준 관리를 가장 중요하게 생각하고 있으며, 데이터 흐름 관리를 가장 중요도가 낮은 활동으로 평가하고 있음을 알 수 있다.

전문가 그룹을 활용한 데이터 품질관리 활동의 가중치와 국방정보화 분야의 설문조사를 통한 조사 결과를 비교하면 <표 7>과 같다.

차이 분석 결과, 전문가 그룹에서는 데이터 모델 관리를 가장 중요한 활동으로 판단하고 있지만 국방정보화 분야에 있어서는 상대적인 중요도가 낮게 나타났다. 이런 차이를 보면, 데이터 품질관리 활동의 상대적인 중요도는 각 기업이나 기관의

고유한 설립목적과 운영방식, 그리고 현재 처한 상황에 따라 상이할 수 있음을 알 수 있다.

표 7. 데이터 품질관리 활동 가중치 비교

데이터 품질관리 활동	전문가 그룹 (A)	국방 분야 (B)	차이값 (A-B)
데이터 관리 정책	0.19	0.2230	-0.0330
데이터 표준 관리	0.18	0.2080	-0.0280
데이터 모델 관리	0.22	0.1383	0.0817
데이터 흐름 관리	0.09	0.0723	0.0177
DB 관리	0.04	0.0886	-0.0486
DB 보안 관리	0.05	0.0788	-0.0288
데이터 활용 관리	0.06	0.0805	-0.0205
요구사항 관리	0.17	0.1105	0.0595

앞서의 결과를 종합해볼 때, 국방정보화 분야에 있어 데이터 품질관리 활동은 민간의 방법과 노하우를 최대한 수용하되 데이터 품질관리 활동의 우선순위에 있어서는 국방의 현실을 고려하여 정책을 수립하고 추진 방향을 모색해야 한다.

3.2.3 국방정보화 분야의 데이터 품질관리 수준 분석

국방정보화 분야의 데이터 품질관리 수준 분석을 위해서 IP(Importance and Performance) 분석을 활용하였으며, 총 30명의 설문결과가 접수되었다.

분석 결과, 기대수준 면에서는 데이터 표준 관리, DB 보안 관리, 데이터 관리 정책 순으로 높게 나왔다. 인지수준 면에서는 DB 보안 관리, DB 관리 순으로 높게 나타났으며, 데이터 활용 관리가 매우 저조하게 나타났다. 차이분석 결과를 보면, 데이터 표준 관리 분야가 가장 큰 차이를 나타내고 있으며, 데이터 관리 정책, 데이터 모델 관리, 요구사항 관리 순으로 차이가 크게 나타났다.

표 8. 데이터 품질관리 활동 수준 분석 결과

구분 \ 활동구분	기대 수준	인지 수준	차이 (인지-기대)
데이터 관리정책	5.77	2.90	-2.87
데이터 표준관리	5.87	2.83	-3.04
데이터 모델관리	5.60	2.73	-2.87
데이터 흐름관리	5.10	2.57	-2.53
DB 관리	5.63	3.27	-2.36
DB 보안 관리	5.83	3.67	-2.16
데이터 활용관리	5.10	2.43	-2.67
요구사항 관리	5.63	2.93	-2.70
평균	5.57	2.92	-2.65

기대수준과 인지수준 결과를 IP격자에 표현하면 다음과 같다. 그림 상의 번호는 데이터 품질관리 활동을 순서대로 표시한 사항이다.

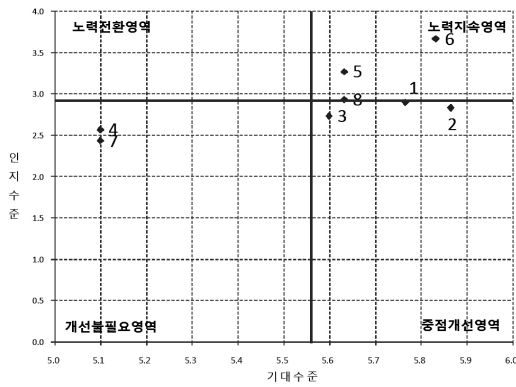


그림 3. IP 분석 결과

국방정보화 분야에 있어서 가장 시급히 개선해야 하는 사항으로는, IP격자상 중점개선영역에 할당된 데이터 관리 정책, 데이터 표준 관리, 데이터 모델 관리로 나타났다. 노력지속영역으로는 DB 관리, DB 보안 관리, 요구사항 관리로 나타났으며, 특히 요구사항관리의 만족도는 만족도 평균값을 미세하게 초과하여 노력지속영역으로 분류되었지만 관심 있게 주의를 기울여야 할 사항으로 판단된다.

4. 품질코스트를 이용한 데이터 품질관리 활동의 자원할당

4.1 DEA 모형

비모수적 효율성 측정방법인 DEA는 다른 효율성 측정방법과는 다르게 사전에 구체적인 함수형태를 가정하고 모수를 추정하는 것이 아니라 선형계획법에 근거하여 평가대상의 경험적인 투입요소와 산출물 간의 자료를 이용하여 경험적 효율 프론티어를 도출한 후 평가대상들이 효율적 프론티어로부터 얼마나 떨어져 있는지의 여부로 비효율성을 측정하는 기법이다. 이 방법은 Charnes, Cooper and Rhodes(1978)가 Farrel의 상대적인 효율성 개념을 새로이 해석하고 이를 다수 투입물과 다수 산출물과의 비율모형(CCR Ratio)으로 연장하여 비선형계획법으로 나타내었다. 이 모형을 DEA 모형이라 하며, 비영리 프로그램의 상대적 효율성을 평가하기 위해서 개발되어 미국 공립학교 교육에 관한 실험연구인 Project Follow Through에 처음 적용되었다.

일반적으로 DEA 모형 중에서 가장 많이 활용되는 모형은 Charnes, Cooper and Rhodes(1978)의 CCR 모형과 Banker, Charnes and Cooper(1984)의 BCC 모형이다. 또한 이 두 모형은 투입요소에 초점을 두는가, 산출물에 초점을 두는가에 따라 투입지향(Input Oriented)과 산출지향(Output Oriented)으로 구별된다.

투입지향모형은 DMU의 산출물 수준이 주어졌을 때, 현재의 산출수준을 유지하면서 투입요소의 사용량을 얼마나 줄일 수 있는가를 파악하고자 하는 경우에 사용하며, 산출지향모형은 DMU에 의해서 사용된 투입요소의 수준이 주어졌을 때, 달성해야 할 산출수준을 파악하고자 하는 경우에 사용한다.

표 9. 투입지향모형과 산출지향모형 비교

모형	요약	
투입지향	효율성 = $\frac{\text{산출}}{\text{투입}}$	← 산출고정 ← 투입최소화
산출지향	효율성 = $\frac{\text{산출}}{\text{투입}}$	← 산출최대화 ← 투입고정

DEA 모형에 대한 판별력은 DMU의 수로 결정이 된다. DEA 모형의 판별력과 관련된 선행연구 중 대표적인 연구로 Banker et al.(1984)은 평가대상인 DMU의 수가 투입요소 수와 산출요소 수의 합보다 3배 이상 커야 변별력이 있다고 검증하였고, Boussofinance et al.(1991)은 평가대상인 DMU의 수가 투입요소의 수와 산출요소수의 곱보다 2배 이상 커야 변별력이 있다고 주장하였으며, Fitzsimmons(1994)는 평가대상 DMU의 수가 투입요소의 수와 산출요소 수의 합보다 2배 이상 커야 변별력이 있다고 하였다. 실제로 DEA 모형을 적용한 많은 연구들의 대부분이 DMU의 수가 투입요소의 수와 산출요소 수의 합보다 2배 이상 커야 변별력이 있다는 기준을 사용하고 있다.

4.2 품질코스트를 이용한 데이터 품질관리 활동의 자원할당

4.2.1 연구모형

데이터 품질관리 활동을 수행하기 위해서는 전문성을 갖춘 인력의 투입, 예산 할당, 관련 소프트웨어 및 하드웨어 등의 다양한 자원이 요구된다. 대부분의 프로젝트와 유사하게 데이터 품질관리 활동에 있어서도 가장 중요하게 고려해야 할 요소는 전문성을 지닌 인력의 투입과 적절한 예산의 분배 문제이다.

본 연구에서는 이러한 문제를 해결하고자 다음과 같이 데이터 품질관리 활동의 자원할당 모형을 개발하였다.

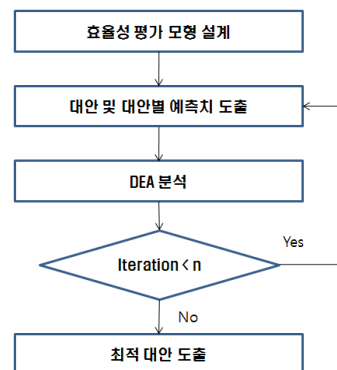


그림 4. 데이터 품질관리 활동의 자원할당 모형

데이터 품질관리 활동의 자원할당 모형은 4단계로 구성된 다. 모형의 세부 단계별 기능을 설명하면 다음과 같다.

(1) 1단계 : 효율성 평가 모형 설계

품질코스트를 투입요소로 하여 데이터 품질관리 자원할당의 효율성을 평가하기 위한 모형을 설계한다.

(2) 2단계 : 대안 및 대안별 예측치 도출

설계된 효율성 평가 모형에 따라 자원할당 대안을 작성하고 각 대안별로 예측치를 도출한다. 대안 작성은 Interval Halving 방법을 이용하며, 예측치 조사는 설문조사, 담당자 면담, 전문가 조사, 델파이 기법 등을 활용한다. 작성된 각각의 대안은 Decision Making Unit이 되며, DMU1, DMU2, ..., DMUn으로 부여한다.

(3) 3단계 : DEA 분석

DEA 분석을 통해서 효율적인 대안을 선정한다. 효율적 대안 중 최적의 대안을 찾고 최적의 대안을 포함하는 이웃 구간을 min, max 구간으로 재설정한다. 사용자가 설정한 반복 횟수를 초과한 경우는 4단계로 진행하며, 그렇지 않은 경우는 2단계와 3단계 작업을 반복 수행한다.

(4) 4단계 : 최적 대안 도출

DEA 분석 결과를 토대로 최적의 대안을 포함한 구간을 제시하거나 최적의 대안을 선정한다.

4.2.2 가상의 데이터를 활용한 데이터 품질관리 활동의 자원 할당

(1) 1단계 : 효율성 평가 모형 설계

본 연구의 효율성 평가 모형은 예방코스트와 평가코스트에 대한 투입이 품질 불량을 줄일 수 있고 이는 제품의 품질수준을 향상시킬 수 있다는 품질코스트의 기본적 개념을 적용하여 설계하였다. 즉, 데이터 품질관리 활동에 있어서 예방과 평가 활동에 해당하는 코스트를 투입함으로써 데이터 품질관리 성숙수준이 향상될 것이라는 개념이다.

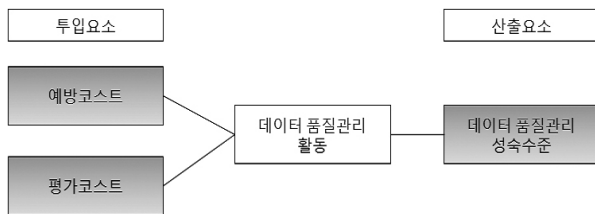


그림 5. 효율성 평가 모형

(2) 2단계 : 자원할당 대안 작성 및 예측치 도출

투입요소별 자원할당 대안 작성은 1단계에서 설계된 효율성 평가 모형에 따라 투입요소별로 자원할당 대안을 세부적으로 작성하는 단계이다. 투입요소에 대한 대안은 다양하게 작성할 수 있다. 투입요소에 대한 예산을 균일하게 주는 방법, 가중치별로 할당하는 방법, 과거의 경험치에 의해 할당하는 방법 등 다양한 대안이 작성된다. 각각의 대안은 DMU1부터 DMUn까지 부여하며, DMU의 수는 투입요소와 산출요소를 합한 값보다 최소 2배 이상이 되도록 작성한다.

본 연구에서는 자원할당 대안을 작성하기 위해서 Interval Halving 방법을 적용한다. Interval Halving 방법은 구간 사이에

총 3개의 점을 비교하여 구간을 줄여나가는 방법이나 DEA 모형의 판별력을 만족시키기 위해 최소 6개의 대안이 필요한 단계로 구간 사이에 총 4개의 점을 추가로 비교하는 방법을 이용하였다.

표 10. 예산 배정 1차 대안(단위 : 억 원)

대안	예방코스트	평가코스트
DMU1	1	9
DMU2	2.6	7.4
DMU3	4.2	5.8
DMU4	5.8	4.2
DMU5	7.4	2.6
DMU6	9	1

총 10억 원을 예방코스트와 평가코스트로 Interval Halving 방법을 이용하여 배정한 결과는 <표 10>과 같으며, 각 대안별로 데이터 품질관리 활동별 세부적 예산 배분을 수행할 수 있다. DMU5를 예를 들면, 데이터 관리 정책 수립 활동은 예방코스트 활동으로 분류되었기에, 예방코스트 7.4억 원에 활동별 가중치 0.038을 곱하고 이를 전체 예방코스트 활동의 가중치 합계인 0.4311로 나누면 0.65억 원이 할당된다.

자원할당 대안이 작성되면 산출요소에 대한 대안별 예측치 도출을 수행한다. 대안별 예측치 도출은 <표 10>에서 제시한 예산 배정 대안에 대해서 산출요소의 값인 데이터 품질관리 성숙수준을 조사하는 과정이다. 대안별 산출요소의 값을 도출하기 위해서는 각 대안별 소분류 활동에 대한 예산 배분 결과를 기초로 설문조사, 담당자 면담, 전문가 조사, 델파이 기법 등을 활용하여 데이터 품질 기준별 데이터 품질관리 성숙수준을 조사하여 평균값을 적용한다.

표 11. 대안별 데이터 품질관리 성숙수준 예측치 1차 결과

대안	투입요소		산출요소
	예방 코스트	평가 코스트	데이터 품질관리 성숙수준 예측치
DMU1	1	9	0.78
DMU2	2.6	7.4	1.2
DMU3	4.2	5.8	1.82
DMU4	5.8	4.2	2.81
DMU5	7.4	2.6	3.2
DMU6	9	1	2.76

(3) 3단계 : DEA 분석

DEA 분석은 2단계에서 작성된 투입요소와 산출요소에 대한 데이터를 DEA 모형을 통해서 분석하는 단계이다.

2단계에서 제시한 6개의 대안을 DEA로 분석하기 위해서 산출지향 CCR-DEA 모형을 적용하였다. 산출지향 CCR-DEA 모형을 적용한 이유는, 투입요소인 예방코스트와 평가코스트는 통제 가능한 요소인데 반하여 이에 대한 성과로 나타나는 데이터

품질관리 성숙수준 예측치는 통제 불가능한 요소이고 산출요소를 최대화 하는 것이 본 문제의 상황에 적합하기 때문이다.

본 연구에서는 DEA 문제를 풀기 위해서 박만희가 개발한 “효율성 및 생산성분석 시스템”을 이용하였다.

DEA 결과에 대한 분석은 효율성 지수를 기준으로 판단한다. 효율성 지수가 1인 DMU는 효율적인 대안이 되고, 1보다 작은 DMU는 비효율적인 대안이 된다.

표 12. DEA 분석결과-1차

대안	효율성 지수	산출 부족분	투영점	효율성 참조 집합
DMU1	1	0	0.78	1
DMU2	0.8238	0.257	1.457	1,4
DMU3	0.8531	0.313	2.133	1,4
DMU4	1	0	2.81	4
DMU5	1	0	3.2	5
DMU6	1	0	2.76	6

<표 12>를 살펴보면, 효율성이 1인 대안은 DMU1, DMU4, DMU5, DMU6이 된다. 이 중 DMU5는 산출지향모형에 따라 가장 효율적인 대안이 된다. <표 12>의 투영점을 살펴보면 DMU5를 중심으로 DMU4와 DMU6사이에 예산 할당의 최적해가 있음을 예상할 수 있다.

Interval Halving 방법에 따라 DMU4와 DMU6을 구간으로 재설정하여 대안을 구성하면 <표 13>과 같다. 반복 횟수를 만족한 경우 4단계로 넘어가고 반복 횟수를 만족하지 못한 경우, <표 13>을 근거로 2단계와 3단계 작업을 재수행한다.

표 13. 데이터 품질관리 활동 예산 배정 대안-2차

대안	예방코스트	평가코스트
DMU1	5.8	4.2
DMU2	6.4	3.6
DMU3	7.1	2.9
DMU4	7.7	2.3
DMU5	8.4	1.6
DMU6	9	1

표 14. DMU별 DEA 최종 결과

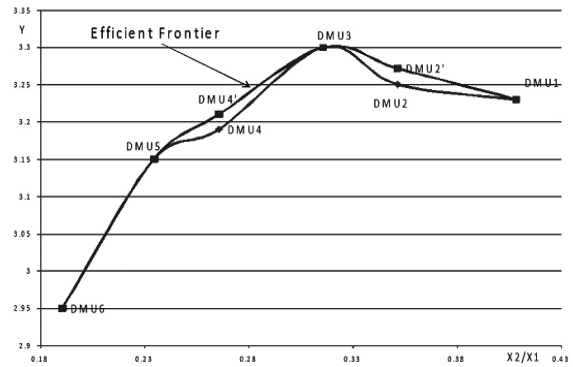
대안	투입		산출 Y	효율성 CRS	산출 부족분	투영점	효율성 참조집합	변수					
	X ₁	X ₂						λ ₁	λ ₂	λ ₃	λ ₄	λ ₅	λ ₆
DMU1	7.1	2.9	3.23	1	0	3.23	1	1	0	0	0	0	0
DMU2	7.4	2.6	3.25	0.9933	0.022	3.272	1,3	0.4	0	0.6	0	0	0
DMU3	7.6	2.4	3.3	1	0	3.3	3	0	0	1	0	0	0
DMU4	7.9	2.1	3.19	0.9938	0.02	3.21	3,5	0	0	0.4	0	0.6	0
DMU5	8.1	1.9	3.15	1	0	3.15	5	0	0	0	0	1	0
DMU6	8.4	1.6	2.95	1	0	2.95	6	0	0	0	0	0	1

(4) 4단계 : 최적 대안 도출

최적 대안 도출은 반복 횟수를 만족하는 경우에 3단계의 DEA 분석 결과를 토대로 최적의 대안 구간을 도출하거나 또는 최적의 대안을 선정하는 단계를 말한다. 예를 들어, 반복 횟수를 만족한 경우의 DEA 분석 결과가 <표 14>라고 가정해보자.

분석결과에 따르면, DMU1, DMU3, DMU5, DMU6은 효율성 지수가 1로써 효율적인 대안으로 나타났다. 비효율적인 대안으로 판명된 DMU2는 DMU1과 DMU3을 참조해야하며, DMU4는 DMU3과 DMU5를 참조해야 하는 것으로 나타났다.

참조집합으로 가장 많은 빈도수를 보인 대안은 DMU3으로 나타났으며, 이들을 참조집합으로 하는 DMU4의 경우는 데이터 품질관리 성숙수준이 3.21이 나와야 하지만 3.19의 예측치로 나타나 0.02만큼의 산출부족분이 발생하여 비효율적인 대안으로 분석되었다.



<그림 6> DEA 결과 그래프

<표 14>에서 제시한 DMU2의 λ 값을 보면 λ₃이 0.6이고 λ₁이 0.4로 DMU1보다는 DMU3에 보다 큰 방향성을 가지고 있는 것으로 나타났다. DMU4는 λ₃이 0.4이고 λ₅이 0.6으로 DMU3보다는 DMU5에 보다 큰 방향성을 가지고 있는 것으로 나타났다. λ 값을 살펴보면, 예산 10억 원을 할당하는 문제에 있어 최적의 해가 DMU2와 DMU3사이에 위치함을 알 수 있다. 이는 <그림 6>를 통해서도 분명히 알 수 있다. X축을 투입요소인 X₂/X₁으로 하고 Y축을 데이터 품질관리 성숙수준 예측치와 투영점으로 하여 그래프를 그리면 투영점을 중심으로 한 빨간색 선이 Efficient Frontier가 되며 그래프상 DMU3 근처에 최적의 해가 존재함을 알 수 있다.

4.3 국방정보화 분야 적용 사례

본 절에서는 제 4.2절에서 제시한 데이터 품질관리 활동의 자원할당 모형을 국방정보화 분야에 실제로 적용해 봄으로써 모형의 실효성을 검증한다.

데이터 품질관리 활동에 6.7억 원을 투입한다고 가정하고, 투입요소별 자원할당 대안을 Interval Halving 방법에 따라 도출하였으며, 도출된 자원할당 대안은 DMU1부터 DMU6까지 부여하였다. 또한, 각 산출요소에 대한 대안별 예측치는 군의 표준화 담당부서와의 인터뷰를 통해서 도출하였으며, 데이터 품질관리 성숙수준 예측치를 조사하기 위해서 데이터 품질관리 활동의 상세 예산 배분 내역을 각 대안별로 제공하였다. 마지막으로, 반복 횟수는 3회로 가정하였다.

(1) 1차 수행 결과

투입요소별 자원할당 대안과 대안별 데이터 품질관리 성숙수준의 예측치 1차 결과는 <표 15>와 같이 도출되었다.

DEA 분석 모형은 투입요소가 고정되어 있고 산출요소를 최대화해야 하는 상황에 따라 산출지향 CCR-DEA 모형을 적용하여 분석하였다.

1차 DEA 분석 결과, DMU1, DMU2, DMU3, DMU5가 효율적인 대안으로 도출되었으며 DMU1의 투영점이 3.5로 가장 높아 가장 효율적인 대안으로 파악되었다.

표 15. 국방정보화 분야의 예산 할당 대안과 데이터 품질관리 성숙수준 예측치 1차 결과

대안	투입요소		산출요소
	예방 코스트	평가 코스트	데이터 품질관리 성숙수준 예측치
DMU1	6.70	0.00	3.5
DMU2	5.36	1.34	3.35
DMU3	4.02	2.68	2.95
DMU4	2.68	4.02	1.8
DMU5	1.34	5.36	1.3
DMU6	0.00	6.70	0

(2) 2차 수행 결과

Interval Halving 방법에 따라, 1차 DEA 분석 결과의 DMU1을 최대값으로 DMU2를 최소값으로 설정하여 대안을 재구성하고 각 대안별 데이터 품질관리 성숙수준 예측치를 조사하여

표 16. 국방정보화 분야 1차 DEA 분석 결과

대안	투입		산출	효율성	산출 부족분	투영점	효율성 참조집합	변수						
	X_1	X_2	Y	CRS				λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	
DMU1	6.7	0	3.5	1	0	3.5	1	1	0	0	0	0	0	0
DMU2	5.36	1.34	3.35	1	0	3.35	2	0	1	0	0	0	0	0
DMU3	4.02	2.68	2.95	1	0	2.95	3	0	0	1	0	0	0	0
DMU4	2.68	4.02	1.8	0.8471	0.325	2.125	3,5	0	0	0.5	0	0.5	0	0
DMU5	1.34	5.36	1.3	1	0	1.3	5	0	0	0	0	0	1	0
DMU6	0	6.7	0	0	0	0	6	0	0	0	0	0	0	1

표 17. 국방정보화 분야 2차 DEA 분석 결과

대안	투입		산출	효율성	산출 부족분	투영점	효율성 참조집합	변수						
	X_1	X_2	Y	CRS				λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	
DMU1	6.7	0	3.5	1	0	3.5	1	1	0	0	0	0	0	0
DMU2	6.47	0.23	3.7	1	0	3.7	2	0	1	0	0	0	0	0
DMU3	6.2	0.5	3.93	1	0	3.93	3	0	0	1	0	0	0	0
DMU4	5.94	0.76	3.83	1	0	3.83	4	0	0	0	1	0	0	0
DMU5	5.67	1.03	3.67	1	0	3.67	5	0	0	0	0	0	1	0
DMU6	5.36	1.34	3.35	0.9656	0.119	3.469	5	0	0	0	0	0	0.9453	0

표 18. 국방정보화 분야 3차 DEA 분석 결과

대안	투입		산출	효율성	산출 부족분	투영점	효율성 참조집합	변수						
	X_1	X_2	Y	CRS				λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	
DMU1	6.47	0.23	3.7	1	0	3.7	1	1	0	0	0	0	0	0
DMU2	6.36	0.34	3.85	1	0	3.85	2	0	1	0	0	0	0	0
DMU3	6.26	0.44	3.95	1	0	3.95	3	0	0	1	0	0	0	0
DMU4	6.15	0.55	3.86	0.9875	0.049	3.909	3,6	0	0	0.6563	0	0	0	0.3438
DMU5	6.05	0.65	3.85	0.9945	0.021	3.871	3,6	0	0	0.3438	0	0	0	0.6562
DMU6	5.94	0.76	3.83	1	0	3.83	6	0	0	0	0	0	0	1

분석한 결과는 <표 17>과 같다.

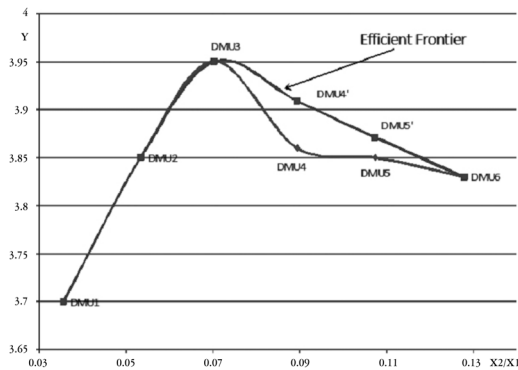
2차 DEA 분석 결과, DMU1, DMU2, DMU3, DMU4, DMU5가 효율적인 대안으로 도출되었으며 DMU3의 투영점이 3.93으로 가장 높아 가장 효율적인 대안으로 파악되었다.

(3) 3차 수행 결과

동일하게, DMU2와 DMU4의 구간을 대상으로 대안을 재구성하고 각 대안별 데이터 품질관리 성숙수준 예측치를 조사하여 분석한 결과는 <표 18>과 같다.

3차 DEA 분석 결과, DMU1, DMU2, DMU3, DMU6이 효율적인 대안으로 도출되었으며 DMU3의 투영점이 3.95로 가장 높아 가장 효율적인 대안으로 파악되었다.

<표 18>을 그래프로 표현하면 <그림 7>과 같다. 그래프 상의 빨간색 선이 Efficient Frontier가 되며 DMU3을 중심으로 한 부근에 최적의 해가 있을 것으로 분석되었다.



<그림 7> 국방정보화 분야의 DEA 결과 그래프

4.3.3 기대효과와 제약사항

국방정보화 분야에 데이터 품질관리 자원할당 모형을 실제로 적용해본 결과, 직관적 방법이나 과거 경험적인 방법이 아닌 과학적인 방법으로 예산을 할당할 수 있고 데이터 품질관리 활동을 화폐가치로 표현함으로써 데이터 품질관리 활동별로 계획 대비 집행 실적을 분석할 수 있다는 의견이 제시되었다.

이러한 기대효과 이외에도 예산 할당에 대한 과학적 근거 확보는 좋지만 예산 할당 결과를 지휘관 및 업무 담당자 등에게 설명할 때는 너무 난해하다는 제약사항이 일부 도출되었다.

5. 결론 및 향후 연구과제

본 연구에서는 품질코스트와 데이터 품질관리 활동 간의 관계를 분석하고, 품질코스트를 이용하여 데이터 품질관리 활동의 자원할당 모형을 제시하였다. 이러한 새로운 접근방식을 시도하는 목적은 그간 주관적 판단 하에 이루어졌던 자원배분

를 상대적이고 미래지향적 관점으로 발전시키는 것이 필요하기 때문이다.

향후 연구과제로는 데이터 품질관리의 예방 및 평가코스트에 대한 투자가 실패코스트에 어느 정도 영향을 미치며, 예방 및 평가코스트의 투자 증대로 인한 기대효과는 어느 정도인지를 파악할 수 있는 실증적인 연구가 요구된다.

참고문헌

Schneiderman, A. M. (1986), *Optimum Quality Costs and Zero Defects : Are They Contradictory Concepts?*, Quality Progress.

Feigenbaum, A. V. (1961), *Total Quality Control, 2nd ed.*, McGraw-Hill, Chapter 5.

Alan, F. Karr, Ashish P. Sanil., and David L. Banks (2006), *Data Quality : A Statistical Perspective*, Science Direct.

Bank, B. (1989), *Principles of Quality Control*, Singapore, 499.

Harrington, H. J. (1987), *Poor Quality Costs*, American Society For Quality Control, 5.

Herb Krasner (1998), *Using the Cost of Quality Approach for Software*, Crosstalk (The journal of Defense Software Engineering).

Grocock, J. M. (1986), *Chain of Quality: Market Dominance Through Product Superiority*, John Wiley and Sons.

Juran, J. M. and Gryna, F. M. (1993), *Quality Planning and Analysis*, 3rd ed., New York : McGraw-Hill, 19.

Juran, J. M. (1974), *Quality Control Handbook, 3rd edition*, New York : McGraw-Hill, 7.

Jack, E. Olson(2006), *Data Quality*.

Jeong, Y-B., Kim, Y-S., and Kim, J-H. (2004), *Development of Web Based Q-Cost System*, Journal of Korea Industrial and Systems Engineering.

Kang, J-H. (1995), *A Study on Classification and Calculation Method of Cost of Quality*, *Journal of Korea Industrial and Systems Engineering*, 18(35), 17-24.

Korea Database Agency (2006), *Data Quality Management Maturity Model (1.0)*.

Korea Database Agency (2006), *The Guideline for Data Quality Management (2.1)*.

Larry, P. English (1999), *Improving Data Warehouse and Business Information Quality : Methods for reducing costs and increasing profits*, Wiley computer publishing.

Lee, K-J. (2002), *A Case Study on Q-Cost System Utilization Method*, Management Study 11.

Oppermann, M., Sauer W., and Wohlrabe, H. (2003), *Optimization of quality costs*, Robotics and Computer-Integrated Manufacturing, 19(1-2), 135-140.

Park, D-H., Ryu H-Y., and Kim, J-S. (2007), *Test Process Improvement by Using Cost of Quality in National Defense Software Development*, Korea Computer Congress.

Park, J-S. and Kim, C-S. (2003), *Data Quality Management Maturity Model*, Database Grand Conference.

Park, M-H. (2002), *A Decision Support System for Stepwise Improvement of Quality Competitiveness*, SungKyunKwan University.

Park, M-H. (2008), *Analysis of Efficiency and Productivity*, Korea Studies Information co.

Crawsey, R. A. (1976), *A Business Performance Measure of Quality Management*, Annual Quality Congress Transactions, American Society for Quality Control.

Chang, S-J., Park, Y-H. and Park, E-H. (1996), *Quality Costs in Multi-stage Manufacturing Systems*, Computers and Industrial Engineering, 31(1-2), 115-118.

Stephen, T. Knox. (1993), *Modeling the Cost of Software Quality*, Digital Equipment Corporation.

**이 상 철**

성균관대학교 산업공학 학사
성균관대학교 산업공학 석사
성균관대학교 산업공학 박사
현재 : 한국국방연구원 국방정보체계관리단
선임전문연구원

**신 완 선**

오클라호마대학교 산업공학 박사
삼성엔지니어링 사외이사
대한산업공학회 부회장
성균관대학교 품질혁신센터 센터장
현재 : 성균관대학교 시스템경영공학부 교수