

데이터 본질 기반의 데이터 분류 방법론

(A Data Taxonomy Methodology based on Their Origin)

최 미 영 [†] 문 창 주 ^{**} 백 두 권 ^{***} 권 주 흠 ^{****} 이 영 무 ^{*****}
 (Mi-Young Choi) (Chang-Joo Moon) (Doo-Kwon Baik) (Ju-Hum Kwon) (Young-Moo Lee)

요 약 조직의 데이터를 효과적으로 관리하는 대표적 방법은 기존 데이터의 공유와 재사용을 촉진하여 데이터의 중복 생산을 방지하는 것이다. 데이터 공유와 재사용의 촉진을 위해서 기존 데이터의 체계적 구조화와 효율적인 검색이 지원되어야 한다. 이러한 점이 고려되지 않은 조직간 단절된 데이터 개발은 데이터 중복을 양산하고 데이터의 품질을 저하시킨다. 데이터 분류는 관리하는 데이터에 대한 체계적 정리를 원하는 데이터 요소의 빠른 검색을 가능하게 한다.

본 논문에서는 데이터 공유, 재사용과 통합을 극대화하고 MDR과 시멘틱 웹에서 효과적으로 사용될 수 있는 본질기반 데이터 분류 방법론을 제안한다. 본질기반 데이터 분류 방법론은 데이터 본질을 기반으로 데이터 분류 구조를 구성하여 업무분류에 독립적인 데이터 분류가 가능하다. 또한 제시된 데이터 분류 구조를 지원하는 데이터 분류 절차를 제시하여 다양한 데이터 요소들을 데이터 분류 구조에 따라 배치하는 방법을 보인다. 사례연구에서는 제안된 데이터 분류 구조와 데이터 분류절차가 효과적으로 실제에 적용될 수 있음을 보였다.

키워드 : 데이터 분류, 데이터 아키텍처, 비즈니스 분류, 데이터 공유, 데이터 통합

Abstract The representative method to efficiently manage the organization's data is to avoid data duplication through the promotion of sharing and reusing existing data. The systematic structuring of existing data and efficient searching should be supported in order to promote the sharing and reusing of data. Without regard for these points, the data for the system development would be duplicated, which would deteriorate the quality of the data. Data taxonomy provides some methods that can enable the needed data elements to be searched quickly with a systematic order of managing data.

This paper proposes that the Origin data taxonomy method can best maximize data sharing, reusing, and consolidation, and it can be used for Meta Data Registry (MDR) and Semantic Web efficiently. The Origin data taxonomy method constructs the data taxonomy structure built upon the intrinsic nature of data, so it can classify the data with independence from business classification. Also, it shows a deployment method for data elements used in various areas according to the Origin data taxonomy structure with a data taxonomic procedure that supports the proposed taxonomy. Based on this case study, the proposed data taxonomy and taxonomic procedure can be applied to real world data efficiently.

Key words : Data Taxonomy, Data Architecture, Business Classification, Data Sharing, Data Consolidation/Integration

· 이 논문은 2009년도 교육과학기술부의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(K20601000001)

논문접수 : 2009년 10월 14일
 심사완료 : 2009년 11월 17일

[†] 학생회원 : 고려대학교 컴퓨터학과
 michelle@paran.com
^{**} 정 회 원 : 건국대학교 항공우주정보시스템공학 교수
 cjmoon@konkuk.ac.kr
 (Corresponding author임)
^{***} 종신회원 : 고려대학교 컴퓨터학과 교수
 baikdk@korea.ac.kr
^{****} 정 회 원 : 공군 군수사령부 전산소 개발실장
 jkweon@gmail.com
^{*****} 정 회 원 : 지식경제부 신산업정책관실 소프트웨어진흥과 사무관
 lym0202@mke.go.kr

Copyright©2010 한국정보과학회: 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.
 정보과학회논문지: 컴퓨팅의 실제 및 래터 제16권 제2호(2010.2)

1. 서론

데이터 관리의 목적은 효율적으로 고품질의 대용량 데이터를 유지하는 것이다. 특히 조직이 복잡해지고 조직 간 상호 데이터 요구량이 높아지면 고품질의 데이터를 유지하고 관리하는 것은 어려운 과제이다. 데이터 중복 최소화화 및 중복된 데이터의 통합은 이러한 과제 해결을 위하여 반드시 필요하다. 데이터 중복 최소화화를 위한 효과적인 방안은 데이터 아키텍처 내의 중요 데이터를 효과적으로 검색하여 데이터의 공유와 재사용을 극대화하는 것이다.

메타데이터 레지스트리(Meta Data Registry : MDR)나 온톨로지 기반의 시맨틱 웹은 데이터의 공유와 재사용을 촉진하는 대표적인 방법들이다. 그러나 이 방법들은 구현하기 어렵고 많은 시간과 비용이 요구된다. 비록 데이터 공유와 재사용이 메타 데이터 레지스트리와 시맨틱 웹을 구축하는 목적 중의 하나이나 단지 데이터 공유와 재사용만을 위하여 이들을 구축하는 것은 비효율적이다. 데이터 분류는 MDR과 시맨틱 웹에 비해서 적은 비용과 시간으로 데이터 공유와 재사용을 극대화한다. 예를 들어 대형 마트에 익숙한 고객은 원하는 상품을 마트의 상품 분류 체계 내에서 빨리 찾을 수 있다. 이와 같은 접근방법은 데이터 공유를 위한 데이터 검색에도 동일하게 적용될 수 있다. 또한 데이터 분류는 유사 데이터를 분류하는 과정에서 중복된 데이터를 식별할 수 있어 중복 데이터 통합에 사용된다.

데이터 분류는 MDR과 시맨틱 웹의 구축을 지원한다. ISO 11179 MDR Part 2 분류(Classification)에서는 특정 분류 체계의 사용을 지정하지는 않지만 어느 형태의 분류 체계이건 데이터 관리를 위해 분류 체계가 필수적으로 요구된다는 것을 강조하고 있다. 시맨틱 웹과 데이터 분류는 상호보완적인 관계로 데이터 분류는 시맨틱 웹의 검색의 범위를 줄일 수 있다. 따라서 비용 효과적 접근을 위해서는 시맨틱 웹이나 MDR 구축을 수행하기 전 단계로 보유한 데이터의 체계적 분류가 선행되어야 한다.

본 논문에서는 데이터 공유, 재사용과 통합을 극대화하고 MDR과 시맨틱 웹에서 효과적으로 사용될 수 있는 데이터 본질에 기반한 데이터 분류 방법론인 본질기반 데이터 분류 방법론을 제안한다. 본질기반 데이터 분류 방법론은 데이터 본질 중심의 데이터 분류 방법론이다. 본질적 데이터는 업무 요구에 따라 생성되는 것이 아니라 해당 업무 영역의 고유 특성을 반영하는 중요 데이터이며 다른 데이터들을 파생할 수 있는 근간 데이터들이다. 따라서 본질기반 데이터 분류 방법론은 데이터 본질을 기반으로 데이터 분류체계를 구성하여 업무

분류에 독립적인 데이터 분류가 가능하다. 업무분류에 의존한 데이터 분류는 다양한 업무적 요구에 따라 데이터 요소의 중복 배치가 가능하므로 결과적으로 데이터의 중복 생산을 야기한다.

데이터 분류에 관한 필요성이 높아지면서 정부와 민간 기업에서 자체적으로 데이터 분류를 시도하고 있다. 대한민국은 범정부 DRM을 통하여 데이터 분류를 시도하고 있는 중이다. 그러나 아직 체계적 분류 구조와 분류를 위한 절차가 논문화 되어 발표된 것은 없다. 본 논문에서 제안하는 본질 기반 데이터 분류 방법론은 데이터를 생성하는 주체와 데이터를 생성하기 위해 사용하는 자원, 그리고 주체가 자원을 이용할 때 발생하는 행위 데이터를 기반으로 분류를 시작한다. 이는 이전의 논문에서 발표되지 않았던 새로운 시각의 접근 방법이다. 분류 구조와 분류 구조의 구성 요소는 대한민국 범정부 DRML0과 그 개정작업에서 그 일부를 사용했으나 본 논문에서 보다 구조화하고 체계적으로 개선하였다. 또한 제안하는 분류 구조에 맞게 데이터를 분류하기 위한 절차적 방법론 역시 처음으로 제시하고 있다. 따라서 그동안 조직의 요구에 맞추어 자의적으로 이루어진 데이터 분류업무에 대한 구체적인 실질적인 가이드 라인을 제공한다.

본질기반 데이터 분류 방법론의 데이터 분류 구조는 업무 주제 영역에 의존하지 않고 고정적으로 존재하는 2레벨의 데이터 영역(Data Area)과 업무 주제 영역에 따라 구성이 가변적인 데이터 클래스(Data Class)로 나누어진다. 본질기반 데이터 분류 방법론에서 분류하는 데이터 대상은 데이터 모델 상의 논리 엔티티 단위이다. 이는 대부분의 조직에서 데이터를 관리하는 단위가 DBMS(DataBase Management System)의 테이블이고 테이블은 데이터 모델의 논리 엔티티의 물리적 표현이기 때문이다. 데이터 분류 절차는 데이터 클래스로 정의된 데이터 분류구조에 업무적 개념이 반영된 데이터 요소를 할당하는 것이다. 본 논문의 본질기반 데이터 분류 방법론은 커널 엔티티, 코드 엔티티, 교차 엔티티, 종속 엔티티로 구분된 데이터 요소들을 8단계의 데이터 분류 절차를 거치면서 데이터를 분류한다. 사례연구에서는 쇼핑몰 데이터와 정부 부처의 데이터를 이용하여 본질기반 데이터 분류 방법론이 데이터 분류, 데이터 공유, 데이터 통합, MDR과의 융합 등에서 효과적으로 사용될 수 있음을 보였다.

본질기반 데이터 분류 방법론을 사용하면 신속한 데이터 검색이 가능하며 기존 데이터의 중복 여부를 빠르게 판별할 수 있다. 신규 데이터 생성 요구가 있다면 데이터 관리자는 데이터 분류를 통하여 중복 데이터 생산인지 반드시 필요한 신규 데이터인지를 판별할 수 있다.

또한 본질기반 데이터 분류 방법론을 이용하면 중복된 데이터들은 동일한 분류영역에 속하게 되어 데이터 관리자는 통합 데이터 후보를 정확하게 식별하고 통합 할 수 있다. 이러한 과정에서 데이터 공유와 재사용을 촉진하고 중복 데이터의 식별이 가능하여 고품질의 데이터를 효율적으로 관리하는 것이 가능하다.

본 논문의 구성은 다음과 같다. 2장에서는 데이터 분류를 위한 정부 차원의 노력과 관련 연구 진행상황을 보인다. 3장은 본질기반 데이터 분류 방법론으로 본 논문에서 제시하는 데이터 분류 구조와 데이터 분류 절차에 대해서 설명 한다. 4장에서는 쇼핑물과 정부부처의 데이터가 본질기반 데이터 분류 방법론에 따라 분류/검색/통합되고 MDR과 융합되는 사례 연구를 보인다. 5장에서는 결론과 향후 연구 방향을 언급한다.

2. 관련 연구

데이터 공유와 재사용의 필요성이 부각되면서 국가차원에서 정부 부처의 정보자원 최적화를 위하여 데이터 분류와 활용을 시도하고 있다. 국내의 경우 2001년에 3월에 제정되어 최근인 2009년 5월까지 개정이 진행되며 시행중인 전자정부법 21조 행정정보공동 이용에 의하여 의무적으로 정부 부처간 정보공유[1]를 해야 한다. 이를 위하여 행정안전부는 정보 자원 최적화라는 목표 아래 범정부 EA(Government wide Enterprise Architecture)구축을 진행중이며 지속적으로 정부부처의 데이터를 수집하여 부처간 데이터공유와 통합을 위한 데이터 분류 구조 정립을 시도하고 있다.

미국 정부는 데이터 관리를 위한 FEA(Federal Enterprise Architecture)의 DRM(Data Reference Model) 2.0에서 데이터 분류의 필요성을 명시하고 있다[2]. 이는 데이터 공유와 교환 구조에 중점을 두던 DRM 1.0[3]보다는 보다 구조적으로 진보된 개념이다. DRM 2.0의 Data Taxonomy는 다른 RM(Reference Model : BRM, SRM, TRM, PRM)들 과의 연결을 위한 Ontology기반의 Data Taxonomy를 요구하고 있다. 미국 FAA(Federal Aviation Administration)에서도 Data Taxonomy의 필요성과 기본 요건을 언급하고 있다. FAA의 Rick Jordan은 메타데이터 표준화 포럼에서 FAA의 데이터를 주제(Party)와 사건(Event)이라는 두 분류로 나누는 것과 비슷한 주제 영역들을 보다 상위 수준에서 모을 것을 제안하였다[4]. 그러나 이에 대한 구체적인 내용은 아직 발표되지 않았고 Data Taxonomy 활용방안에 대한 연구도 필요하다.

데이터 분류에 관한 대부분의 이전 연구들은 데이터 분류에 대한 필요성과 효용성을 언급하지만 이에 대한 체계적인 데이터 분류 구조 방법론이나 데이터 분류를

위한 구체적인 절차를 보이는 논문들은 아직 발표되고 있지 않다. Gutierrez, Serrano 등의 논문은 정보 기술과 업무 목적과의 연계를 기반으로 데이터 분류에 대한 여러 가지 방안을 제시하였으나 분류에 대한 구체적 절차나 방법론은 언급하지 않았다[5]. Malcolm Chisholm의 논문은 데이터 분류는 정보를 보다 빠르고 정확하게 검색하는데 유용하다고 주장하나 데이터 분류의 목적을 데이터 검색에만 한정하였다[6]. Tim Bryce는 그의 블로그에 ‘데이터 분류의 혜택’이라는 글을 올렸고 여기에서 그는 데이터 분류는 여러 개의 시스템에서 표준화, 공유, 재사용을 위해 사용할 수 있다고 주장하였다[7]. 그는 데이터가 단지 응용 개발자의 관점에서 정의될 때 중복이 발생하고, 데이터는 데이터 소유의 성격에 의해 분류되어야 한다고 주장한다. 그러나 분류의 필요성과 혜택을 언급할 뿐 분류를 위한 방법론이나 구체적 절차를 소개하는 것은 아니다. Bradley, Curry 등의 논문은 정성적 데이터 분석을 위한 분류(taxonomy), 주제(theme), 이론(theory)이라는 개괄적 구조만을 소개하고 있다[8]. 이 논문은 Data 수집과 분석은 경험 있는 한 사람에 의해 수행되며 다른 이에 의해 검증, 보완하는 것을 제안하고 있다.

3. 본질기반 데이터 분류 방법론

본질기반 데이터 분류 방법론은 데이터를 업무 목적에 의존하지 않고 본래의 특성에 따라 분류한다. 즉, 데이터를 업무적 관점이 아닌 본질의 태생적 관점으로 규정하는 것이다. 데이터는 업무적 필요성에 의해 생성되기 때문에 업무 기반의 데이터 분류는 데이터 검색 시 적합한 면에서 유리하다. 그러나 업무와 데이터는 근본적으로 다대다(M:M, Many to Many)의 관계를 갖기 때문에 이러한 분류는 필연적으로 데이터의 중복을 발생시킨다. 즉, 하나의 업무는 여러 데이터를 사용하고 한 데이터는 여러 업무에서 사용된다. 또 다른 이유는 업무는 시간이 지남에 따라 자주 분할되거나 통합 되는 불안정한 대상이다. 업무가 분할되어 세분화 되는 경우 기존 데이터는 약간의 특성만 추가된 별도 데이터로 생성되어 데이터의 중복이 발생한다. 업무가 통합될 때 중복되는 되는 데이터들이 종종 그대로 존재하게 되고 중복된 데이터 간의 불일치는 전반적인 데이터의 품질을 저하 시킨다. 그러나 업무적 관점이 아닌 데이터 본질에 기반한 데이터 정의는 데이터의 존재 목적에 초점을 둔다. 예를 들어 통신 회사에서 가입자, 실사용자, 요금납부자 등으로 구분하여 데이터를 생성하는 것은 개별 업무 관점의 데이터 관리이다. 즉, 하나의 고객인 홍길동이 역할에 따라 여러 모습으로 생성된 것이다. 이러한 데이터 관리보다는 홍길동이 하나의 고객 개체로 먼저

존재하고 고객의 역할에 따라 각각 다른 관계 데이터를 생성하는 것이 본질 기반의 데이터 정의이다.

3.1 데이터 분류 구조

대표적인 분류구조들[9,10]의 공통적 특성은 대상 집합을 같은 목적으로 나누는 것으로 대개 한 분류 그룹 내에서 계층 구조를 갖는다. 상위 분류에서는 분류간 구분되는 성격이 뚜렷해야 한다. 반대로 한 분류 영역에서 분류가 세분화되면 세분화된 분류 영역 내 구성원들 간의 유사성은 높아진다. 한 분류체계 내에서 하위계층의 속성은 상위계층의 속성을 상속받고 거기에 몇 가지가 추가된 것이다. 분류구조들의 공통적인 특성을 기반으로 본질기반 데이터 분류 방법론의 데이터 분류 구조는 아래의 항목들을 준수한다.

- 1) 분류 영역의 계층, 정의 방법 등이 일관된 원칙과 기준을 가져야 한다.
- 2) 한 분류 영역 내에서 상위 분류의 속성이 하위 분류로 계승되어야 한다.
- 3) 분류구조의 계층성 유지를 위하여 한 계층에는 2개 이상의 분류영역 수가 존재하는 것을 권장한다.

본질기반 데이터 분류 방법론의 데이터 분류 구조는 그림 1과 같다. 데이터의 본질 특성이 데이터 분류에 반영 되어 데이터 요소 할당에 있어 중복이 발생하지 않도록 데이터 영역(Data Area)은 상위 계층(레벨1)에 주체(Party), 자원(Resource), 활동(Action) 세가지 영역이 있고 하위계층(레벨2)으로 주체는 개인(Individual)과 조직(Organization)으로, 자원은 유형(Visible)과 무형(Invisible)으로, 활동은 준비(Plan)와 진행(Progress)로 나누어진다. 데이터 클래스(Data Class)는 데이터 영역의 하위로 업무 주제 영역에 따라 구성 항목이나 계층의 깊이가 달라진다. 예를 들어 통신 업무 영역에서의

시설 분류는 통신 전자 장비를 분류하는 것이나 공공업무에서는 건물이나 정류장, 항구 등을 분류하는 것이다. 데이터 요소(Data Elements)는 분류의 대상이 되는 데이터 단위로 공유, 재사용, 통합의 대상이다. 본질기반 데이터 분류 방법론에 따른 데이터 분류는 6개의 데이터 영역(개인, 조직, 유형자원, 무형자원, 준비, 진행)인 정적 구조에 업무 영역의 특성에 따라 도출된 데이터 클래스를 가변적인 계층구조로 구성하고 단말 데이터 클래스에 데이터 요소를 배치하는 것이다.

3.1.1 데이터 영역

데이터 영역은 그림 1의 좌측에 사각형으로 표현되었다. 데이터는 일반적으로 누군가(주체) 무엇으로 혹은 무엇을 대상(자원)로 행위(활동)하면서 발생한다. 주문이란 고객이라는 주체가 상품이라는 대상에 대해 구매라는 활동을 하는 것이다. 이러한 관점에서 데이터 분류 구조의 가장 상위 부분을 주체(Party), 자원(Resource), 활동(Action)이라는 세가지 영역으로 구분하고 레벨1 데이터 영역이라 한다. 이러한 시도는 Linné's가 움직이는 것과 그렇지 않은 것으로 가장 상위의 분류구조를 동물계와 식물계로 분류한 것과 유사하다[10]. 데이터 영역 역시 아래의 특성에 따라 각 영역간의 구분을 명확히 하였다.

- 1) **주체(Party)** : 활동을 일으키는 당사자로 데이터를 생성하는 개인 혹은 단체
- 2) **자원(Resource)** : 주체가 활동을 위하여 사용하는 대상 혹은 활동의 대상
- 3) **활동(Action)** : 분명한 목적이나 동기를 가지고 사고, 선택, 결심을 거쳐 의식적으로 행하는 인간의 의지적인 언행과 그 결과. 활동의 범위는 하나의 업무 단위로 한정한다. 즉, ACID (Atomicity,

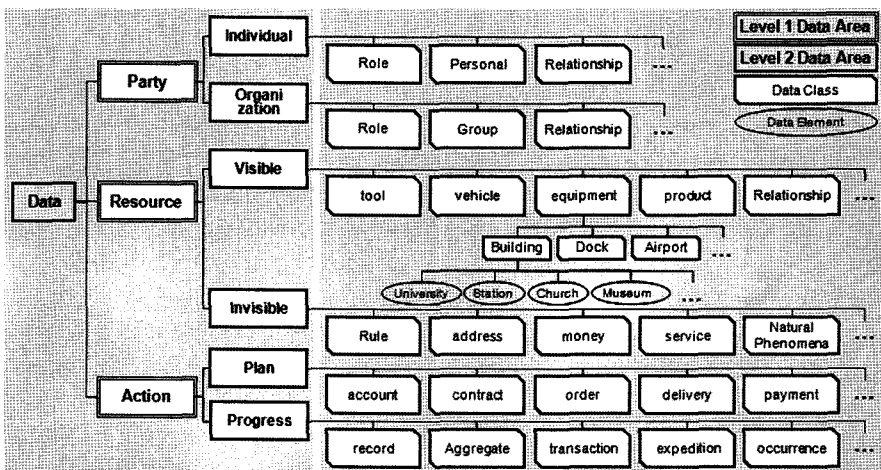


그림 1 데이터 분류 구조

Consistency, Isolation, Durability) 특성을 보장하는 최소의 업무 단위이어야 한다.

그림 2는 주체, 자원, 활동의 관계를 보인다. 활동이 주체와 자원에 의해 발생하므로 주체와 자원은 활동보다 선행해서 발생하는 데이터이다. 주체와 자원은 해당 데이터가 스스로 존재하는 즉, 데이터 생성의 원인이 되는 부모 데이터를 가지지 않는 데이터들의 집합이다. 따라서 주체와 자원은 데이터의 본질 특성이 강하다. 상대적으로 이들에 의해서 파생되는 활동 데이터는 업무의 고유특성을 반영하는 성향이 강하다. 데이터를 보는 관점에 따라서 주체와 주체가 활동을 만들거나 자원과 자원이 활동을 만들어 데이터를 생성할 수 있다. 예를 들어 개인 간의 계약이나 개인과 조직과의 계약은 주체와 주체 간에서 생성된다.

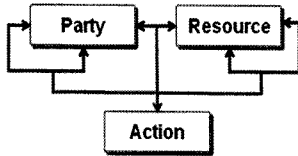


그림 2 데이터 영역 관계도

실세계의 객체(Object)들은 직관적으로 유형의 객체와 무형의 객체로 나누어진다. 이러한 개념은 주체, 자원, 활동에 모두 적용될 수 있다. 유형으로 존재하는 주체(Party)는 개체(Individual)이고 하나 이상의 개체들이 모여 생성된 무형의 주체는 조직(Organization)이다. 즉, 가족, 회사, 그룹 등은 무형이지만 데이터를 생성하는 주체이다. 그러므로 주체라는 데이터 영역은 개체와 조직이라는 하위 데이터 영역을 갖는다. 자원 데이터 영역은 실세계에서 유형자원과 무형자원으로 나누어진다. 활동 데이터 영역은 행위 결과의 데이터 집합인 유형의 진행(Progress)과 행위 수행 전 무형의 계획(Plan)으로 나누어진다. 진행과 계획으로의 분류는 어느 업무에서도 공통으로 나타나므로 업무분류는 독립적이다. 따라서 주체, 자원, 활동에 공통으로 유형 무형의 양분화된 개념을 적용하여 아래와 같은 하위 데이터 영역을 정의하고 이를 레벨2 데이터 영역이라 정의한다.

1) 주체(Party)

개체(Individual) : 활동 즉, 행위를 일으키는 구체적 개별 개인. (Ex. 고객, 청구인, 학생 ...)

조직(Organization) : 하나 이상의 개체가 목적을 갖고 모인 집단. (Ex. 대학교, 보험사, 통신사 ...)

2) 자원(Resource)

유형자원(Visible Resource) : 눈에 보이고 실체가 있는 자원 (Ex. 건물, 시설물, 상품, 도구 ...)

무형자원(Invisible Resource) : 눈에 보이지 않고 실체가 없는 자원 (Ex. 에너지, 규제, 표준, 권리, 자금, 가상공간, 주소, 지식 ...)

3) 활동(Action)

계획(Plan) : 주체가 자원을 분배 받아 활동을 준비한다. (Ex. 예산, 주문, 계약, 계좌 ...)

진행(Progress) : 시간의 흐름에 따라 계획이 수행되는 내역. (Ex. 계좌거래, 성적, 결과기록, 예산집행 ...)

계획은 하나 이상의 진행이 일어날 것을 기대한다. 예를 들어 계좌 데이터가 생성이 되면 하나 이상의 계좌 거래 데이터가 발생할 것을 기대한다.

3.1.2 데이터 클래스

그림 1의 우측박스 내부의 육면체는 데이터 클래스와 데이터 클래스의 계층구조를 나타낸다. 데이터 클래스는 데이터 영역의 하위로 데이터 요소가 할당되는 공간이다. 데이터 분류를 하고자 하는 업무 영역의 특성, 복잡성, 규모에 따라 다양한 데이터 클래스들이 정의된다. 데이터 클래스들은 계층 구조를 가지고 있어 범위가 넓은 데이터 클래스는 계층구조 상위에 위치하고 데이터 범위가 좁은 데이터 클래스는 계층구조 하위에 위치한다. 즉, 상위의 데이터 클래스일수록 업무에 독립적이며 하위의 클래스일수록 업무의 특성이 반영된다. 이러한 계층구조의 데이터 클래스는 데이터를 검색할 때 검색의 범위를 줄여 주는 역할을 한다.

데이터 클래스의 정의는 데이터 영역에 배치된 데이터 요소를 업무 목적에 따라 그룹화 한 후 정의해야 한다. 그러나 업무에 따라 반드시 있어야 할 데이터 요소들이 있으므로 이를 기반으로 한 필수 데이터 클래스를 먼저 정의할 수 있다. 예를 들어, 농수산물 도소매를 위한 쇼핑몰과 소규모 의류 통신 판매를 위한 쇼핑몰의 데이터 요소들은 차이가 있으나 고객, 상품, 주문 등은 쇼핑몰에는 공통적으로 있어야 하는 데이터 요소들이므로 이들이 분류되는 데이터 클래스는 미리 선언될 수 있다. 하지만 농수산물 도매에는 실제 시장과 창고를 관리해야 하나 소규모 통신 판매에는 별도의 시장이나 창고가 없는 경우가 대부분이므로 이들의 정의는 분류과 정에서 수행해야 한다.

3.1.3 데이터 요소

데이터 요소는 데이터 클래스 계층구조 상에서 가장 하위 데이터 클래스의 한 인스턴스(instance)이며 데이터를 사용하는 과정에서 공유, 재사용, 통합의 대상이 되는 단위이다. 본질기반 데이터 분류 방법론에서 데이터 요소의 수준(granularity)은 ISO 11179 MDR의 데이터 요소 정의와 같은 엔티티(Entity) 단위 수준으로 정의한다. 즉, 데이터 요소는 하나 이상의 식별 속성을 포함한 두 개 이상의 속성과 두 개 이상의 데이터 인스

턴스를 가져야 한다[11]. DAMA(Data Management Association)[12]에서는 속성 수준으로 데이터 요소를 정의한다. 그러나 실제계의 대부분의 데이터들은 두 개 이상의 속성으로 구성된 DBMS의 테이블로 관리하기 때문에 본질기반 데이터 분류 방법론에서는 데이터 요소를 속성 수준으로 정의하지 않는다. 그리고 데이터 분류에 있어 데이터 요소는 DBMS의 테이블과는 정확히 일치하지 않는다.

DBMS의 데이터 테이블은 성능상의 이유, 혹은 잘못된 논리 엔티티 정의로 인해 종종 비정규화된 형태로 존재한다. 하나의 테이블에 여러 데이터 요소가 혼합해서 존재하거나 한 데이터 요소가 여러 테이블에 분할되어 존재하는 경우도 많다. 그러므로 독립된 정보 그룹이 비교적 정규화된 형태(일반적으로 3차 정규화까지를 권장)로 존재하는 것을 데이터 요소라고 할 수 있다.

데이터 분류의 1차적 목적은 공유, 재사용 및 통합 대상 확인에 있다. 따라서 데이터 요소 결정시 중요한 고려 사항은 데이터 분류 목적에 맞는 데이터 요소를 할당해야 한다. 즉, 조직의 모든 데이터를 분류하는 것보다는 데이터 분류 목적에 부합하는 데이터 만을 분류하는 것이 분류의 복잡성을 줄일 수 있다. 경우에 따라 데이터 클래스와 데이터 요소는 같은 이름을 가질 수 있다. 실 세계에서는 활동데이터 영역에 해당하는 많은 논리 엔티티들이 업무와 같은 이름을 갖기 때문이다. 주문, 배송, 계약, 결제, 예산 등이 그 예이다. 이때는 행위의 이름이 데이터 클래스의 이름이자 데이터 요소의 이름으로도 사용된다.

3.2 데이터 분류 절차

실세계의 데이터는 문서, 디지털 매체, 사람의 기억 등 여러 형태로 존재할 수 있겠지만 본질기반 데이터 분류 방법론의 공유, 재사용 그리고 통합 대상으로서의 데이터는 DBMS에서 관리하는 데이터로 한정한다. 따라서 본질기반 데이터 분류 방법론의 분류 대상인 데이터 요소는 데이터 모델링의 논리 엔티티 단위이다. 아래의 절차에서 언급하는 엔티티는 분류 과정상의 데이터 요소와 동일한 의미로 사용한다.

데이터 모델의 엔티티들은 크게 커널 엔티티, 코드 엔티티, 교차 엔티티, 종속 엔티티로 나눌 수 있다.

- 1) 커널 엔티티(Kernel Entity) - 사람, 자원, 장소와 같이 업무 종속적이지 않고 본래부터 존재하는 엔티티. 해당 데이터를 탄생시키기 위한 부모 엔티티가 존재하지 않는다[13].
- 2) 코드 엔티티(Code Entity) - 커널 엔티티와 같이 업무 종속적이지 않고 부모가 존재하지도 않는다. 그러나 관리하는 속성이 2~3개로 비교적 적고 구성하는 데이터 레코드의 양도 많지 않다.

- 3) 교차 엔티티(Associative, Relationship, Intersection Entity) - 커널 엔티티간의 다대다 관계를 해소하기 위한 엔티티로 업무적 목적에 의해 발생한다. 고객과 상품의 다대다 관계는 주문이라는 교차 엔티티에 의해 해소된다.

- 4) 종속 엔티티(Dependent Entity) - 엔티티의 1차 정규화 과정에서 발생한다. 회원이 독서, 음악감상, 낚시 등 하나 이상의 취미를 갖는다면 해당 속성은 복수 값을 가지므로 1차 정규화 과정에서 취미를 코드화한 코드 엔티티와 회원 취미라는 자식 엔티티를 생성한다. 종속 엔티티는 교차 엔티티와의 구분을 위하여 부모 중 하나가 코드 엔티티이거나 코드를 생성할 수 있는 데이터들로 한정한다. 또한 엔티티의 일부 속성에 대한 데이터 값의 변경 이력을 관리하는 이력 엔티티도 이에 속한다.

본질기반 데이터 분류 방법론에서 분류하는 데이터 요소들은 이 네 종류의 엔티티를 대상으로 하며 커널 엔티티, 코드 엔티티, 교차 엔티티, 종속 엔티티로 구분하여 분류한다. 본질기반 데이터 분류 방법론의 데이터 분류 절차는 2가지 과정의 8개 단계로 구성된다. 먼저 과정(Phase) 1은 5단계로 구성되며 하향식(Top-down)으로 커널 엔티티와 교차 엔티티를 레벨1과 레벨2 데이터 영역에 배치하는 것이다. 다음으로 배치된 데이터 요소를 상향식(Bottom-up)으로 그룹화 하여 데이터 클래스를 정의하고 정의된 데이터 클래스에 데이터 요소를 조정 배치하는 과정 2는 3개의 하위 단계(Step)로 구성되었다. 본질기반 데이터 분류 방법론에서는 이 8개의 과정을 모두 거쳐야 데이터 분류 구조의 완성과 데이터 요소의 배치가 완성된다. 그림 3은 2개의 과정과 내부의 8개의 단계를 통하여 데이터 요소가 제일 하위의 데이터 클래스에 배치되는 절차를 보인다.

과정(Phase) 1 : 데이터 영역에 데이터 요소 배치

단계(Step) 1 : 주체 데이터 영역에 커널 엔티티 배치. 커널 엔티티 중 사람, 조직, 또는 사람 및 조직이 업무를 수행하기 위해 부여 받은 역할에 해당하는 데이터들을 주체(Party) 데이터 영역에 배치(配置)한다. 고객, 사업체, 요금납부자, 배달자 등이 이에 속한다.

단계 2 : 자원 데이터 영역에 커널 엔티티 배치. 커널 엔티티 중 사람이나 조직이 활동을 하기 위해 사용하는 요소들을 자원(Resource) 데이터 영역에 배치한다. 도구, 건물, 자동차 등의 눈에 보이는 유형자원도 있지만 에너지, 법령, 표준, 지식 등도 무형의 자원에 해당한다. 직위, 직무, 직급, 상품구분 등의 코드 엔티티들도 모두 자원 데이터 영역에 배치한다.

단계 3 : 활동 데이터 영역에 교차 엔티티 배치. 교차 엔티티들은 커널 엔티티간의 다대다 관계를 해소하

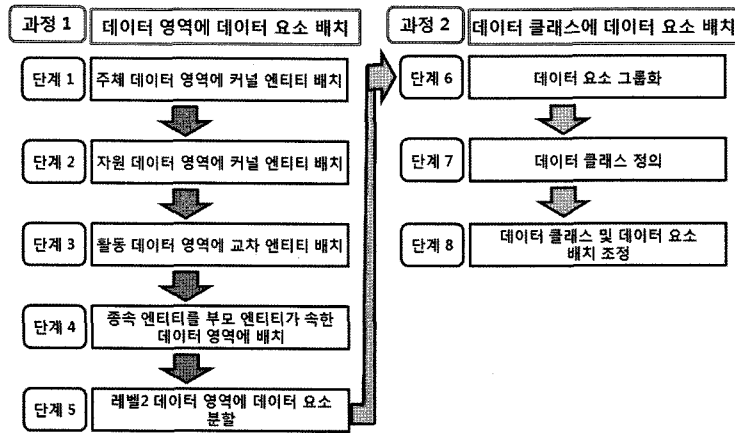


그림 3 데이터 분류 절차

는 과정에서 대부분 주체와 자원의 활동을 나타내므로 1차적으로 활동 데이터 영역에 배치한다. 그러나 일부 교차 엔티티는 주체나 자원으로 재배치될 수 있으며 이는 과정2에서 참고사항 3)을 반영하여 수행하도록 한다.

단계 4 : 종속 엔티티를 부모 엔티티가 속한 데이터 영역에 배치. 종속 엔티티도 교차 엔티티와 같이 두 개 이상의 부모를 갖는다. 그러나 데이터의 성격이 한쪽 부모에만 편향되므로 종속 엔티티는 그 부모가 속한 데이터 영역에 배치한다. 예를 들어 상품가격변동이력과 같은 이력 성격의 데이터들도 종속 엔티티로 상품이 속한 자원 데이터 영역에 배치한다.

단계 5 : 하위 데이터 영역에 데이터 요소 분할. 레벨1 데이터 영역의 3개 영역에 배치된 데이터 요소들을 각각 6개의 레벨2 데이터 영역으로 분할 재배치 한다. 데이터의 구분이 익숙하다면 1단계에서부터 레벨2 데이터 영역에 배치하는 것도 가능하다.

과정 2 : 데이터 클래스에 데이터 요소 배치

단계 6 : 데이터 요소 그룹화. 과정1에서 레벨2 데이터 영역 내 배치된 데이터 요소들을 업무 성격에 따라 그룹화한다. 데이터 요소들의 그룹화는 업무 영역 내의 분류 목적에 의존 한다. 예를 들어 건설교통부에서는 자동차를 배, 항공기와 같은 운송기 그룹으로 분류하나 관세청에서는 토지, 건물과 같은 과세단위 그룹화한다.

단계 7 : 데이터 클래스 정의. 그룹화된 데이터 요소들을 데이터 클래스로 정의한다. 이 때 데이터 클래스의 이름은 업무 영역에 의존하기 때문에 비슷한 성격의 데이터 요소들로 그룹화되었어도 업무 영역에 따른 데이터 분류의 데이터 클래스 이름은 다를 수 있다. 농약, 제조제, 비료는 농사를 위한 약품이지만 환경보호 업무에서는 독성물질로 명명한다. 반면에 같은 데이터 클래스 이름이 업무 영역에 따라 다른 데이터 요소들로 그

roup화되기도 한다. 백화점의 상품은 유형으로 고객이 운반 가능한 물건 만을 의미한다면 관광에서의 상품은 지역의 특산물 외에 그 지역의 자연 경관이나 심지어 물이나 공기까지도 상품이라고 정의할 수 있다. 또한 데이터 클래스 이름은 데이터 영역이 다르다면 한 분류 구조 내에서 중복이 허용된다. 즉, 관계라는 데이터 클래스는 주체 데이터 영역에도 존재하고 자원 데이터 영역에도 존재할 수 있다.

단계 8 : 데이터 클래스 및 데이터 요소 배치 조정. 이미 정의된 데이터 클래스를 세분화하여 더 하위의 데이터 클래스나 형체 데이터 클래스를 분화하는 등의 데이터 클래스 재정의가 가능하다. 데이터 클래스간 데이터 요소 배치를 조정한다. 필요 시 데이터 요소는 데이터 영역간 이동을 할 수 있다.

다음은 8개 단계의 데이터 분류 절차를 진행하면서 참고해야 할 사항들이다.

- 1) 무형 자원 정의 - 토지, 시설물, 주소 등은 '장소'의 개념으로 객관화된 자원으로 해석한다. 사이버 공간, 이메일 주소, IP주소, 좌표 등은 무형 자원 데이터 영역에 배치한다.
- 2) 코드 엔티티 정의 - 종종 코드 엔티티들이 주체 영역의 데이터와 혼동되는 경우도 있다. 3.2에서 언급한 코드 엔티티 정의와 다르게 구체적 정보를 관리해야 한다면 주체 성격의 데이터 요소가 된다. 그러나 대부분 코드 엔티티들은 구조가 간단하고 코드화한 데이터 값에 의미를 두기 때문에 자원 데이터 영역에 할당하는 것이 더 합당하다.
- 3) 교차 엔티티 배치 조정 - 과정 1의 단계 3에 의해 모든 교차 엔티티들을 1차적으로 활동 데이터 영역에 배치하였다. 그러나 다대다 재귀적 자기 참조를 해소하기 위한 교차 엔티티들의 경우와 같이 부모가

속한 데이터 영역에 할당해야 하는 교차 엔티티들도 있다. 예를 들어 패키지 상품은 여러 개의 상품이 하나의 다른 상품을 구성하고 한 상품은 여러 패키지에 구성될 수 있다. 이때 상품은 다대다의 재귀적 순환 관계를 갖게 되므로 이 관계는 별도의 상품구성이라는 엔티티를 생성하여 해소해야 한다. 이 과정에서 상품구성을 나타내는 엔티티는 부모인 상품이 속한 자원 데이터 영역에 할당하는 것이 가장 바람직하다.

4. 사례 연구

사례 연구에서는 본질기반 데이터 분류 방법론이 데이터 분류, 공유 데이터 후보 검색, 데이터 통합, MDR과의 연동에 효과적으로 사용될 수 있음을 보인다.

4.1 쇼핑물 데이터 분류

이 절에서는 인터넷 쇼핑물의 논리 데이터 모델 정보를 이용하여 데이터 분류를 수행한다. 한 업체는 소규모의류 인터넷 쇼핑물을 운영하고 있었으나 사업 확장으로 인한 시스템 개선을 앞두고 기존의 데이터 요소들을 정비하고자 한다. 인터넷 쇼핑물의 기존 데이터에 대한 논리 데이터 모델이 존재하므로 이를 3.2에서 정의한 데이터 분류 절차에 따라 일부 데이터들을 분류한다.

1) 과정 1 : 데이터 영역에 데이터 요소 할당

단계 1 : 주체 데이터 영역에 커널 엔티티 할당

- 쇼핑물의 고객, 회원, 직원, 상품관리자, 배송업체담당자, 납품업체담당자, 납품업체, 배송업체에 해당하는 데이터들은 상품에 대한 활동을 일으키는 주체 데이터 들이다.
- 이들은 데이터를 생성하기 위한 부모 엔티티가 필요하지 않은 개체나 조직에 해당하는 커널 엔티티 들이다. 이들을 주체(Party) 데이터 영역에 배치한다.

단계 2 : 자원 데이터 영역에 커널 엔티티 할당

- 쇼핑물에서 판매 대상이 되는 상품은 본래부터 존재해야 하는 커널 엔티티이다. 상품을 배송하는 주소를 위한 우편번호, 쇼핑물 운영을 위한 소비자약관 등도 업무 종속적이지 않은 커널 엔티티이다. 이들을 자원(Resource) 데이터 영역에 배치한다.
- 키워드와 금융기관 코드 엔티티를 자원 데이터 영역에 배치한다.

단계 3 : 활동 데이터 영역에 교차 엔티티 할당

- 주문, 계약 등은 고객과 상품간에 혹은 상품과 납품업체간에 업무 목적에 의해 생성된 교차 엔티티들이다. 교차 엔티티를 활동(Action) 데이터 영역에 배치한다.
- 고객 상담이라는 엔티티는 고객이 상품관리자와 상담한 내역을 관리하는 교차 엔티티로 이를 활동 데이터 영역에 배치한다.

이터 영역에 배치한다.

- 고객배송지는 한 고객이 동시에 여러 곳에 상품을 배송하기를 원하거나 한 군데 이상의 배송 주소를 관리하기 위한 데이터로 고객과 우편번호와의 교차 엔티티이다. 따라서 활동 데이터 영역에 배치한다.

단계 4 : 종속 엔티티를 부모 엔티티가 속한 데이터 영역에 할당

- 회원 취미는 회원이라는 부모 엔티티의 종속되는 데이터이므로 그 부모가 속한 주체 데이터 영역에 배치한다.
- 상품가격변동이력이란 이력 데이터를 부모인 상품이 속한 자원 데이터 영역에 배치한다.

그림 4는 레벨1 데이터 영역에 데이터 요소를 배치한 결과이다.

Party	Resource	Action
- 고객	- 상품	- 주문
- 회원	- 금융기관	- 고객배송지
- 직원	- 우편번호	- 계약
- 상품관리자	- 소비자약관	- 고객상담
- 배송업체	- 키워드	
- 납품업체	- 상품가격변동이력	
- 회원취미		
- 납품업체담당자		
- 배송업체담당자		

그림 4 데이터 요소 배치(과정1의 단계1~과정1의 단계4)

단계 5 : 하위 데이터 영역에 데이터 요소 분할

- 단계 1에서 레벨1 주체 데이터 영역에 배치한 데이터 요소를 개체와 조직으로 구분된 레벨2 데이터 영역으로 분할한다.
- 이 경우 배송업체와 납품업체를 주체 아래의 조직으로 배치하고 나머지는 개체 데이터 영역에 배치한다.
- 단계 2에서 우편번호, 소비자약관은 무형 자원 데이터 영역에 배치하고 나머지는 유형 자원 데이터 영역에 배치한다.
- 키워드, 금융기관 코드 데이터는 무형 자원 데이터 영역에 배치한다.
- 단계 3에서 배치한 데이터 요소들 중 주문, 계약은 계획 데이터 영역에 배치한다. 고객상담은 계획은 없고 진행만 있는 데이터로 이를 진행 데이터 영역에 배치한다.

그림 5는 레벨1 데이터 영역에 배치된 데이터 요소들이 레벨2 데이터 영역으로 재배치된 결과이다.

2) 과정 2 : 데이터 클래스에 데이터 요소 할당

Party		Resource		Action	
Individual	Organization	Visible	Invisible	Plan	Progress
- 고객 - 직원 - 회원 - 회사 - 사원 - 관리자 - 배송업체 - 납품업체 - 담당자 - 담당자	- 납품업체 - 배송업체	- 상품 - 상품 - 가격 - 동이력	- 우편번호 - 키워드 - 소비자 - 약관 - 금융 - 기관	- 주문 - 계약 - 고객 - 배송지	- 고객상담

그림 5 데이터 요소 배치(과정1의 단계5)

단계 6 : 데이터 요소 그룹화

- 레벨2 데이터 영역에 배치된 데이터 요소들을 쇼핑물 업무 성격에 맞게 그룹화 한다.
 - 고객, 회원은 상품을 구입하는 개체에 대한 정보이고 직원, 배송업체담당자, 납품업체담당자는 쇼핑물 운영을 위한 개체들이다. 이들을 두 그룹으로 나눈다.
 - 키워드는 금융기관과 같은 코드들과 같이 그룹화한다.
- 아래의 그림 6은 단계 6에 따라서 데이터 요소들을 그룹화한 결과이다.

단계 7 : 데이터 클래스 정의

- 쇼핑물 업무를 위한 데이터 클래스를 정의한다.
- 회원은 아이디번호, 계좌번호 등 개인 정보를 더 제공하는 일부 고객이다. 회원과 고객이 있는 그룹을 고객이라는 데이터 클래스로 정의한다.
- 직원, 상품관리자, 배송업체담당자, 납품업체담당자는 쇼핑물을 운영하는 직원들의 다양한 역할들이다. 이들을 위해 운영자라는 클래스를 정의하였다.
- 키워드, 금융기관 그리고 그 외의 코드들을 위해 코드라는 데이터 클래스를 정의하였다.
- 재고 현황을 위해서는 계획 데이터 영역의 재고 데이터 클래스를 재고 조사는 진행 데이터 영역의 같은 이름의 재고 데이터 클래스를 정의하여 배치한다.

단계 8 : 데이터 클래스 및 데이터 요소 배치 조정

- 고객배송지는 고객과 우편번호와의 교차 엔티티로 과정1의 단계3에서 활동 데이터 영역에 배치하였다. 그러나 고객배송지는 업무적 활동의 계획이나 진행에 속한다기보다는 배송이라는 활동을 생성하기 위한 자원 데이터에 속한다. 고객배송지를 자원의 주

소 데이터 클래스로 이동 배치한다.

그림 7은 그룹화한 데이터 요소에 데이터 클래스를 정의하고 각 데이터 클래스에 데이터 요소를 조정 배치한 결과이다.

4.2 공유 데이터 후보 검색

다음 사례는 완성된 데이터 분류를 이용하여 공유 가능한 데이터를 검색하는 경우이다. 공유데이터 후보 검색은 목적하는 데이터를 데이터 분류표에서 빠르게 찾는 것이다. 이는 업무에서 요구하는 데이터를 결정하고 기존의 데이터 분류표에 같은 목적의 데이터가 존재하는지를 확인하는 절차로 이루어진다. 즉, 업무에서 필요로 하는 엔티티 후보를 먼저 도출하여 기존 데이터 분류표의 데이터 클래스에 배치한 후 같은 데이터 클래스 내에서 공유 가능 데이터를 결정하면 된다. 이번 사례에서는 업무만 정의된 상황에서 엔티티 후보를 추출하여 이미 만들어진 데이터 분류표에 배치하는 과정을 보인다. 결과적으로 추출된 엔티티 후보가 배치된 데이터 클래스에서 유사 데이터가 존재함을 확인하고 기존 데이터를 공유하기로 결정하는 과정을 보인다.

4.1에서 언급한 쇼핑물에 이벤트 관리 업무를 신규로 추가하고자 한다. 시스템 구축 전에 회사의 데이터 관리자는 추가하고자 하는 업무에 필요한 데이터 요소를 업무 설명에서 추출하려고 한다. 정교하지는 않지만 1차 추출한 데이터 요소들을 이미 구성한 데이터 분류표에 대응시킨다면 공유 가능한 데이터 요소를 찾을 수 있어 데이터의 중복 발생을 예방할 수 있다. 아래의 업무 설명에서 3.2에서 언급한 엔티티 종류에 따라 신규 요구가 있는 데이터 요소를 종류별로 추출하고 과정1과 과정2

Party		Resource		Action	
Individual	Organization	Visible	Invisible	Plan	Progress
- 고객 - 회원 - 회원 - 취미 - 직원 - 상품 - 관리자 - 배송 - 업체 - 담당 - 납 - 품 - 업체 - 담당 - 자	- 납품업체 - 배송업체	- 상품 - 상품 - 가격 - 동이력	- 우편번호 - 소비자 - 약관 - 금융 - 기관 - 키워드	- 주문 - 계약 - 고객 - 배송지	- 고객상담

그림 6 데이터 요소 그룹화(과정2의 단계6)

Party				Resource				Action			
Individual		Organization		Visible		Invisible		Plan		Progress	
class	DE	class	DE	class	DE	class	DE	class	DE	class	DE
고객	-고객 -회원 -회원취미	업체	-납품업체 -배송업체	상품	-상품 -상품가격 -변동이력	주소	-우편번호 -고객 -배송지	계약	- 계약	상담	-고객상담
						규정	-소비자 규정	주문	- 주문		
운영자	-직원 -상품관리자 -배송업체 담당자 -납품업체 담당자					코드	-금융기관 -키워드	재고	-재고현황	재고	-재고조사

DE : 데이터 요소(Data Element)

그림 7 데이터 클래스 정의(과정2의 단계7 ~ 과정2의 단계8)

의 데이터 분류 절차에 따라 배치한다. 이번 사례 설명에서는 종류별 엔티티를 추출하는 과정의 분류 절차를 포함하여 설명한다.

이벤트 관리 업무 : 여름 휴가철을 수영복 판매 증가를 위한 이벤트나 명절을 위한 아동의 특별 의상을 이벤트, 혹은 매년 쇼핑몰 오픈 기념일을 위한 이벤트 등 다양한 이벤트가 존재할 수 있다. 이벤트 관리자는 특정 기간에 이벤트를 개최하여 이벤트별 상품 종류 선택 및 할인율을 결정한다. 한 고객은 한 이벤트에서 복수개의 항목에 응모하고 당첨될 수 있다.

- 1) 커널 엔티티(Kernel Entity) - 업무를 발생하는 이벤트 관리자, 업무의 대상이 되는 이벤트 상품은 각각 주체와 자원에 해당한다. 이들을 추출하여 운영자, 상품 데이터 클래스에 배치한다.
- 2) 코드 엔티티(Code Entity) - 이벤트는 여름 휴가철, 가을 명절 등 여러 종류가 있고 쇼핑몰 오픈 기념일과 같은 이벤트는 매년 발생한다. 이벤트 종류를 관리하기 위한 이벤트 종류 코드 엔티티를 추출하여 코드 데이터 클래스에 배치한다.
- 3) 교차 엔티티(Associative, Relationship, Intersection Entity) - 이벤트 응모자는 이벤트와 고객간의 관계에 의해 발생한 교차 엔티티이므로 이를 활동 데이터 영역에 배치한다. 그러나 이벤트 응모자는 이벤트라는 활동의 계획이나 진행 데이터는 아니므로 이를 개체 데이터 영역에 관계라는 데이터 클래스를 선언한 후 해당 데이터 클래스로 이동한다.
- 4) 종속 엔티티(Dependent Entity) - 이벤트 당첨자는 이벤트 응모자 중 이벤트 성격에 부합하는 기준에 의해 선택된 회원의 일부이다. 이벤트 응모자라는 엔티티에서 당첨이라는 속성을 두어 관리할 수도 있으나 하나 이상의 당첨 기준이 있을 수 있으므로 이벤트 당첨자는 종속 엔티티로 구분될 수 있

고 이를 이벤트 응모자가 있는 데이터 클래스에 같이 배치한다.

이 과정에서 이벤트 관리자는 이벤트별 상품과 상품 할인율을 관리한다. 초기에는 이벤트 관리자를 운영자 데이터 클래스에 배치하였다. 그러나 실제 업무에 있어서도 이벤트 관리 업무의 대부분이 상품 관리와 유사하므로 이벤트 관리자라는 별도 엔티티를 생성하지 않고 상품관리자 데이터를 공유하기로 하였다.

그림 8은 업무로부터 직접 데이터 요소를 도출하여 배치 절차 수행 후 쇼핑몰 데이터 분류이다. 그림에서 상품관리자(이벤트 담당자)라고 표현한 데이터 요소는 실제 상품관리자라는 데이터 요소만 존재하는 것이다.

4.3 데이터 통합

데이터 분류가 완료되었다면 데이터 분류표에 배치된 데이터 요소들을 확인하여 데이터 통합을 고려할 수 있다. 특히 데이터 아키텍처적 관점이 아닌 개발자의 관점에서 오랫동안 점진적으로 진행된 시스템일수록 일반적으로 중복 데이터의 사용 빈도가 높다. 데이터 본질에 기반한 데이터 분류는 동질의 데이터들이 같은 클래스에 모일 수 있게 하고 이를 기반으로 통합 데이터 후보를 확인할 수 있다. 같은 클래스에 모인 데이터 요소들의 메타 정보를 확인하여 비슷한 성격의 데이터 요소로 판명되면 통합 후보가 될 수 있기 때문이다.

사례의 쇼핑몰의 경우 기존 데이터 모델은 논리 중심적이 아닌 물리 테이블 위주로 데이터 요소가 선언되어 있다. 고객, 회원, 직원, 상품관리자, 배송업체담당자, 납품업체담당자, 납품업체, 배송업체 등의 데이터들을 데이터 분류 작업을 통하여 데이터 클래스에 배치하였더니 상품관리자, 배송업체담당자, 납품업체담당자, 이벤트 담당자들 간의 유사성, 그리고 납품업체와 배송업체 데이터들의 유사성을 확인할 수 있었다. 따라서 상담, 배송업체, 납품업체, 이벤트 들을 구분하는 업무구분이라는 코

Party				Resource				Action							
Individual		Organization		Visible		Invisible		Plan		Progress					
class	DE	class	DE	class	DE	class	DE	class	DE	class	DE				
고객	-고객 회원 회원취미	업체	납품업체 배송업체	상품	-상품 상품가격 변동이력 -이벤트 상품	주소	우편번호 고객배송지	계약	계약	상담	-고객상담				
운영자	-직원 -상품관리자 (이벤트담당자) -배송업체 담당자 -납품업체 담당자		코드		금융기관 키워드 -이벤트 종류	규정	소비자약관	주문	주문			재고	-재고현황	재고	재고조사
						관계	-이벤트 응모자 -이벤트 당첨자								

DE : 데이터 요소(Data Element)

그림 8 쇼핑몰 데이터의 공유 데이터 후보 검색

Party				Resource				Action							
Individual		Organization		Visible		Invisible		Plan		Progress					
class	DE	class	DE	class	DE	class	DE	class	DE	class	DE				
고객	-고객 회원 회원취미	업체	납품업체 배송업체 업체	상품	-상품 상품가격 변동이력 -이벤트 상품	주소	우편번호 고객배송지	계약	계약	상담	-고객상담				
운영자	-직원 -상품관리자 (이벤트 관리자) -배송업체 담당자 -납품업체 담당자 -업무별 담당자		코드		금융기관 키워드 -이벤트 종류	규정	소비자약관	주문	주문			재고	-재고현황	재고	재고조사
						관계	-이벤트 응모자 -이벤트 당첨자								

그림 9 쇼핑몰 중복 데이터 통합

드를 생성하여 업무별 관점에서 분리되어 생성되었던 업무 담당자 엔티티들을 업무별담당자라는 엔티티로 통합하였다. 또한 납품업체와 배송업체는 업체라는 엔티티를 생성하여 통합하고 이 과정에서 업체구분 코드를 생성하였다. 그림 9에서 통합 전 데이터 요소는 지워지고 새로 생성된 데이터 요소는 굵은 글씨체로 표현하였다.

4.4 데이터 분류와 MDR(Meta Data Registry)의 옹합

공유 및 재사용 데이터 검색의 범위를 한 기업에서가 아닌 정부 단위로 확장할 수 있다. 다음의 예는 범정부 단위에서 정부 부처간 데이터 공유를 위한 예를 보인다. 조직이 다르다면 모든 데이터를 확인하여 공유 가능성을 확인하는 것은 매우 어렵다. 조직은 데이터 분류표나 데이터에 대한 메타 정보 수준은 공개 가능하나 내부

데이터는 조직간 데이터 교환에 대한 조약 후가 아니면 공개하지 않는 것이 대부분이기 때문이다. 검색한 공유 대상 데이터의 메타 데이터 정보를 MDR에서 검색하여 세부 정보를 확인할 수 있다면 공유 대상 데이터에 대한 신뢰도와 표준에 의한 데이터 상호 교환의 효과를 더 높일 수 있다.

보건복지부는 데이터 분류를 유지하고 있으며 식품 안전을 위해 현재는 보유하고 있지 않은 '위해성 수입 식품' 데이터 구축을 희망한다. 보건복지부는 '위해성 수입 식품' 데이터가 유형 자원 데이터 영역에 있어야 한다고 판단하였다. 그리고 해당 데이터를 처음부터 새로 구축하는 것보다는 다른 부처에 유사 데이터가 있다면 이를 활용하기로 하였다. 이를 위하여 다른 부처의 데이

	Party				Resource				Action			
	Individual		Organization		Visible		Invisible		Plan		Progress	
	class	DE	class	DE	class	DE	class	DE	class	DE	class	DE
농림수산부					약품	농수산물성분	규정	검역관련 국제기준			검사	농수산물품질검사결과 축산물위생감시결과 수입검역결과
					성분	동물용의약품 농약						
식품의약품안전청					식품	가공식품 수입식품					검사	식품위생점검결과 의약품안전관리점검결과 의약품안전평가결과 식품성분검사결과 마약관리현황
					약품	의약품 마약류						
					성분	위해성분						
행정안전부					약품	의약품						
					독물	독극물						

DE : 데이터 요소 (Data Element)

Data elements	농수산물품질검사결과	Data elements	수입검역결과	Data elements	식품성분검사결과
Action - Progress	Inspect	Action - Progress	Inspect	Action - Progress	Inspect
Attached	농림수산부	Attached	농림수산부	Attached	식품의약품안전청
Definition	농수산물의 품질 및 위해성분 검사 결과...	Definition	수입식품에 대한 성분 검사 결과	Definition	국내에서 판매되는 식품의 성분 검사 결과
Create	품질 검사 항목에 따라 품목 당 표준 검사 결과로 생성	Create	매일 약 50개의 수입식품을 검사	Create	신규 식품이나 의의가 떨어진 식품에 대한 검사
Update	수정하지 않고 검사 결과가 달라지면 이력 생성	Update	수정하지 않고 검사 결과가 달라지면 이력 생성	Update	수정하지 않고 검사 결과가 달라지면 이력 생성
Delete	삭제 없음	Delete	삭제 없음	Delete	삭제 없음
Period	영구 보존	Period	영구 보존	Period	영구 보존
Amount	약 50,000,000	Amount	About 1,000,000	Amount	About 100,000,000
Attributes	Keys : 식품종류, 검사항목, 검사일	Attributes	Keys : 수입품목, 검사항목, 검사일	Attributes	Keys : 식품항목, 검사항목, 검사일
Name	Domain	Mandatory	Name	Domain	Mandatory
식품종류	string	Yes	수입품목	string	Yes
검사항목	string	Yes	검사항목	string	Yes
검사일	date	Yes	검사일	date	Yes
...

MDR (Meta Data Registry) - ISO 11179를 따른 데이터 구조

그림 10 3개 부처의 데이터 분류표와 메타 데이터 비교

터 정보를 검색하고자 한다.

보건 복지부는 자부처의 데이터 분류에서 '위해성 수입 식품'이 유형 자원 데이터 영역에 있으므로 타부처의 데이터 분류도 유형 자원 데이터 영역부터 검색을 시작하였다. 농림수산식품부, 식품의약품안전청, 행정안전부 등에서 관리하는 데이터 분류에서 '수입식품', '농수산물 성분', '독물', '위해성분' 등이 유사하긴 하나 바로 사용할 수 있는 공유 데이터로는 적당하지 않았다. 따라서 활동 데이터 영역에 존재하는 것에서 '농수산물품질검사 결과', '수입검역결과', '식품성분검사결과'의 MDR에 기록된 메타 데이터를 확인하고 여기에서 공유 대상 데이터를 추출할 수 있다고 판단했다. 농림수산식품부는 검사나 검역 활동 후의 결과를 보관하고 있다. 여기에 식품의약품안전청과 행정안전부에서 관리하는 '위해성분'과 '독물'에 특정 기준치를 반영하여 '위해성'이 있다고 판단하는 것은 보건 복지부에서도 수행할 수 있다고 판단하였다.

그림 10은 농림수산식품부, 식품의약품안전청, 행정안전부의 데이터 분류표에서 '농수산물품질검사결과', '수

입검역결과', '식품성분검사결과'를 검색한 후 이들 데이터의 메타 정보를 확인한 결과이다. 여기에서 사용한 메타 데이터 표준은 ISO 11179 규약을 따른 것이다. 이러한 일련의 과정이 범정부 단위의 데이터 관리 위원회에서 권한이 있는 데이터 관리자에게 웹등을 통하여 제공된다면 범정부 단위의 데이터 공유는 활발해지고 중복 데이터 생산 가능성은 줄어들 것이다.

5. 결론

대규모 조직의 데이터 관리에는 데이터 관리 효율성과 데이터 품질 향상을 위하여 데이터 중복을 억제하고 기존 데이터의 공유와 재사용을 할 수 있는 방법이 필요하다. 이를 위해서 기존 데이터의 체계적 구조화와 효율적인 검색이 지원되어야 한다. 데이터 분류는 보유한 데이터의 정리를 통하여 원하는 데이터의 빠른 검색을 가능하게 한다. 데이터 검색을 위한 시맨틱 웹과 메타데이터 저장소는 빠르고 정확한 데이터 관련 정보의 검색을 가능하게 하지만 초기 구축과 유지 보수에 막대한 비용이 필요하다. 이에 비하여 데이터 분류는 시맨틱 웹

이나 MDR에 비하여 저비용으로 단시간에 구축할 수 있다.

본질기반 데이터 분류 방법론은 데이터 발생에 기초한 데이터 본질 특성에 따른 데이터 분류 체계를 제안한다. 본질기반 데이터 분류 방법론의 데이터 영역과 데이터 클래스는 간단하지만 견고한 기초 데이터 분류 구조를 제공한다. 또한 본질기반 데이터 분류 방법론의 분류 구조를 지원하는 데이터 분류 절차를 이용하여 다양한 데이터 요소들을 제안된 분류 체계에 맞게 배치하는 방법을 보인다. 이러한 절차는 본질기반 데이터 분류 방법론이 다양한 업무 카테고리에서 활용될 수 있는 확장성과 일반성을 제공해 준다. 사례연구에서는 본질기반 데이터 분류 방법론이 효과적으로 실제 쇼핑몰 관리 업무에 적용될 수 있음을 보인다.

향후 연구에서는 기존의 시맨틱 웹과 MDR에서 요구하는 데이터 분류에 본질기반 데이터 분류 방법론이 효과적이며 효율적으로 대응될 수 있도록 구체적인 연구를 진행할 것이다.

참 고 문 헌

[1] Korea e-government Law Article 9, 11, 13, 21, partially revised in February 29, 2008.

[2] FEA DRM, "The Data Reference Model Version 2.0," FEA (Federal Enterprise Architecture) US, November 17, 2005.

[3] FEA DRM, "Data Reference Model 1.0," FEA Program, Office of Management and Budget, vol. 1, ver. 1.0, 2004.

[4] Jordan, R., "A Methodology for Developing a Taxonomy - A Subject Oriented Approach," International Symposium on Ontology-Metamodeling, 2006.

[5] A. Gutierrez, J. Orozco, A. Serrano, W. T. Golden, Acton, K. Conboy, "A ADEVELOPING A TAXONOMY FOR THE UNDERSTANDING OF BUSINESS AND IT ALIGNMENT PARADIGMS AND TOOLS," Proceedings of the Sixteenth European Conference on Information Systems, 2008.

[6] M. Chisholm, "Governance for Taxonomic Reference Data," <http://www.tdan.com/view-articles/9379>, IRMUK and the Data Governance Conference Europe 2009, 2009. (get at March 30 2009)

[7] T. Bryce, "The Benefits of a Data Taxonomy," <http://it.toolbox.com/blogs/irm-blog/the-benefits-of-a-data-taxonomy-4916>, July 11 2005 (get at April 1 2009).

[8] E. Bradley, L. Curry, K. Devers, "Qualitative data analysis for health services research: developing taxonomy, themes, and theory," Health Services Research. Volume 42 Issue 4, pp.1758-1772, Blackwell Publishing Ltd., 2007.

[9] Biological classification, http://en.wikipedia.org/wiki/Biological_classification, (get at October 27 2009).

[10] F. A. Stafleu, "Linnaeus and the Linnaeans : The spreading of their ideas in systematic botany," pp. 1753-1789, Utrecht, 1971.

[11] C. Fleming, B. Halle, "Handbook of Relational Database Design," p.14, Addison-Wesley Professional, 1989.

[12] "The DAMA Dictionary of Data Management," The Data Management Association, 1st Edition 2008.

[13] S. Alagić, "Relational database technology," p.92, Springer-Verlag, 1986.



최 미 영

1992년 포항공과대학 전자계산학과(석사). 2005년~현재 고려대학교 컴퓨터학과 박사과정. 2004년~현재 ISO/IEC JTC1/SC32 전문위원회(전문위원). 관심 분야는 데이터 아키텍처, 데이터 모델링, 상용 DBMS 성능 튜닝, 데이터 정보전략설계, 정보 보안 및 분석



문 장 주

2004년 고려대학교 컴퓨터학과 박사. 2005년 고려대학교 정보보호대학원 연구교수 2006년 건국대학교 컴퓨터응용과학부 조교수. 2006년~현재 건국대학교 항공우주정보시스템공학과 조교수. 관심분야는 실시간 임베디드 시스템, 컴퓨터 보안, 시스템통합, 정보보호, 데이터베이스



백 두 권

1974년 고려대학교 수학과(학사). 1977년 고려대학교 대학원 산업공학과(석사). 1983년 Wayne State Univ. 전산학과(석사), 1985년 Wayne State Univ. 전산학과(박사). 1989년~2007년 (사)한국정보과학회(이사/평의원/부회장). 1986년~현재 고려대학교 컴퓨터-전파통신공학과(교수). 1991년~현재 (사)한국시뮬레이션학회(이사/부회장/감사/회장/고문). 1991년~현재 ISO/IEC JTC1/SC32 전문위원회(위원장). 2001년~현재 (사)도산아카데미(원장) 2002년~2004년 고려대학교 정보통신대학(초대학장). 2004년~2005년 (사)한국정보처리학회(부회장) 2009년 고려대학교 정보통신대학 학장. 관심 분야는 메타데이터, 소프트웨어공학, 데이터공학, 컴포넌트 기반 시스템, 메타데이터 레지스트리, 프로젝트 매니지먼트등



권 주 흡

1999년 Wayne State Univ. 전기/컴퓨터공학학과(석사). 2005년 고려대학교 컴퓨터학과 박사. 관심분야는 special emphasis on reasoning on taxonomic structure and ontology integration를 이용한 온톨로지 기반의 지식 표현, 엔터프라이즈 아키텍처, 시멘틱 웹



이 영 무

2001년 KAIST 전기전자공학과(박사). 현재 지식경제부 사무관. 관심분야는 임베디드 시스템, 모델링 & 시뮬레이션, 시스템 통합