메타데이터 기반 데이터 통합 관리 동향에 관한 연구*

강양석** · 홍순구*** · 이영상**** · 허진석****

A Study on Metadata-Driven Data Integration*

Yang-Suk Kang** · Soon Goo Hong*** · Young-Sang Lee**** · Jin-Suk Heo*****

■ Abstract ■

It is essential for companies to manage massive data for dealing with large volume of transactions and customers' needs. To this end, the companies have operated data warehouse with many complex tools for data gathering and reporting to the end-users. However, the data from the heterogeneous tools at the various sources cannot be exchanged because of the different interfaces. Therefore, the data cannot be controlled with integrated manner, and furthermore the companies do not focus the quality of data resulting in the data quality problem. Thus, this study suggests how to manage massive data with a metadata. In particular, we investigate current status of metadata management, its appliance, and perspectives. The contribution of this research is to apply the metadata management system to the real world and to suggest its management procedure.

Keyword : Metadata, Data Quality, Data Warehouse, Interoperability

논문투고일 : 2008년 09월 13일 논문수정완료일 : 2008년 12월 06일 논문게재확정일 : 2008년 12월 08일
* 이 논문은 동아대학교 학술연구비 지원에 의하여 연구되었음.
** 동아대학교 경영정보과학부
*** 동아대학교 경영정보과학부, 교신저자
**** 이노베이티브데이터솔루션즈(주)

1. 서 론

글로벌 경쟁과 복잡한 의사 결정의 필요성으로 인해 데이터 관리는 기업의 성패를 가름하는 열쇠가 되고 있다. 양질의 데이터를 통한 기업의 전략적 의사결정과 시장 환경 변화에 대한 유연한 대처는 기업의 경영 이익을 극대화 할 뿐만 아니라 위기 상황 속에서 기업이 현명하게 대처할 수 있도록 돕는다. 그러나 나날이 증가하는 데이터의 양과 복잡해지는 사용자들의 요구 사항으로 인해 데이터 관리의 필요성이 대두되고 있다.

이러한 양질의 데이터를 전사적인 관점에서 관리하기 위해 메타데이터를 기반으로한 데이터 관리 방안이 연구되고 있다. 메타데이터는 데이터에 대한 데이터로 전체의 데이터에 대한 요약된 정보를 제공하며, 데이터 표준화 이후의 메타데이터는 데이터에 대한 생성 규칙, 오너쉽(ownership), 사용 목적 등에 대한 정보를 제공한다. 사용자는 이를 참조하여 새로운 시스템을 개발하거나 현재의 데이터가 어떠한 용도로 생성되었는지 등 데이터에 관한 상세한 정보를 얻을 수 있다. 프로세스적인 측면에서는 전사적인 논의를 거친 데이터 표준화 작업과 이를 반영한 메타데이터의 참조를 통해서만 데이터의 생성 및 사용을 허용함으로써 데이터 품질 관리 문제를 해결할 수 있게 된다.

본 논문에서는 메타데이터의 의미 변화를 통시적인 관점에서 파악하고 기업의 데이터 관리를 위한 메타데이터 관리 방안을 제시 한다. 이를 위한 연구 방법으로 국내외 문헌 조사 하고 업체에 적용된 구현 사례를 살펴 본다.

2. 선행 연구

2.1 메타데이터의 정의 및 의미 변화

메타데이터의 일반적인 정의는 데이터에 대한 데이터를 뜻한다. 데이터 관리 환경이 변함에 따라 많은 연구자들이 메타데이터를 다양한 관점에서 정의를 내리고 있다(<표 1> 참조)[1, 2, 4, 5, 8, 10, 11].

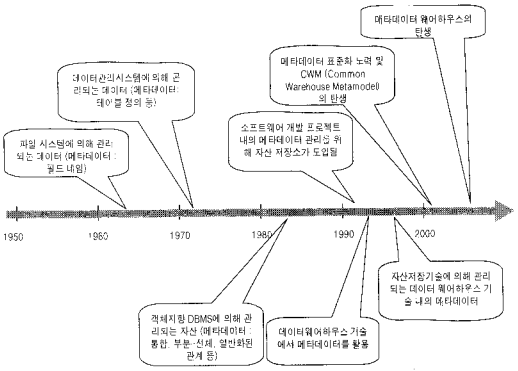
<표 1>의 Kal(2001)의 정의에서 메타데이터를 한정 짓는 일이 어려움을 알 수 있다. 본 연구에서의 메타데이터는 전사적인 관점에서 기업의 경영활동을 지원하기 위한 정보시스템에서 발생되는 데이터에 대한 데이터로 정의한다.

Sen(2002)에 의한 메타데이터의 역할 변화는 [그림 1]과 같다. 파일 시스템에서의 메타데이터는 해당 파일이 어떤 내용을 담고 있는지를 요약하여

〈표 1〉 연구자에 따른 메타데이터의 정의

연 구 자	설 명
Inmon(2003)	데이터를 설명하는 데이터
Kal(2001)	데이터에 대한 기술적 데이터이므로 정의에 근거하여 거의 모든 종류의 데이터를 메타데이터라고 정의 할 수 있음
오삼균(2002)	메타데이터들은 유형에 관계없이 주어진 자원을 파악할 수 있는 다중관점을 제공
남영광(2005)	데이터에 관한 데이터로써 정보자원의 속성을 기술하는 데이터
Eckerson(2004)	일관성 있는 참조(consistent reference)를 위한 데이터이며 기록에 대한 핵심적인 집합(core set)을 의미
ANSI의 NISO(2004)	정보 자원의 기술, 설명, 위치 또는 갱신, 사용 또는 관리를 용이하게 하기 위한 구조화된 정보
Sen(2002)	메타데이터의 관리는 전통적인 데이터베이스와 데이터웨어하우스 관리 문제에 있어서 부차적인 것이었지만 오늘날의 기업들은 도구 및 데이터의 통합과 변화 관리(change management)를 위해 메타데이터를 필요로 함

간략하게 보여준다. 1970년대 초의 데이터베이스 관리시스템에서 메타데이터는 테이블에 대한 정의 등을 통해 해당 테이블이 어떠한 목적으로 작성되었는지에 관한 정보를 제공하였다. 1980년대 초에는 데이터베이스 내의 통합 및 규칙 등에 메타데이터가 활용되었다. 1990년대 초에는 소프트웨어 개발에 메타데이터가 사용되기 시작하여 1990년대 중반에는 데이터웨어하우스 내에서 도구들 간의 통합을 위해 사용되었다. 2000년대 초에는 메타데이터의 사용 영역이 확장됨에 따라 각각의 영역 간의 메타데이터들의 통합 문제가 대두되었고, 이를 해결하고자 기업과 기업 간의 메타데이터 교환을 위한 CWM(Common Warehouse Metamodel) 등의 상호 운용성에 대한 표준화 노력이 대두되었다. Sen(2004)이 제시한 [그림 1]에서는 최초 원천 데이터로부터의 부가적인 정보 수준의 메타데이터의 쓰임이 2000년대에 들어서는 메타데이터 웨어하우스라는 명칭이 등장하기까지에 이르러, 메타데이터의 활용 범위와 메타데이터 자체의 복잡도가 비약적으로 증가하였음을 알 수 있다.



출처 : Sen, Metadata Management : Past, Present and Future, p. 154, 2004.

[그림 1] 메타데이터의 과거, 현재 및 미래

2.2 데이터 관리시스템의 변화

데이터 관리시스템은 기존 시스템의 문제점들을 개선하기 위해 파일처리 시스템에서 데이터베이스 시스템을 거쳐 데이터웨어하우스로 변화 하였다[3]. 과거에 비해 오늘날은 관리 대상이 되는 데이터의 양과 관리의 복잡도가 증가하였으며, 이를 해결하기 위해 다양한 도구들이 데이터 관리에 사용된다[7]. 이러한 각각의 도구들을 통합 관리 할 수 있는 방안이 메타데이터를 통한 데이터 통합 관리이다(<표 2> 참조).

2.3 메타데이터와 관련된 데이터 품질문제

메타데이터와 관련된 기업 내 데이터 품질 문제는 <표 3>과 같다. 이러한 데이터 품질문제는 메

〈표 2〉 데이터 관리시스템의 변화

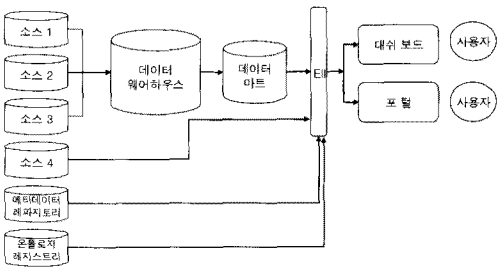
	시스템 명	내 용	특 징
과거 ↓ 현재	파일처리 시스템	개별 응용프로그램에 종속적인 데이터 파일이 존재	-
	데이터베이스 시스템	조직내 필요한 데이터를 통합하여 여러 사용자들이 공동으로 사용하는 시스템	• 데이터 통합 • 데이터의 중복 및 불일치 개선 • 응용프로그램 독립적
	데이터웨어하우스	경영 및 정책 결정에 필요한 정보 처리 기능을 효율적으로 지원하는 전사적 데이터베이스	• 의사결정 업무 지원 • 주제 중심적이며 시계열 자료 포함
	메타데이터 기반 통합 관리시스템	다양한 도구들 간의 커뮤니케이션과 서로 다른 형태 및 언어로 기술된 데이터들의 통합을 위한 메타데이터의 활용	• ETL 도구, Reporting 도구 등 데이터 웨어하우스 내의 서로 다른 도구에 대한 이해 및 데이터의 통합 관리 지원

타데이터를 통해 해결가능하기 때문에 기업에서 메타데이터 관리시스템의 도입 및 운영이 필요하다.

3. 메타데이터 기반 데이터 관리 방안 및 적용

3.1 메타데이터 기반의 데이터 관리 방안

원천 데이터의 증가와 함께 관리의 필요성이 있는 메타데이터의 양과 복잡도 또한 증가함에 따라 이에 대한 관리 방안이 필요하게 되었다. 메타데이터는 종류 및 쓰임새가 다양하며 Mehra(2005)는 메타데이터의 종류를 문법적 메타데이터(Syntactic metadata), 구조적 메타데이터(Structural metadata) 및 의미적 메타데이터(Semantic metadata)로 구분하고 온톨로지를 포함한 메타데이터 아키텍처를 제시하였다[9]. Mehra의 구분은 메타데이터가 포함하는 정보가 구조, 배열 및 의미 중 어떤 것과 연관되어 있는지에 따른 구분이다. Mehra에 의한 물리적인 메타데이터 레파지토리의 구조는 [그림 2]와 같다.



출처 : Mehra, V., Building a Metadata-Driven Enterprise; A Holistic Approach, 2005.

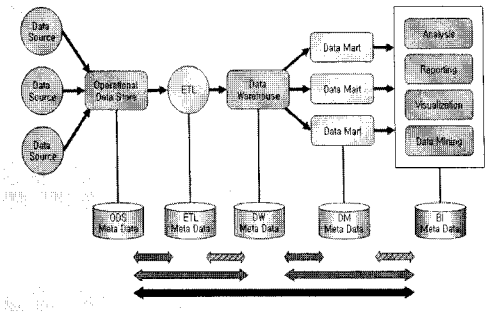
[그림 2] Mehra에 의한 메타데이터 통합 아키텍처

〈표 3〉 메타데이터와 관련된 기업 내 데이터 품질 문제

적용 분야	As-Is	문제점	To-Be
데이터 표준화	• 표준화 지침은 개발자에 의해 임시적 · 임의적으로 적용 • 표준화 관리 시스템의 부재	• 표준화 되지 않은 데이터로 인한 정합성 문제 • 자료의 신뢰성 문제	• 전사적인 데이터 표준화 지침 수립 • 메타데이터 관리시스템 구축을 통한 데이터 관리의 기본 틀 마련
모델링 관리	• 관리의 부재 또는 부분적 관리 • 개발 프로젝트의 종료 후 자료 갱신의 어려움 • 자동화 관리 시스템의 부재	• 부분 관리로 인한 통합 차원의 데이터에 대한 이해 부족	• 데이터 표준화 관리시스템과 연동하여 모델링 관리 도구 구축
데이터 품질 관리	• 단발적인 품질 관리 정책의 수행 • 데이터 품질관리 자동화 시스템의 부재	• 품질 관리 수행시 품질관리 검증 프로그램으로 인한 시스템 부하 발생	• 전사적 차원의 데이터 품질관리에 대한 인식 및 실행 필요
관리 조직	• DA(Data Administrator)의 부재 • 데이터 품질관리위원회의 부재	• 관리 조직의 부재	• 관리 조직에 대한 투자 및 전문가 양성
관리 프로세스	• 데이터 관리 프로세스 부재 • 개발계→ 테스트계→ 운영계로의 데이터 이관 프로세스가 명확하지 않음 • 자동화 관리 시스템의 부재	• 데이터에 대한 책임 소재를 알 수 없어 문제 발생시 대응하기 어려움	• 워크플로우(workflow) 체계 수립 및 강제성을 띤 시스템화 필요
통합 데이터 관리	• 통합 데이터 관리에 대한 전략 부재로 인해 개별 시스템 필요시 독립적으로 도입	• 모델 관리, 메타관리 및 품질 관리 등 분산 관리를 통해 통합 관리의 필요성	• 데이터 표준화, 데이터흐름, 데이터 품질, 영향도 등의 측면에서 통합적으로 데이터에 대해 모니터링 할 수 있는 통합데이터 관리시스템 구축이 필요

[그림 2]에서 EII(Enterprise Information Integration)은 기업내 정보를 통합하기 위한 도구를 의미한다. 통합 도구는 통합 도구를 위한 별도의 데이터 저장소인 소스 4를 가지고 있다. [그림 2]의 좌측 하단에 메타데이터 레파지토리는 사내 데이터로부터 발생하는 메타데이터를 저장하는 장소이다. 레파지토리 하단의 온톨로지 레지스트리는 메타데이터를 위한 표준화 절차의 결과인 용어 표준화와 관련된 정보를 담고 있다.

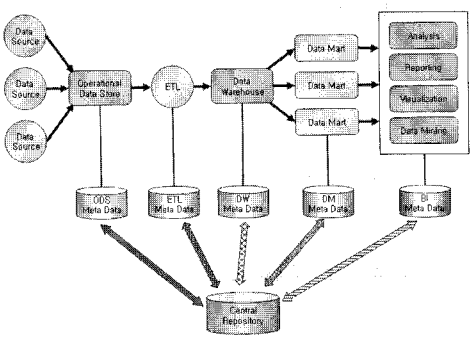
국내의 메타데이터 기반 솔루션 보유 업체에 따른 메타데이터 관리 방법은 브릿지 방식, 중앙집중식 방식의 2가지 관리 방법을 예로 들 수 있다. 먼저 [그림 3]은 브릿지 방식의 메타데이터 관리 아키텍처를 표현한다[7].



[그림 3] 브릿지 방식의 메타데이터 관리 아키텍처

브릿지 방식의 메타데이터 관리 아키텍처는 데이터웨어하우스 내의 데이터 발생 원천으로부터 개별적으로 독립된 메타데이터 저장소를 보유한다. 저장된 메타데이터는 전사적인 활용을 위해 상호 교환이 이루어져야 하는데 이를 위해 [그림 3]에서 표현된 조정 인터페이스에 따라 각각의 저장소의 메타데이터 교환이 이루어진다. 브릿지를 통한 교환 방식은 각각의 ODS, ETL 등의 메타데이터 저장소가 서로 다른 저장소와의 통신을 위해 모든 인터페이스를 이해하여야 한다는 단점이 있다.

[그림 4]는 중앙집중식 메타데이터 레파지토리의 구조를 표현한다. 기업 내에서 발생되는 메타데이터는 각각의 메타데이터 저장소를 거쳐 중앙의 메타데이터 레파지토리에 저장된다. 이 방식의 경우 각각의 메타데이터 저장소는 서로 다른 메타데이터 저장소와의 통신을 위한 별도의 인터페이스를 필요로 하지 않으며, 중앙의 Central Repository와 통신 할 수 있는 인터페이스만을 필요로 한다. 한편 중앙집중식 레파지토리의 방식에서도 역시 이기종 도구 간의 이해를 위해 각각의 데이터 포맷 및 교환 방식 등을 분석할 필요성이 있다.

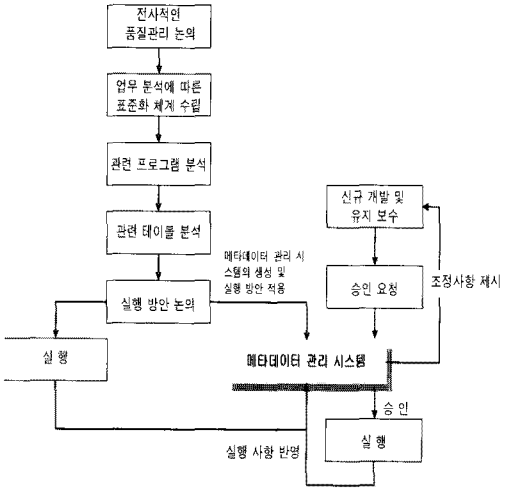


[그림 4] 중앙집중식 메타데이터 레파지토리

3.2 메타데이터 적용 절차

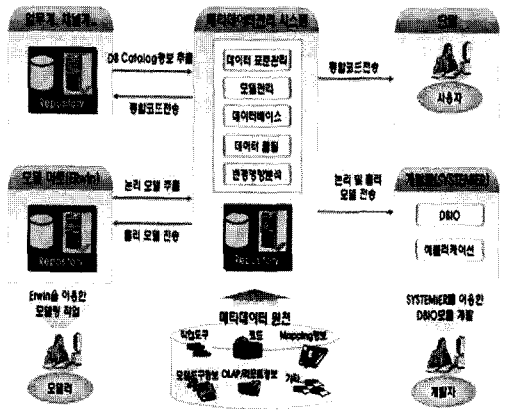
메타데이터의 적용 절차는 [그림 5]와 같다.

먼저 기업은 메타데이터의 적용을 위해 전사적인 품질관리 논의를 거쳐 데이터 표준화 과정을 수행한다. 표준화 이후 관련 프로그램 및 테이블을 분석하고 이에 대한 실행 방안을 논의한다. 실행 방안은 기존의 DBMS를 이용한 조정 및 제약사항들을 표준화 논의 결과에 맞게 등재하는 것과 새로운 메타데이터 레파지토리를 생성하여 새로운 규칙을 적용하는 것 등으로 구분된다. 실행 방안에 따라 생성된 메타데이터 관리시스템은 해당 데이터와 관련된 포괄적인 정보를 담고 있으며, 신규 데이터의 생성 및 소멸에 메타데이터 관리 시스템이 관여하게 된다. 생성된 메타데이터 관리시스템은 전사적인 것이며 새로운 규칙과 기존의 메타데이터를 모두 이해할 수 있어야 한다. 이는 데이터 관리를 위한 새로운 논의 사항과 기존의 데



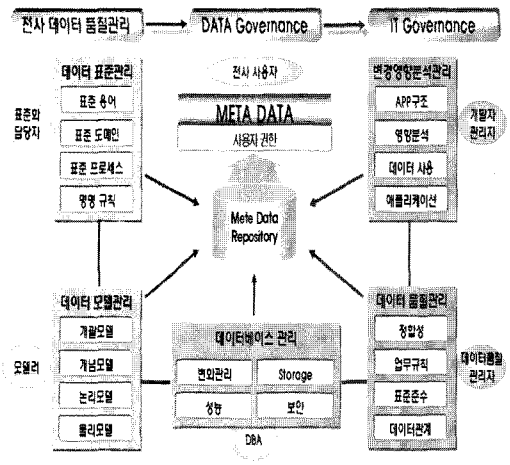
[그림 5] 데이터 품질관리를 위한 메타데이터의 적용 절차

이터 관리시스템과의 연결을 통해 가능하며, 기업의 적용 사례에 따른 데이터 관리시스템과 메타데이터 관리시스템의 연관도는 [그림 6]과 같다[10]. [그림 6]에서 관리시스템은 업무계 및 채널계로부터 데이터베이스 카탈로그 정보 등을 추출한다. ERWin 등의 모델 도구에서는 논리 모델을 추출하며 이러한 기존의 메타 정보들은 표준화 과정을 거친 관리시스템 메타 정보와 결합되어 데이터 관리를 위한 새로운 규칙에 반영된다. 관리시스템은 전사적인 데이터를 이해할 수 있으며, 관리시스템



[그림 6] 메타데이터 관리시스템과 기존 시스템과의 연관도

의 규칙에 따라 데이터의 생성 및 배포가 허용되므로 메타데이터 관리시스템은 데이터의 품질 관리를 위한 기초적인 틀이 된다. 또한 관리시스템은 사내 데이터에 대한 이해를 바탕으로 변경 데이터의 발생시 해당 데이터가 기존의 데이터들에 대해 어느정도 영향을 미치는지 등에 대한 변경영향분석을 수행할 수 있다. 메타데이터 관리 시스템에서는 데이터 품질 관리를 위한 다양한 담당자가 관여하며, 이러한 관리 영역 및 책임 소재는 [그림 7]과 같다[10].



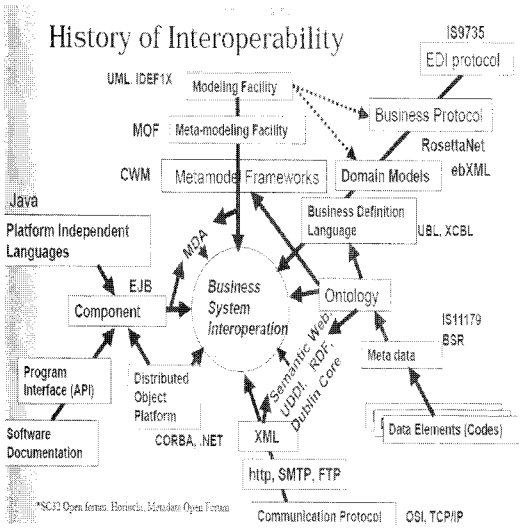
[그림 7] 메타데이터 레파지토리를 기준으로 나타낸 메타데이터 관리 시스템의 구성도

각각의 영역에서의 데이터 관리 담당자는 전사적으로 합의된 데이터 품질 관련 규칙을 메타데이터 저장소로 전송하여 데이터에 대한 통합 관리를 유도하고, 사용자는 메타 시스템으로부터 부여 받은 권한 수준에 따라 이러한 정보를 활용한다[10]. 특히 [그림 7]의 상단 우측의 변경영향분석관리란 특정 메타데이터의 변경시 대상 데이터가 전체 시스템에 미치는 파급 정도를 파악하여 이를 관리하는 것이다.

3.3 기업과 기업 간의 메타데이터 교환

한편 기업 내 메타데이터 관리를 넘어선 업계와

업계 간의 메타데이터의 교환을 위한 연구가 필요하다. 업계 간의 메타데이터 교환은 특정 연구자 및 기업들에 의해 연구되고 있으며 CWM(Common Warehouse Metamodel) 등과 같은 기업과 기업 간의 표준이 이러한 예가 될 수 있다. 메타데이터의 상호 운용성의 역사는 [그림 8]과 같다[6].



출처 : Horiuchi, H., ISO/IEC 19763, Framework for metamodel interoperability, 2006.

[그림 8] 메타데이터의 상호 운용성(interoperability)의 역사

[그림 8]에서 최초의 메타데이터들은 모델링 도구의 사용, 소프트웨어 개발에 있어서의 문서화 등 각각의 업무에 대해서만 편의성과 재사용성을 높일 수 있는 방법들로 고안되었다. 이러한 개념들은 확장과 정립을 통해 분야별로 메타모델 프레임워크, 컴포넌트, 온톨로지 등의 개념에 이르렀는데 [그림 8]에서와 같이 최종적인 개념에서는 각각의 메타데이터에 대한 개념이 서로 간의 영역에 간섭하여 사용되고 있음을 알 수 있다.

[그림 8]에서 시사하는 바는 분야별 메타데이터의 활용도가 증가하였다는 점과 궁극적으로는 메타데이터의 상호 교환을 실현할 수 있는 표준이 마련되어야 한다는 것이다.

4. 결 론

본 연구에서는 관련 선행연구와 메타데이터 관리시스템을 운영하는 IT서비스 업체의 관리 방안에 대한 사례 연구를 통해 전사적인 데이터 통합 관리 동향에 대해 살펴보았고, 이를 위한 관리 절차를 제시하였다. 선행 연구를 통해 메타데이터는 기업내 데이터의 통합 관리 및 품질 향상을 위한 방안으로 유용하다는 것과 메타데이터의 활용 및 이의 중요성이 과거에 비해 매우 높아졌음을 알 수 있었다. 연구의 시사점은 메타데이터의 관리 현황과 동향을 파악함으로써 메타데이터에 대한 중요성을 인식시키고 향후 기업의 데이터 품질 관리를 위한 고려 사항과 이러한 메타데이터 관리를 위한 실행 절차를 제시하였다는 것이다. 향후에는 기업의 메타데이터 적용사례와 기업간 데이터 통합관리를 위한 표준에 대한 심도 깊은 연구가 필요하다.

참 고 문 헌

[1] 남영광, "ISO/IEC 11179에 따른 산업기술정보 메타데이터 표준화", 「정보관리연구」, 제36권, 제1호(2005), pp.57-75.

[2] 오삼균, "메타데이터의 국제적인 적용사례와 호환성에 관한 연구", 정보통신부, 2002.

[3] 최덕원, 이동철, 「경영정보 데이터베이스 시스템」, 2000.

[4] Ahmed, K., D. Ayers, M. Birbeck, and J. Cousins, *Professional XML Metadata*, Wrox Press, 2차 인용 자료, 2001.

[5] Eckerson, W. W., "Mastering Metadata", *Business Intelligence Journal*, Vol.9, No.4 (2004).

[6] Horiuchi, H., ISO/IEC 19763, Framework for Metamodel Interoperability, 2006.

[7] IDS, 내부자료, 2006.

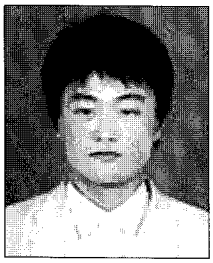
[8] Inmon, W. H., The Importance of Meta Data, www.billinmon.com, 2003.

[9] Mehra, V., "Building a Metadata-Driven Enterprise; A Holistic Approach", *TDWI*, 2005.

[10] NISO, *Understanding Metatadata*, 2004.

[11] Sen, A., "Metadata Management : Past, Present and Future", *Decision Support Systems*, Vol.37, No.1(2004), pp.151-173.

◈ 저 자 소 개 ◈

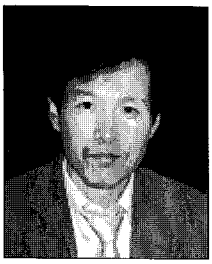


강 양 석(kangyangsuk@naver.com)

동아대학교 경영정보과학부를 졸업하고 동대학원 석사 학위를 취득하였다. 메타데이터, 웹 접근성, RFID 미들웨어 개발 프로젝트에 참여하였고, 주요 관심분야는 Data Governance, RFID/USN, Green IT 등이다.

홍 순 구 (shong@dau.ac.kr)

영남대학교에서 경영학사, 미국 네브라스카 주립대학교에서 경영학 석사 및 박사 학위를 취득하고, 현재 동아대학교 경영정보과학부의 부교수로 재직 중이다. 주요 관심분야는 Data Warehousing, IS Evaluation, ERP, Web Accessibility 등이다.



이 영 상 (yslee@datastreams.co.kr)

경북대학교에서 전자공학 학사, 미국 미시간 주립대학교에서 전자공학 석사, 한국과학기술원(KAIST)에서 전자공학과 박사 과정을 수료하고, 현재 데이터처리전문기업인 (주)데이터스트림즈의 대표이사로 재직 중이다. 주요 관심분야는 ETL기반의 real-time data processing, Metadata기반의 데이터 통합 관리, Data Warehousing 등이다.

허 진 석 (jshuh@ids.co.kr)

수원대학교에서 전자계산학 학사를 취득하고, 다양한 산업분야의 DW/RM 프로젝트에 참여 했으며, 현재 데이터처리 전문기업인 (주)데이터스트림즈에 재직 중이다. 주요 관심분야는 Metadata기반의 데이터 통합관리, Data Warehousing, CRM 등이다.