

데이터 웨어하우징을 위한 메타데이터의 構造

여러 해 동안 데이터 웨어하우징을 위한 메타데이터의 중요성을 피력한 자료들이 출판되어 왔다. 간단히 말해 메타데이터는 데이터에 관한 정보이며, 상용(商用) 이용자들에게 뿐만 아니라 데이터 웨어하우징 관리자들과 개발자들을 위해서 중요한 역할을 한다. 메타데이터가 없다면 상용 이용자들은 자신들이 여행하는 도시에 관한 정보가 없는 관광객들 같은 꼴일 것이며, 데이터 웨어하우징 관리자들은 자신들이 관리하는 도시의 크기와 그 도시가 얼마나 빠르게 팽창하는지에 관한 정보가 없는 행정 관리들과 같은 꼴이 될 것이다.

메타데이터의 중요성에도 불구하고, 메타데이터는 다수의 데이터 웨어하우징 프로젝트들 중 가장 도외시(度外視)되는 부분으로 여전히 남아있다. 관계자들은 “나중에 걱정하지” 하는 식으로 메타데이터의 문제에 접근한다.

데이터 추출/변형(抽出/變形) 툴들은 메타데이터 관련 부분의 변환 작업의 대부분을 담당한다. 몇몇 front-end query tool들 역시 그러한 툴들과 데이터베이스 엔진들로부터 메타데이터를 읽어 들일 수 있다. 그러나 케이스 툴들과, 상속 데이터, front-end tool들, 데이터베이스 엔진들 및 데이터 웨어하우징 관리 툴 등에는, 사용자들의 차질없는 접근을 위해 통합되어야 할 고립된 중요한 메타데이터가 다수 있다. 또한, 그러한 상이(相異)한 source들에 위치한 메타데이터의 대부분은 오래되었거나 서로 일치하지 않는다.

메타데이터를 합병하고, 관리하며, 최신의 것으로 유지하려는 노력의 부족은 데이터 웨어하우징 프로젝트의 신뢰성과 성공을 위협하는 몇몇 문제점들을 초래할 수 있다.

· 해당 웨어하우스의 신뢰성의 감소를 유발하는 그릇된 정보

상기한 문제점을 잘 보여주는 예가 어느 한 회사

의 다양한 하부 조직들이 보유한 재고의 유형을 연구하는 어느 법인 의사결정지원시스템(DSS) 분석가의 경우이다. 그는 어느 날 갑자기 중장비 부서의 설명이 불가능한 재고 증가를 감지했다. 그 분석가는 해당 중장비 부서의 관리자와 상의했고, 그 관리자는 직원들에게 원인을 물어보아야 했다. 며칠간의 조사와 생산성의 악화를 거친 뒤, 그들은 문제의 기간중 당 회사의 재조직을 위한 짧은 시간동안 해당 중장비 데이터에 배전(配電) 트랜스들이 포함되어 있었다는 사실을 알게 되었다.

이 문제는 데이터 웨어하우징 부서와 영업부서의 직원인 상기의 법인 DSS 분석가 사이에 많은 고충을 야기했다. 그러한 문제는 해당 기간중 중장비의 정의에 관한 내용에 맞는 메타데이터가 있었다면 피할 수 있었을 것이다.

상기의 문제점은 우리로 하여금 메타데이터는 최신의 것이어야 하고 자세해야 할 뿐 아니라, 올바른 해석과 내용에 맞아야 한다는 또 다른 시각을 가져다 주었다.

■ 내용(Context) : 상기의 예에서 사용된 질의도구가 중장비와 문제의 기간중 보유했던 것들에 대한 올바른 정의를 지니고 있었다면 그러한 유래하지 못한 상황은 쉽게 피할 수 있었을 것이다. 현재

의 메타데이터 기술은 사용자가 질의의 결과를 찾아볼 때 내용 메타데이터를 제공하지 않는다.

Contextual 메타데이터는 하나의 질의에 대한 결과와 함께 해당 데이터가 작성되었을 당시에 존재한 데이터간의 관계들을 포함한 모든 내용 정보를 제공하여야 한다. 이러한 일은 메타데이터 관리 도구들과, 메타데이터 저장소들, 그리고 front-end tool들 사이의 올바른 통합을 필요로 한다.

■ 버전관리(versioning) : 과거의 메타데이터 version들을 보존하지 않고서 올바른 contextual 데이터를 제공하기는 불가능하다. 위에서 언급된 예에서 보듯이, 최신의, 그리고 상세한 메타데이터를 보유하는 것도 문제의 기간동안을 위한 적절한 메타데이터 version이 보존되지 못했다면 아무 쓸모가 없을 것이다.

조직들이 합병(合併), 주식 배분(配分), 사업의 매입(買入) 및 내부구조의 재편(再編) 등으로 인한 사업상의 여러 변화들을 거침에 따라, 사업의 실체들은 변화를 겪는다. 이러한 변화는 상황에 맞게 서로 다른 저장 메타데이터 version을 유지함으로써 세심하게 관리되어야 한다.

■ 데이터의 질에 관한 메타데이터 : 대부분의 조직들은 열악한 검증 법칙들과, 특정 분야의 단순한 오용(誤用)과 과용(過用)으로 인한, 현존 데이터와 외부 데이터가 지닌 품질의 현실을 감내(堪耐)해야 한다. 산업 현장에는 이러한 경우에 부합하는 수많은 예들이 존재한다.

데이터의 질을 분석하고 그 결과들을 연관된 메타데이터로 문서화하는 일은 데이터 웨어하우징 프로젝트의 초기 단계나 상호작용 기간 동안 매우 중요하다.

· 올바른 정보 습득의 불가능

적절한 메타데이터가 없이는 사용자들은 데이터 마트가 데이터 웨어하우징내에 이미 존재하는 특정 정보조차도 알지 못할 것이다.

이러한 문제가 먼저 논의된 문제점보다 더욱 심

각하게 들리겠지만, 필자는 이 문제가 성격상 앞서 기술한 문제보다는 덜 심각하다고 생각한다. 그릇된 정보는 아무런 정보를 얻지 못하는 것보다 더 심각한 문제이다.

· 생산성 감소

급격한 생산성의 감소를 유발하는 상황의 와중에서 사용자는 올바른 방법으로 적합한 정보를 얻을 수 없다. 예를 들어, 어느 회사의 판매담당 부사장이 해당 회사의 마지막 자금 조달 사업 이래로 참여한 고객들의 평균 수익성을 알고자 한다고 가정하자. 만약 평균 수익성과 이것을 계산하는 법 및 고려되는 자금조달 사업이 어느 것인지를 정의하는 메타데이터가 없다면, 그는 그러한 정의들을 연구하기 위해 많은 시간을 소비할 것이며 그것은 생산성의 감소를 초래할 것이다.

· 예기치 않은 작동

질의 수요 예측을 위한 정보의 부족은 장시간을 필요로 하는 질의들을 질의 수요가 한산한 시간대에 실행되도록 예정하는 시스템의 능력을 손상시킨다. 이러한 문제는 시스템의 예기치 않은 작동을 야기할 수 있다.

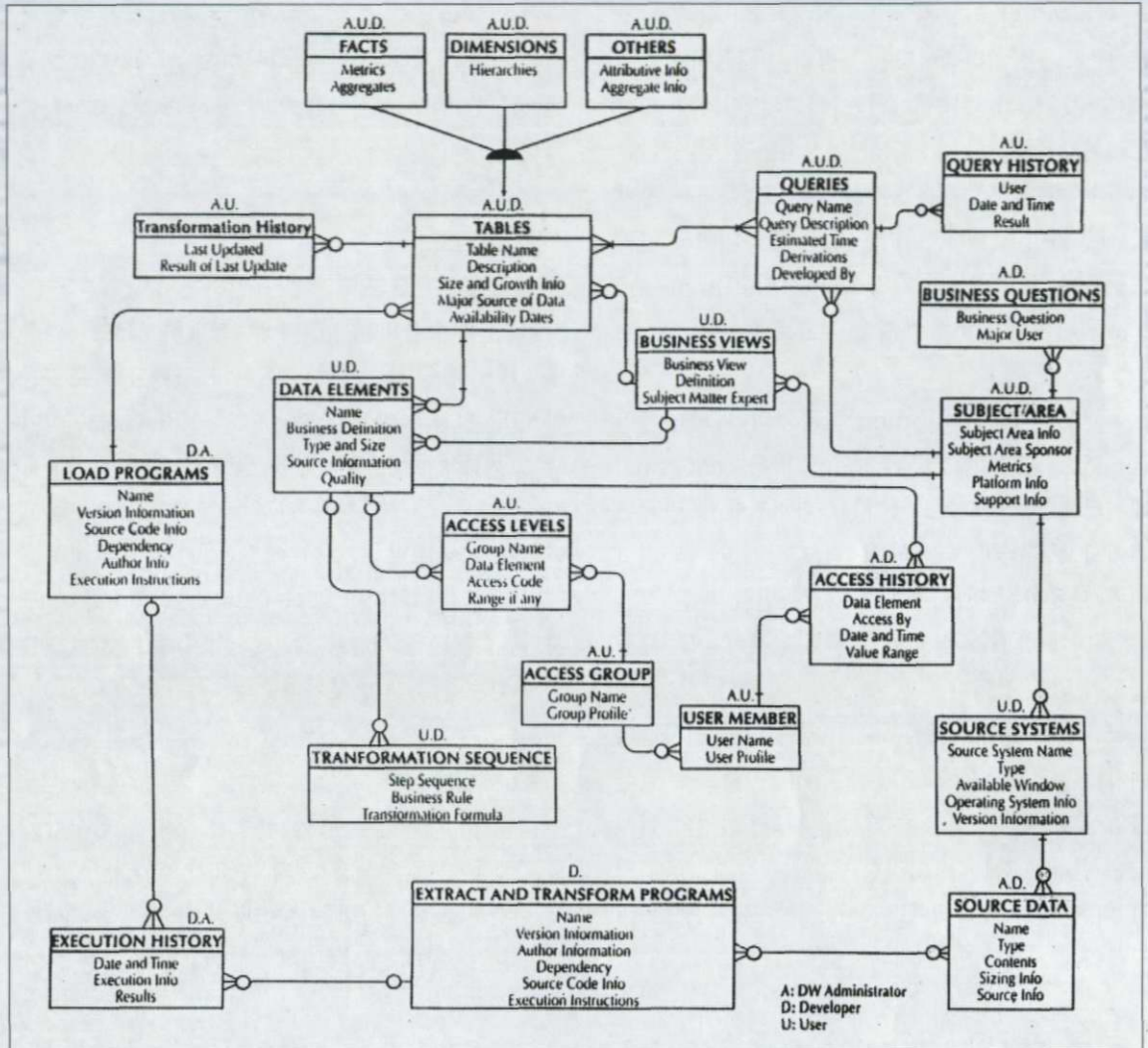
그런 문제를 피하기 위해서 메타데이터는 적어도 외관상 복잡하고 장황(張皇)한 질의들의 추정 실행 시간과 예상되는 결과 열의 수에 관한 정보를 지니고 있어야 한다. DW 환경하에서 그러한 일반적이지 않은 질의들 전부를 예측하기란 거의 불가능하다는 점은 이해할 만하다. 그러나 다양한 질의들에 관한 간단한 약력 파일을 제작하면 상기의 문제점을 완화시키는데 큰 도움을 줄 수 있다.

· ROI

데이터 마트와 데이터 웨어하우징의 사용량에 관한 메타데이터가 없이 ROI를 결정하고 해당 프로젝트를 지탱하기는 힘들 것이다.

· 시스템의 개선과 향상 작업상의 난제

데이터 웨어하우징 개발자들은 현재 version의



〈그림 1〉 메타데이터의 각 구조간 관계

웨어하우스내의 내용을 모르고서는 해당 데이터 마트나 데이터 웨어하우스의 개선, 향상 작업 및 유지할 수 없을 것이다. 그러한 사실은 신속한 데이터 마트들에 대한 사용자의 요구를 충족시키고자 하는 현재의 추세 하에서 더욱 그렇다.

적절한 메타데이터 구조가 수반되지 않으면, 사용자들은 데이터 마트들로부터 모순되고 일정하지 않은 정보를 얻을 것이다.

메타데이터의 내용들

〈그림 1〉에서 제시된 독립 구조간의 관계를 나타내는 그림은 데이터 웨어하우스를 위한 메타데이터 저장부(貯藏部)의 내용들을 보여준다. 메타데이터에는 세가지 광의(廣義)의 범주가 있다.

1. 상용(商用) 이용자들을 위한 메타데이터

메타데이터는 AAA(American Automobile Association)에서 발행한 여행 안내서같이 어디에서 어떤 정보를 찾을 수 있고, 어떤 방식으로 접근할 수 있으며, 이런 정보에 접근하기 위해서 얼마나 시간이 걸리는지, 그리고 사용자가 해당 정보를 얻었을 때 어떤 품질의 정보를 기대할 수 있는지를 보여준다. 〈그림 1〉에서 'U' 자 표시된 독립구조들은 사용자들의 사업을 위해 아주 중요함을 나타낸다.

2. 데이터 웨어하우스 관리자들을 위한 메타데이터

데이터 웨어하우스의 효율성을 보장하고, 유지하며, 정착시킬 의무가 있는 데이터 웨어하우스 관리자는 그의 지루한 작업들을 자기 자신의 특별한 목

적에 부합하는 메타데이터를 이용하여 보다 간편하게 만들 수 있다.

그러한 메타데이터는 <그림 1>에서 'A'자로 표시된 것들에 추가하여 프로파일과 성장성 분석 매트릭스를 포함하고 있다.

3. 데이터 웨어하우징 개발자를 위한 메타데이터

개발자들을 위한 메타데이터는 데이터 웨어하우징 개발자들의 데이터 마트 유지 능력에 영향을 끼친다. 갱신된 메타데이터 없이 개발자들은 모순된 정보의 군락(群落)들로 발전하기 쉬운 데이터 마트들을 개선하고 유지하지 못할 것이다. <그림 1>에서 'D'자로 표기된 메타데이터의 구성체들은 이러한 개발자들에게 특별한 관심사일 것이다.

메타데이터를 위한 구조

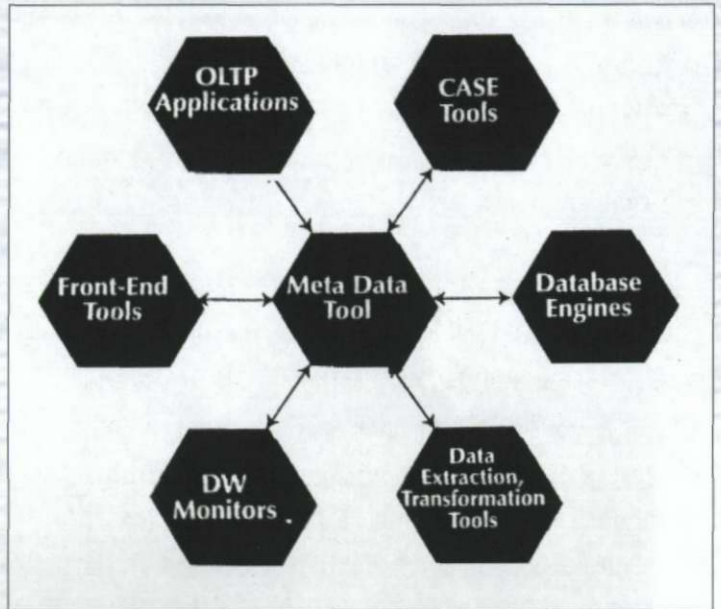
독자(讀者)가 단일한 데이터 마트를 만들거나 거대한 기업용 데이터 웨어하우징을 제작하든지 간에, 구조는 그것들에 대한 계획과 설계과정에 있어서 필수 불가결한 요소이다.

이와 같은 프로젝트 초기 과정에서 장기간의 구조 개발 과정은 그 과정을 통하여 데이터 웨어하우징 팀을 인도하여 미래를 위한 개발 시각을 결정하는데 도움을 준다. 메타데이터의 구조는 다음과 같은 성격을 지녀야 한다.

· 강제성

메타데이터는 OLTP 시스템에서조차도 중요 존재였다. 그러나 OLTP 시스템의 경우 대부분의 메타데이터는 프로그래머들과 분석가들로 구성된 IT 집단에게 요구되었다. 이러한 시스템상에서는 최종 사용자들이 대부분의 메타데이터를 필요로 하는 데이터 웨어하우징의 초창기부터 인지된 중요성을 결코 얻지 못했다.

만약 사용자들이 한 웨어하우스에서 어떠한 정보가 사용가능한지, 그러한 정보는 무엇을 의미하고 어떻게 이런 정보에 접근할 수 있는지를 모른다면 그들은 전혀 갈피를 잡지 못할 것이다. 데이터 웨어하우징 사용자들에게는 단지 상세한 메타데이터 뿐



<그림 2> 메타데이터와 메타데이터들의 경로관계

만 아니라 상세한 contextual 메타데이터 또한 제공되어야 한다.

앞에서도 설명했듯이, 상세하고 갱신된 contextual 메타데이터 정보가 없다면 사용자들은 잘못된 판단을 초래할 수 있는 애매하고 그릇된 정보를 얻을 수 있다. 그러므로 어느 데이터 웨어하우징 프로젝트를 위한 메타데이터의 구조는 추가적인 대상이 아니고, 전체 구조에서 필수 불가결하고 훌륭하게 설계된 부분이어야 한다.

· 개방성(다양한 범주의 틀들로부터 정보를 얻고 작업할 수 있는 능력)

메타데이터에 있어서 필수적인 조건은 정보를 항상 갱신된 상태로 유지하는 것이다. 만약 메타데이터 저장소가 독점적이고 다른 데이터 웨어하우징 구조의 구성 부분들과 정보를 교환하지 않는다면, 해당 메타데이터의 갱신 절차는 많은 시간을 소모할 것이며, 해당 메타데이터가 낡은 정보를 간직하는 결과를 초래할 것이다. <그림 2>가 보여주듯이, 메타데이터 틀은 메타데이터의 다양한 경로들과 문제없이 작동할 수 있어야 한다.

*케이스 틀들:

*애플리케이션들을 위한 내/외부로부터의

source 데이터:

- *연관적 혹은 다차원적 데이터베이스 엔진들:
- *데이터 추출, 변형, 정화 tool들:
- *DW 감시자:
- *Front-end tool들

기업들이 행한 두 가지 노력은 통일된 메타데이터 규격의 제정을 이끌어 냈다. 95년에 설립된 메타데이터 코알리션에 의한 그 첫번째 노력은 MIDS라고 불리는 메타데이터 교류규정(메타데이터 교환 양식)을 만들어 내었다. Viasoft's Rochade repository, The Intellidex Warehouse Control Center, ETI's ETI.EXTRACT Tool Suite 등의 몇몇 기업들이 MDIS 1.0을 준수하기로 발표하였다. MDIS를 준수하는 제품을 내놓기로 약속한 다른 기업들은 IBM, Apertus Carleton, Brio, Platinum, SAS 등이다.

Microsoft Corp.에 의한 두번째 노력은 객체 관리 그룹의 통합모델링 언어(Unified Modeling Language) 규격을 기반으로 한다. 개방 정보 모델(Open Information Model(OIM))이라고 불리는 The Microsoft Repository는 객체들의 설명에 관한 정보를 공유할 수 있게 하는 틀들을 위한 공통된 형식과 애플리케이션 개발의 주기를 통해 필요한 구성 부분들을 제공한다.

Microsoft는 또한 모든 형태의 데이터 웨어하우징 메타데이터를 지원하기 위해서 OIM을 확장하는 계획도 발표했다: 정보의 추출 및 변형 법칙, 데이터 설계 및 데이터 모델 정보 등을 포함하고 있다.

Microsoft는 OIM의 이러한 확장 문제를 이 회사의 데이터 웨어하우징 상대들과 협동으로 실행할 것을 계획하고 있다. 몇몇 기업들은 OIM의 확장부의 제작을 지지한

다는 뜻을 발표한 상태이다.

· 정보의 동조와 갱신의 용이함

만약 메타데이터 관리자가 메타데이터를 갱신하는데 아주 많은 단계를 거치기를 요구하면, 이것은 오류가 발생할 확실한 방법이다. 메타데이터의 구조는 가능한 어떤 위치의 메타데이터 source를 가지고서도 정보의 자동적 갱신을 촉진해야 한다. 메타데이터의 동조와 갱신을 위한 명확한 기준 절차가 있어야 하는 것이다. 메타데이터를 갱신하고 동조시키는 절차에는 세가지 범주가 있어야 한다.

▷ 정보의 초기 획득. 이 절차는 다양한 범주의 틀들과 논리적으로 중앙에 위치한 메타데이터 저장소간의 메타데이터 교환을 위해 설계되어야 한다. 정보의 초기 획득은 일반적으로 프로젝트의 개발 주기 동안 발생한다. 이러한 것은 메타데이터 저장부의 계속적인 갱신작업을 위해 필요한 절차를 개발하는데 이용되어야 한다.

▷ 계속적인 갱신 작업. 데이터 웨어하우징이 작동을 시작하고 나면, 해당 메타데이터의 계속적인 갱신 작업을 위해 자동화된 처리과정을 개발해야 한다.

▷ 일관성 검사. 이 과정은 본 기사의 초반부에 제시된 다양한 정보원들의 갱신된 날짜와 시간에 기초하여 해당 메타데이터가 일관적이고 정확한지를 확인하기 위해 이용되어야 한다.

만약 누군가 수동으로 메타데이터를 갱신해야 할 경우, 해당 과정이 그 메타데이터를 고수하는지를 확인하기 위한 충분한 검사와 동기가 있어야 한다. 규정을 준수하는 일 이외에 메타데이터의 일관성과 완성도를 검사하는 주기적인 절차가 필요하다. <그림 3>은 데이터 웨어하우징

프로젝트의 여러 단계의 기간 동안 중요한 메타데이터 source들을 보여 준다.

· 확장성

메타데이터 구조가 손쉽게 확장될 수 있어야 함은 대단히 중요하며, 특히 점점 더 일반화 되고있는 "포괄적으로 사고하며 지엽적으로 행동한다"

	Business Analysis	Data Analysis & Mapping	Architecture Design	Development & Implementation	Maintenance & Support
CASE	✓	✓	✓		✓
Database Engines				✓	✓
Data Extraction, Cleansing & Transformation		✓	✓	✓	✓
Front-End Tools				✓	
Data Warehouse Administration Tools					✓
Legacy Source Applications	✓	✓			✓

<그림 3> 데이터웨어하우징 프로젝트의 기간별 메타데이터 소스

는 방식의 데이터 마트 접근법에 있어서 더욱 그러하다. 더 많은 데이터 마트들이 추가됨에 따라, 메타데이터의 구조는 분산 형식이든 중앙집중 방식이든 간에 손쉽게 확장될 수 있어야 한다.

· 장래성

메타데이터는 웹 기술, 데이터 웨어하우징 등의 결합과 함께 보다 더 중요한 위치를 차지할 것이다.

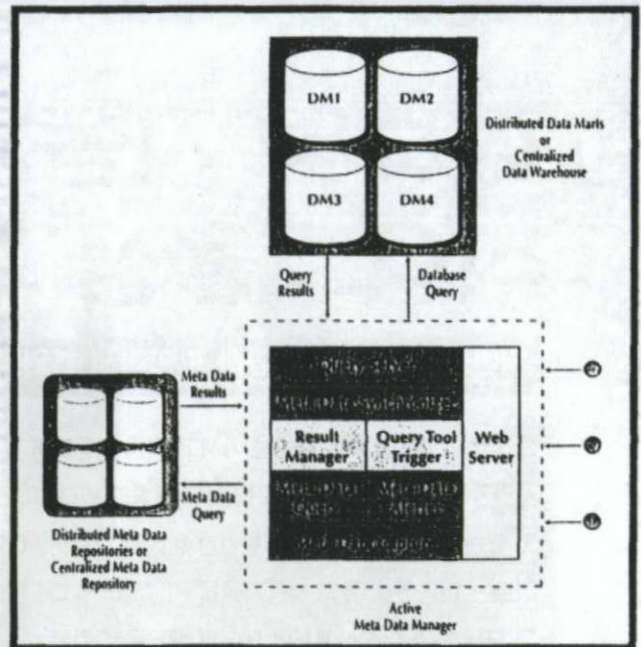
이러한 결합은 인터넷이나 조작 시스템상 혹은, 데이터 웨어하우징 환경에서건 사업정보로의 접근을 위한 단일 창구 역할을 할 메타데이터 브라우저의 출현을 야기할 것이다. 메타데이터는 어떤 종류의 구조하에서도 중요한 구성 부분으로 자리잡을 것이다. <그림 4>의 도형은 미래의 메타데이터 구조에 관한 하나의 논리적 설명이다.

논리적 메타데이터 저장부는 사업상의 필요성이나 조직의 요구에 따라 분산 또는 집중될 수 있다. 이것은 객체들과 관련 데이터들로 구성된 데이터 스토어들의 모임과 흡사하다.

활동중인 메타데이터 관리자는 메타데이터 구조의 핵심 구성 부분이며, 다음과 같은 부분으로 구성될 때 이상적이다.

- 메타데이터 획득 : 다양한 source들로부터 메타데이터의 초기 획득 과정.
- 메타데이터 동기화 : 메타데이터를 갱신하기 위한 과정들.
- 메타데이터 검색 엔진 : 이것은 사용자들이 메타데이터의 검색과 접근을 위한 사용자 환경이 될 것이다.
- 메타데이터 결과 관리자 : 이 부분은 메타데이터 검색으로부터 얻은 결과들을 처리하고 사용자가 적합한 선택을 하는 것을 가능하게 한다.
- 메타데이터 경고자 : 이것은 푸시기술의 한 분야로, 사용자 프로파일에 의거하여 해당하는 가입자들에게 메타데이터의 내용에 새로운 변화를 통보할 것이다.
- 메타데이터 질의 유도자 : 이 부분은 메타데이터 관리자에서 사용자가 선택한 내용을 바탕으로 질의 틀이 데이터 웨어하우징이나 다른 source로부터 데이터를 구할 수 있게 유도한다.

그러한 구조는 사용자들이 최종적으로 그들이 절실히 필요로 하는 정보를 접했을 때 사용자들이 무엇을 얻을 것



<그림 4> 미래의 메타데이터 구조

인지를 알 수 있도록 보장한다. 데이터 웨어하우징 시장이 신속히 발달함에 따라 이러한 구조가 현실화되는 것은 더욱 앞당겨질 것이다.

최근에 the Meta Data Coalition은 앞서 설명한 the Meta Data Coalition의 메타데이터 상호 교환 규정 (MDIS)과 Microsoft의 OIM 기반의 저장소를 이어주기 위한 프로그램의 시험버전을 제작했다.

이 프로그램은 MDIS를 방식의 파일들과 Microsoft의 자료 저장부(貯藏部)간의 쌍방향 정보 교환을 가능하게 할 것이다. 이것은 메타데이터의 동적인 갱신에 바탕을 둔 이상적 메타데이터 구조의 완성을 향한 커다란 첫 걸음인 것이다. 