

데이터 오너 관리를 통한 데이터 품질 향상

박지수

고려대학교 컴퓨터정보통신대학원

e-mail : beetle75@korea.ac.kr

Improving data quality through Data Owners management

Ji-Soo Park

Graduate School of Computer and Information Technology, Korea University

요 약

데이터 품질 기준은 반드시 현업의 입장에서 바라봐야 하며, 현업의 마인드가 데이터 품질에 가장 결정적인 영향을 미친다. 이에 따라 데이터 품질을 향상시키기 위해서는 현업이 데이터 품질 관리에 직접 참여할 수 있는 연구가 필요하다. 본 연구에서는 데이터 값(Data Value)에 대한 데이터 오너 (Owner)를 부여하여 데이터 품질 오류 시 현업이 직접 데이터 품질 관리 프로세스에 참여 할 수 있는 방안을 제시하였다. 데이터 품질 관리 프로세스는 데이터 품질 대상 및 기준을 정의하고 측정, 분석, 개선하는 방법이다. 본 연구에서 제시한 데이터 오너 관리 방안은 보다 효율적인 데이터 품질 관리 프로세스를 개선 시킬 수 있을 것이다.

1. 서론

최근 데이터 관리에 있어 가장 주목 받고 있는 단어는 단연 ‘품질 향상’이다. 부정확한 데이터는 중요한 비즈니스 결정에 막대한 악영향을 끼친다는 인식이 향상됐기 때문이다. 게다가 요즘 요구되고 있는 여러 컴플라이언스 준수 등은 질 낮은 데이터를 바탕으로는 해결할 수 없는 과제이다. 이제 문제는 데이터 통합보다는 데이터 품질이라는 소리다.[1]

질 낮은 데이터로 인한 손실은 실로 막대한 것이다. 전 세계적으로 고객 데이터 품질이 떨어져 발생하는 비용이 61 조에 달하고 있고, 국내 기업의 데이터 품질관리 수준은 데이터 품질관리 성숙수준 5 단계[5] 중 도입 단계인 1 단계를 조금 넘는 것으로 조사됐다. 그러나 금융과 통신 분야는 데이터 관리가 잘되는 편이지만 제조업과 공공부문은 데이터 품질관리 중요성에 대한 인식이 떨어진다. 기업이 데이터 품질에 대한 인식을 높일 수 있는 방법은 CEO에게 데이터 품질 수준을 정량적으로 제시하고 객관적인 품질 검사 결과를 보여주는 것이다.[2]

최근 정부에서 조사한 자료에 따르면 조사대상 117개 공공기관의 50%가 데이터 중복으로 인한 시스템 성능 저하 및 데이터 관리의 비효율을 경험하고 있는 것으로 나타났고, 메타 데이터 시스템을 구축한 기업은 불필요한 데이터를 60% 정도 줄일 수 있는 것으로 보아 데이터 관리는 반드시 필요하다. 또한 이러한 데이터 관리 중 데이터 품질 기준은 반드시 현업의 입장에서 바라봐야 하며, 현업의 마인드가 데이터의 품질에 가장 결정적인 영향을 미친다는 것을 잊어서는 안 된다.[3]

본 연구에서는 데이터 값(Data Value)에 대한 데이터 오너 (Owner)를 부여하여 데이터 품질 오류 시 현업

이 직접 데이터 품질 관리 프로세스에 참여 할 수 있는 방안을 제시하였다. 제시된 방안은 데이터 품질 관리 프로세스의 개선을 통해 데이터 품질 향상에 기여할 것이다.

본 논문의 구성은 2장에서 데이터 품질 확보에 대한 설문 결과를 분석하여 데이터 품질 관리 현황에 대해 기술하고, 3장에서 데이터 품질 향상을 위한 데이터 오너 관리 방안을 제시한다. 4장에서는 품질관리 시스템 구축 사례에 제시된 방안을 적용해, 제시된 방안의 효율성에 대해 입증하고, 마지막으로 5장에서 결론 및 향후 과제로 본 논문을 마무리한다.

2. 데이터 품질 관리 현황

2.1 설문 방법

A 컨설팅사의 홈페이지를 이용하여 “데이터 품질 확보 방안”에 대한 설문을 2007년 7월 2일부터 2007년 8월 31일까지 실시하였다.

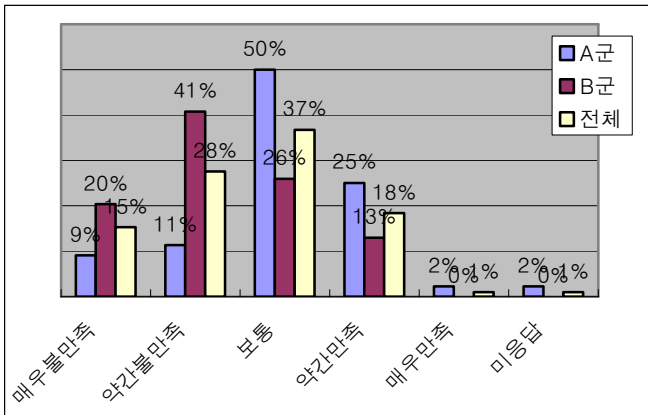
설문에 참여한 응답자 수는 총 98명이며, 응답자의 담당 업무를 기준으로 A, B 두 그룹으로 분류하여 설문 결과를 분석하였다. A군 응답자수는 44명, B군 응답자수는 54명이다.

- 1) A군 : IT 관리 및 운영 업무와 IT 개발 업무
- 2) B군 : IT 관리 및 운영, IT 개발 업무 이외의 업무

2.2 설문 결과

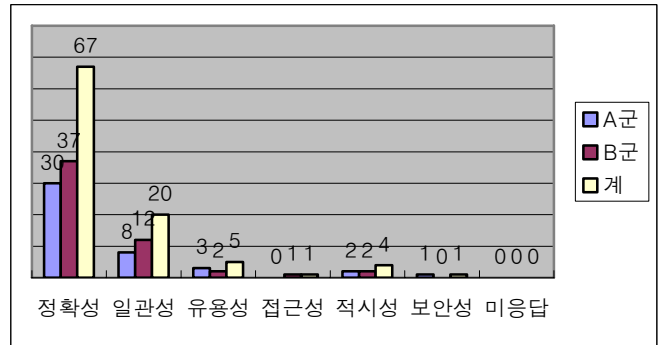
1) 데이터 품질 만족 현황

기업에 제공되는 데이터의 품질 만족도는 만족 20%(약간 만족 18%, 매우 만족 2%), 불만족 43%(약간 불만족 28%, 매우 불만족 15%)로 나타났으며, 만족은 2배 이상 A군이 좋게, 불만족은 B군이 3배 이상 크게 나타났다.



(그림 1) 데이터 품질 만족도

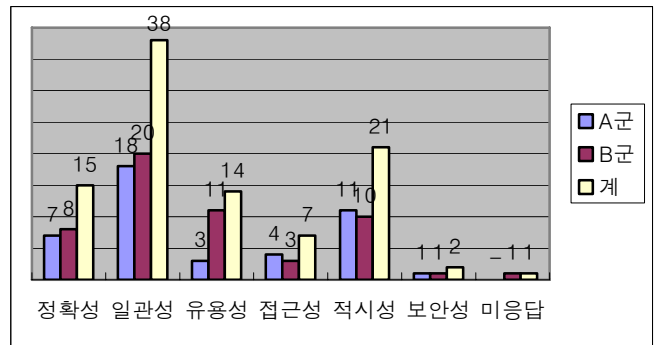
데이터 품질 기준 우선 순위는 A, B 군 모두 1 순위로 정확성을, 2 순위로 일관성으로 나타났다.



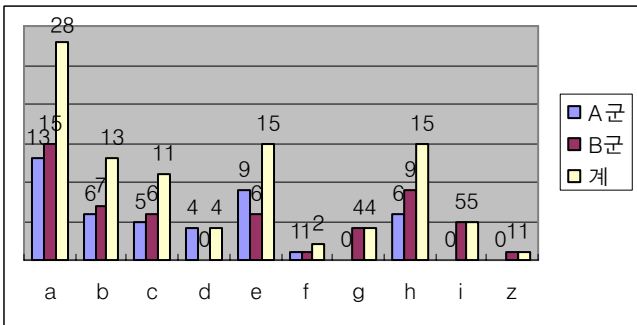
(그림 4) 데이터 품질 우선 순위 : 1 순위

2) 데이터 저 품질 원인

데이터 품질이 확보되지 않는 원인으로 A 군은 1 순위로 데이터 입력 오류 혹은 불성실한 입력, (그림 2) 2 순위로 시스템 복잡성 증가로 인한 데이터 관리 어려움으로 나타났고, (그림 3) B 군의 1 순위는 A 군과 같은 결과로, (그림 2) 2 순위는 데이터의 중복적 관리로 인한 데이터 불일치로 나타났다. (그림 3)



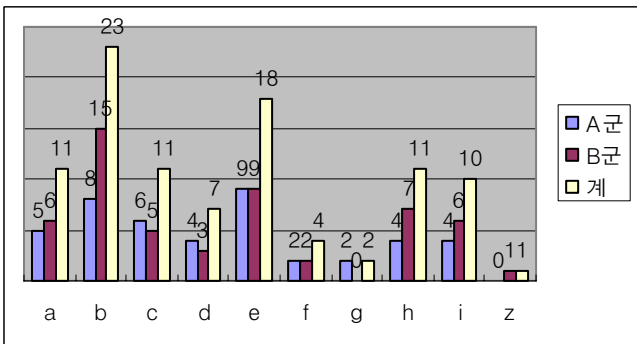
(그림 5) 데이터 품질 우선 순위 : 2 순위



(그림 2) 데이터 저 품질 원인 : 1 순위

4) 데이터 확보를 위한 프로젝트 수행 및 효과
기업에서 데이터 확보를 하기 위해 수행하는 프로젝트 1 순위는 Data Warehousing, 수행 예정인 프로젝트 1 순위는 Data Quality이며, Data Governance 는 예정 없음이 1 순위로 나타났다. (그림 6)

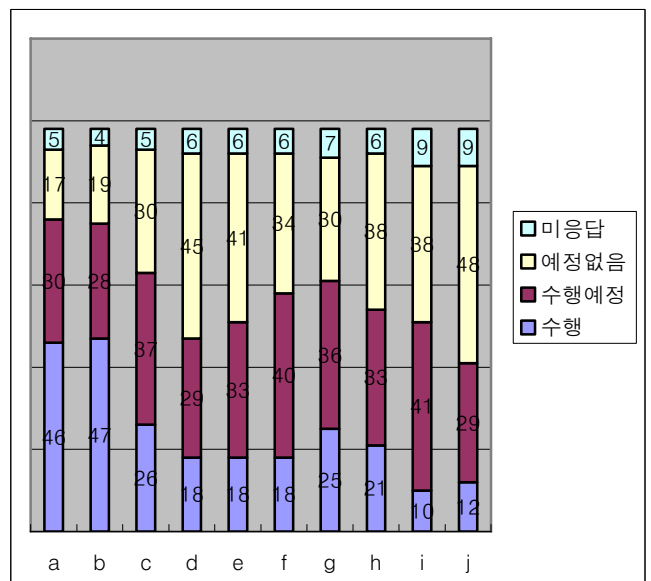
수행한 프로젝트 중 효과가 좋은 프로젝트는 Data Quality, 통합 DB 순으로 나타났고, 효과가 낮은 프로젝트로는 Data Warehousing, Business Intelligence, Meta Data 로 나타났다. (그림 7)



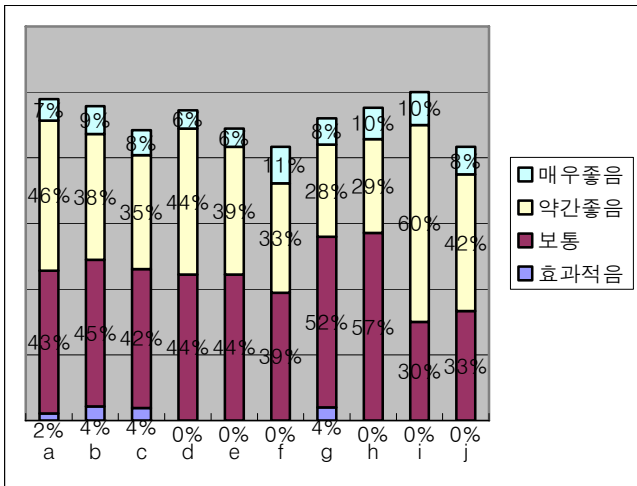
(그림 3) 데이터 저 품질 원인 : 2 순위

- a : 데이터 입력 오류 혹은 불성실한 입력
- b : 데이터의 중복적 관리로 인한 데이터 불일치
- c : 빈번한 정보요건 변경
- d : 데이터의 기하급수적인 증가
- e : 시스템 복잡성 증가로 인한 데이터 관리 어려움
- f : 전산에서의 잘못된 데이터 관리
- g : 여러 사람에 의한 데이터 조작
- h : 데이터에 대한 관리 책임자 부재
- i : 부문별 데이터 관리로 인한 혼선
- z : 미응답

3) 데이터 품질 기준 우선 순위



(그림 6) 데이터 확보 프로젝트 현황



(그림 7) 데이터 확보 수행 프로젝트 효과

- a : 통합 DB
- b : Data Warehousing
- c : Business Intelligence
- d : Customer Data Integration
- e : Enterprise Application Integration
- f : Data Architecture
- g : Metadata
- h : Master Data
- i : Data Quality
- j : Data Governance

5) 데이터 품질 관리 프로젝트 관련

전사적으로 데이터 품질 관리 프로젝트를 추진하는 기업은 30.6%이고, 프로젝트 추진 주체로는 IT 와 현업이 동시에 추진하는 경우가 63.3%, IT 조직이 추진하는 경우가 30.0%, 현업이 추진하는 경우가 6.7%로 나타났다. 데이터 품질 관리 프로젝트를 전사적으로 추진하지 않는 이유로는 담당조직의 부재, 관련자 참여 미흡, 필요성을 못 느낀다는 순으로 나타났으며, 현업에서 데이터 관리 역할을 수행하고 있는 기업은 48.0%로 나타났다. 데이터 품질 관리 프로젝트를 수행하면서 가장 어려운 점으로는 프로젝트의 넓은 범위, 오류 데이터에 대한 근본적 원인 해결이 어렵다는 순으로 나타났다.

3. 데이터 오너십 관리 방안

2 장에서의 데이터 품질 관리 조사 결과를 종합해보면, 대부분의 기업이 데이터 품질에 만족하지 못하고 있으며, 데이터를 확보하기 위해 많은 프로젝트를 수행했지만 그 효과는 미흡한 것으로 나타나고 있다. 이들 프로젝트들이 효과를 보지 못하는 원인은 유형에 따라, 또는 엄밀한 판단 없이 수행하는 경우도 있고, 총체적인 접근보다는 부분적인 문제 해결에 그친 경우도 많기 때문이다.[3]

반면, 프로젝트의 효과가 가장 좋은 Data Quality 프로젝트는 프로젝트 범위와 저 품질 데이터의 근본적인 원인 해결이 어렵다는 이유 및 데이터 품질 기준이 모호하다는 이유로 4.1%만 수행하고 있음을 알 수 있다.

본 연구는 데이터 품질관리 대상이 되는 구성 요소

중 데이터 값의 오너를 정의하고, 데이터 품질 관리 프로세스[5] 중 데이터 오너십 관리 활동을 데이터 품질관리 시스템 구축 사례에 적용하였다. 여기서 데이터 오너십 관리란 데이터에 대한 책임과 권한을 가진 오너를 중심으로 데이터 품질 관리 제반 절차나 과정을 정의하고 실행하는 모든 활동을 의미한다.[5] 제시된 방안은 효율적이고 효과적인 데이터 품질 관리 프로세스를 개선시킴으로써 데이터 품질 향상의 기반을 마련한다. 또한 데이터 품질 관리 시 큰 이슈가 발생할 경우 이를 해결하기 위한 데이터 관리 위원회 또는 데이터 중재 관리 위원회 같은 조직 구성안을 제시한다.

본 연구에서는 <표 1>과 같은 데이터 값의 세부 관리 대상의 오너(owner)를 정의한다.

<표 1> 데이터 값의 세부 관리 대상

조직	데이터 값	세부 관리 대상
DA (개념적 관점)	표준 데이터	표준 단어(Word) 사전, 표준 도메인(Domain) 사전, 표준 용어(term) 사전, 표준 코드
Modeler (논리적 관점)	모델 데이터	데이터 참조 모델, 개념 데이터 모델, 논리 데이터 모델, 물리 데이터 모델에 대한 메타 데이터 및 DBMS 객체 정보
DBA (운영적 관점)	관리 데이터	흐름 관리 데이터
User (운용적 관점)	업무 데이터	운영 데이터, 분석 데이터

데이터 값 중 관리 데이터, 업무 데이터의 세부 관리 대상의 관리 방법[4]에서 도출할 수 있는 관리 요소는 생성 및 매핑 규칙이다. 생성 및 매핑 규칙의 경우 모델 데이터 오너 정의에서 도출한 관리 요소 중 자료 생성 규칙, 자료 정제 규칙, 데이터 이관 규칙에서 관리할 수 있으므로 별도로 제시하지는 않는다.

3.1 표준 데이터 오너 정의

표준 데이터의 세부 관리 대상에 대한 오너를 정의하여 관리 방법[4]에서 도출한 관리 요소에 변경이 발생하거나, 데이터 표준 검토 시 이슈가 발생할 경우 DA 또는 IT 담당자가 오너와 협의하여 해결(승인 처리 등) 할 수 있도록 한다. 세부 관리 대상이 변경될 경우 생성일자와 폐기일자를 관리하여 데이터 품질 측정 시 폐기된 표준 데이터를 따르고 있는 과거 데이터의 경우 해당 오너와 협의하여 개선안을 수립한다.

<표 2> 표준 데이터 샘플

표준 데이터 세부 관리 대상	설명	등록자	오너	생성일자	폐기일자	관리 방법에서 도출한 관리 요소

3.2 모델 데이터 오너 정의

모델 데이터의 세부 관리 대상에 대한 오너를 정의하고, 세부 관리 대상 중 컬럼의 경우 관리해야 하는 관리 요소를 <표 3>과 같이 도출하였다. 관리 요소의 도출 기준은 컬럼 자체의 메타에 관한 변경 사항과 실 데이터의 처리 시 필요한 요소를 도출한 것이다.

<표 3> 모델 데이터 샘플

컬럼	컬럼	소속	생성	오너	핵심	컬럼	컬럼	실	데이	자료	자료	데이
정의	테이블	자	자	항목	생성	변경	데이터	데이터	생성	정제	데이터	데이터
				여부	일자	일자	발생	관	규칙	규칙	관	관
							일자	자			칙	칙

도출한 관리 요소들은 4 장에서 품질 관리 구축 사례에 적용하여 기술하기로 한다.

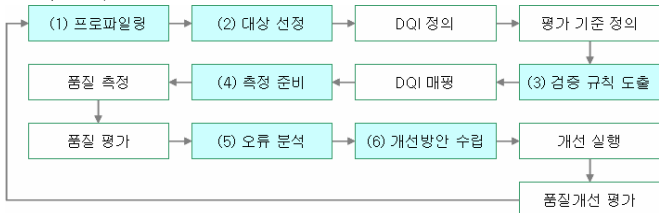
3.3 데이터 품질 관리(중재) 위원회

데이터 오너를 정의하는 가장 큰 이유는 데이터를 관리함에 있어 이슈 발생 시 보다 정확하고 신속하게 해결하기 위해서이다. PMBOK 에서 프로젝트에 큰 이슈가 발생하면 이해당사자들이 모여 해결책을 강구하는 변경관리위원회(Change Control Board-CCB)처럼, 데이터 관리에 큰 이슈가 생겼을 경우, 데이터 이슈에 대한 통제력/교섭력을 가진 CIO 또는 데이터 품질 관리 최종의사결정자를 리더로 해당 이슈에 관련된 데이터 오너들이 모여 이슈 해결을 하는 방안을 제시한다.

4. 데이터 품질관리 시스템 구축 사례

4.1 데이터 품질 관리 프로세스 구현 사례

(그림 8)은 I 은행에서 데이터 품질 관리 시스템 구축 시 정의한 품질 관리 프로세스이다. 3 장에서 도출한 컬럼 데이터의 오너와 관리 요소들을 해당 프로세스에 적용해 어떻게 프로세스를 개선 할 수 있는지 살펴본다.



(그림 8) I은행의 품질 관리 프로세스

(1) 프로파일링

프로파일링 대상 선정 시 모델 데이터에서 데이터 오너가 정의해 놓은 핵심 품질 항목 여부를 기준으로 대상을 선정할 수 있어, 대상 선정 시 소요 시간을 단축 시키며, 품질 관리 핵심 이슈의 대상을 선정할 수 있다. 또한 실 데이터에 대한 접근이 불가능한 경우에 모델 데이터에서 관리하는 데이터를 이용하여, 프로파일링 대상으로 할 수도 있다.

(2) 대상 선정

프로파일링의 결과 도출된 데이터 품질 핵심 이슈에 따라 핵심 품질 항목을 선정, 대상을 정의 한 후 우선 순위 선정 시 오너와 협의하여 순위를 선정하면 된다.

(3) 검증 규칙 도출 및 (4) 측정 준비

검증 규칙 도출은 가장 시간이 많이 소요되며 현업의 적극적인 참여 없이는 이루어지기 힘들다. 그러나 데이터 오너들이 사전에 정의한 자료 생성규칙과 데이터 이관 규칙을 사용할 수 있다면 보다 빠른 시간 내에 정확하고 필요한 업무 규칙을 도출할 수 있을 것이다. 또한 검증 규칙 도출 시 데이터 오너가 정의

해 놓은 규칙 이외의 추가 또는 변경 규칙이 발생한다면 데이터 오너는 자료 생성 규칙 및 데이터 이관 규칙을 변경관리 하여 최신성 및 정확성을 유지할 수 있다.

(5) 오류 분석 및 (6) 개선 방안 수립

검증 규칙 도출만큼 많은 시간과 노력이 필요한 작업이다. 데이터 오너가 중심이 되어 오류 데이터 특성 파악 및 오류 요인을 분석하여 개선 방안을 수립한다며 보다 근본적인 원인을 찾아낼 수 있으며, 개선 방안 수립 시, 보다 빠른 의사결정을 내릴 수 있을 것이다. 또한 개선 방안은 데이터 모델의 자료 정제 규칙에서 관리함으로써 향후 유사한 오류가 발생 시 보다 빨리 대처할 수 있을 것이다.

기존 구축된 품질 관리 프로세스 중에는 현업이 참여해야 하는 작업이 많다. 이때, 데이터에 대한 오너십이 문제가 되어 현업 선정 시 많은 어려움을 겪게 되고, 개선 방안 도출 시 빠른 의사결정을 내리기 힘들다. 데이터 값의 오너를 정의하여 관리함으로써 기존 프로세스 보다 빠르고, 보다 정확한 품질 관리를 할 수 있을 것이다.

5. 결론 및 향후 과제

데이터 품질의 확보에 있어서 현업의 역할은 필수적이다. 현업이 참여하지 않는 데이터 관련 프로젝트는 성공을 보장할 수 없으며, 데이터 관리 최고 책임자도 현업이어야 한다.[3]

본 논문에서 제시된 데이터 오너 관리 방안은 현업이 데이터 관리 주체가 되게 함으로써 데이터 품질을 향상 시킬 수 있는 방안을 제시하였다. 제시된 방안은 각각의 데이터 값의 세부 관리 대상[4]에 오너를 정의하고, 관리하는 요소를 도출하여 관리 함으로써, 데이터 품질 기준 정의, 품질 측정, 오류 분석, 개선 방안을 도출하는 데이터 품질 관리 실무에 기여할 수 있다.

향후에는 제시된 오너 관리 방안을 적용한 모델을 메타데이터로 관리하여 데이터 품질관리 시스템 구현 시 조직의 역할과 책임에 대한 명확화와 체계화를 확보하여야 한다. 또한 저 품질 원인의 가장 큰 비율을 차지하고 있는 데이터 입력 오류 혹은 불성실한 입력을 방지할 수 있는 연구가 필요하다.

참고문헌

- [1] 송지혜, “기업 경쟁력 넘어 생존력 ‘좌우’”, Network Times, 2006.3
- [2] 비즈테크, “데이터 품질 관리 전담 기관 필요하다”, 2007.9.20
- [3] 이진우, “기업 가치, 데이터 품질이 좌우한다”, 컴퓨터월드, 2007.9
- [4] 한국데이터베이스진흥센터, “데이터 품질관리지침 (Ver 2.1)”, 2006.9
- [5] 한국데이터베이스진흥센터, “데이터 품질관리 성숙 모형 (Ver 1.0)”, 2006.9
- [6] 2e Consulting, “데이터 품질관리 시스템 목표시스템 정의서”, 2007.4