

## 데이터 통합을 위한 XMDR 기반의 데이터 허브 시스템

문석재<sup>○</sup> 엄영현 국윤규 정계동 최영근  
광운대학교 컴퓨터 과학과

{msj1568<sup>○</sup>, class76, ykkook, gdchung, ygchoi}@kw.ac.kr

### Data Hub System based XMDR for Data Integration

SeokJae Moon<sup>○</sup> Y.H. Eum Y.G. Kooj G.D. Jung Y.G. Choi  
Department of Computer Science, KwangWoon University

#### 요 약

데이터 통합은 기업의 각 조직과 주요 업무, 핵심 애플리케이션에서 발생하는 물리적인 데이터 소스들을 표준 규칙과 메타데이터에 여과시켜 중복성을 제거하고, 오직 데이터 통합 및 단일 뷰를 정확하게 제공하기에 어려움이 따른다. 특히, 이기종 시스템이나 다양한 애플리케이션에서 나오는 대량의 데이터를 종류와 형식에 관계없이 호환이 가능하도록 지속적으로 통합하며, 정확한 정보를 실시간으로 동기화하여 제공할 수 있는 자동화된 정보의 통합이 관건이다. 따라서 본 논문에서는 레거시 시스템간의 데이터를 협업할 때 실시간으로 변화는 데이터를 일관성 있게 유지하기 위해서 데이터 협업 메커니즘을 제안한다. 또한 XMDR을 이용하여 협업에 의한 데이터 통합에서 발생하는 의미적 상호 운용성의 문제점을 해결하는 XMDR 기반의 데이터 허브 시스템을 구축한다.

#### 1. 서 론

대부분 기업 환경의 정보 시스템들은 지역적으로 분산되어 있고, 이기종 데이터 소스들로 구성되어 있다. 데이터 통합은 기업의 각 조직과 주요 업무, 핵심 애플리케이션에서 발생하는 물리적인 데이터 소스들을 표준 규칙과 메타데이터에 여과시켜 중복성을 제거하고, 오직 단일의 데이터 및 단일 뷰를 정확하게 제공하는 데에서 출발한다[2]. 특히, 이기종 시스템이나 다양한 애플리케이션에서 나오는 대량의 데이터를 종류와 형식에 관계없이 호환이 가능하도록 지속적으로 통합하며, 정확한 정보를 실시간으로 동기화하여 제공할 수 있는 자동화된 정보의 통합이 관건이다[1][6]. 하지만 많은 정보를 모으고 처리해야 하는 협업 환경에서 요구되어지는 다른 응용 시스템과의 연계과정에서 다른 시스템 혹은 파트너에게 정확성이 검증된 정보를 정확하게 공유하는데 어려움이 따른다. 또한 사용자가 요구하는 통합된 정보를 제공하기 위해 사용자에게 단일 뷰를 제공하기에도 어려움이 따른다. 따라서 이러한 문제를 해결하기 위해서는 레거시 시스템들 간에 구축된 데이터베이스들은 각각 독립성을 유지하면서 하나의 인터페이스처럼 투명성을 제공해야 한다.[3] 그러나 정보 검색 및 레거시 시스템간의 데이터 통합에 있어서 의미적 상호 운용성이라는 문제가 대두된다. 본 논문에서는 기존 레거시 시스템을 유지하면서 정보 공유 및 통합 검색을 할 수 있는 데이터 허브 시스템을 ISO/IEC 11179에서 제안한 XMDR의 개념을 이용하여 구축한다. 이 시스템은 크게 5가지 범주의 서

비스 영역으로 분리된다. 첫 번째, 정보 접근의 적시성을 제공하여 기업 내·외부에 흩어져 있는 서로 다른 시스템들의 하나의 인터페이스를 통해서 정보를 공유할 수 있는 환경을 제공한다. 두 번째, 데이터 허브의 핵심 역할 중 하나로 단일 뷰를 제공하여 기업 정보의 전사 컨센서스를 확보한다. 다시 말해, 동일한 의미와 동일한 표현 형식으로 제공된다는 것이다. 세 번째, 데이터의 실시간 양방향성, 실시간 양방향성은 데이터 허브의 허브가 의미하는 것처럼, 허브 앤 스포크라는 정보 흐름의 양방향성을 제공한다. 네 번째, 체계화된 규칙 기반의 데이터 정제 및 관리를 제공한다. 정보의 전사 공유 환경을 제공하기 위해서는 체계화되고 자동화로 관리되어야 한다. 마지막으로 외부 정보의 활용으로 급변하는 시장 환경에서 유연성을 갖춘 데이터 허브 시스템을 구축한다. 따라서 사용자에게 유지하면서 하나의 인터페이스처럼 투명성을 제공하는 데이터 허브 시스템을 제안한다. 본 논문의 구성은 다음과 같다. 제2장은 관련연구를 제3장에서는 데이터 통합을 위한 XMDR 기반의 데이터 허브 시스템을 기술한다. 제4장에서는 XMDR 모델 구성에 대해 기술한다. 제5장에서는 시스템의 적용과 타 시스템과의 비교 분석에 대해 기술하고 마지막으로 제6장에서 결론 및 향후 연구에 대해서 기술한다.

#### 2. 관련연구

최근 수십 년간 기업 업무의 효율성 재고와 자동화를 위한 방안으로 다양한 IT 인프라를 구축해왔다. 시스템의 도입 형태를 살펴보면, 단위 업무별로 필요할 때마다

일부 기능이나 산업에 특화된 Best of Breed 형태의 포인트 솔루션을 전략적으로 선택하여 구현해왔다[6]. 그러나 이러한 고급 기능들은 사용자의 생산성에 더 많은 비용을 발생시켰으며, 결과적으로 정보의 Silo 현상을 초래하여 시스템간 데이터의 장벽과 고립이 심화되었다. 사용자가 고객 데이터를 획득할 때 전체가 아닌 일부 데이터만을 액세스할 수 있고, 동일한 고객 데이터가 여러 가지 형태로 표현되는 문제를 안고 있다는 것이다. 종합적이고 신뢰할 수 있는 정확한 고객 정보를 제공할 수 없다. 따라서 효과적인 실시간 의사결정은 기대할 수조차 없는 실정이다.

### 3. 데이터 통합을 위한 XMDR 기반의 데이터 허브 시스템

#### 3.1 시스템 개요

본 논문에서 제안한 XMDR 기반의 데이터 허브 시스템의 전체 시스템 개요는 그림1과 같다. 이 시스템은 협업에 참여하는 레거시 시스템들의 데이터를 상호간의 실시간으로 동기화할 수 있다.

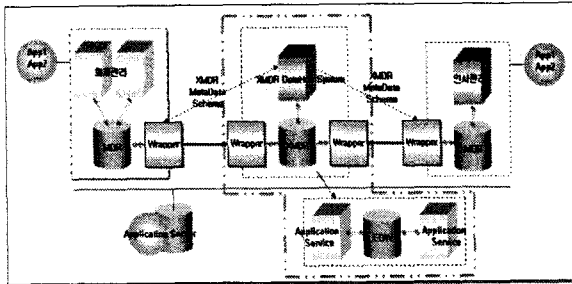


그림1. 시스템 개요

그림1에서 보는 바와 같이 XMDR 데이터 허브 시스템은 EDW와 같이 특정 서비스 애플리케이션을 제공하기 위해 각 회계관리, 인사관리, 참고관리 등의 레거시 시스템들이 협업에 필요한 데이터를 Wrapper를 통해 실시간으로 변화된 데이터를 일관성 있게 유지하기 위한 협업 시스템 구성이다. 일반적으로 데이터 통합에서는 데이터에 대한 일관성을 유지하는 방법이 Snap Shot, DB Link, 일괄처리, Replication, Internal Trigger 등의 방법을 사용한다. 하지만, 이런 데이터 동기화 방법들은 수시로 변화는 데이터를 실시간으로 처리할 수 없는 단점이 있다. 그래서 데이터의 신속성 및 신뢰성에 대한 문제가 제기된다. 그림1은 위의 단점을 해결한 시스템이다. 이러한 데이터 통합은 XMDR 데이터 허브 시스템을 구성함으로써 레거시 시스템간의 데이터 공유가 가능하고, 검

색 정보를 단일 뷰로 제공하며 데이터의 무결성을 보장함으로써 통합된 정보 가치를 보장해 줄 수 있다. 레거시 시스템간의 데이터를 협업할 때 실시간으로 변화는 데이터를 일관성 있게 유지하기 위해서 데이터 협업 메커니즘은 3.2절에 설명한다.

#### 3.2 실시간 데이터 협업 메커니즘

실시간 데이터 협업 메커니즘은 레거시 시스템과 이기종 애플리케이션을 활용하거나 확장하는 경우 사용되는 메커니즘으로 EDW, EAI와 같은 통합된 단일 데이터 보관을 운영하고 모든 데이터를 중앙화하도록 지원한다. 특히, 기존 데이터 동기화 방식인 Snap Shot, DB Link, 일괄처리, Replication, Internal, Trigger와의 다른 처리 방식인 본 논문에서 제안한 실시간 데이터 협업 메커니즘은 그림2와 같다.

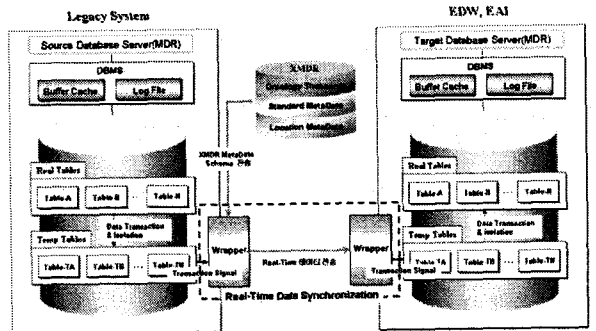


그림2. 실시간 데이터 협업 메커니즘

그림2에서 보는 바와 같이 레거시 시스템에 Real Tables는 실제 레거시에서 운용되는 데이터 테이블이다. Temp Tables는 현재 EDW, EAI에 필요한 협업 데이터 테이블을 구성한 것이다. 만약 Real Tables에서 변경된 데이터가 발생하는 경우 즉각 협업 중인 Temp Tables에 데이터 트랜잭션 된다. 그리고 Temp Tables에서 트랜잭션이 발생할 때마다 트랜잭션 신호로 Wrapper에 알린다. Wrapper에서는 Temp Tables에서의 변경된 데이터를 4장에 설명된 XMDR에서 메타데이터 스키마를 참조하여 EDW, EAI측 Wrapper에 XML문서를 전송한다. EDW, EAI측 Wrapper는 XML문서를 파싱하여 Temp Tables에 데이터 트랜잭션을 하게 되고, Temp Tables는 현재 통합적으로 운용되고 있는 Real Tables에 트랜잭션 하여 실시간으로 데이터의 일관성을 유지한다. 위와 같은 방식은 기존 Snap Shot, DB Link, 일괄처리, Replication, Internal Trigger의 단점인 데이터의 신속성 및 신뢰성을 해결할 수

게 된다. 또한 모든 데이터 처리 과정을 트랜잭션 기법을 이용하기 때문에 데이터 동기화시 오류 및 장애등의 문제도 해결할 수 있게 된다.

4. XMDR 모델 구성

본 논문은 ISO/IEC 11179의 XMDR을 바탕으로 XMDR의 모델의 구성을 제시한다. 모델 구성은 그림3처럼 온톨로지 시소러스, 표준 메타데이터와 위치 메타데이터로 구성된다. 이 XMDR은 레거시 시스템간의 협업에서 발생하는 메타데이터 스키마 이질성, 실제 데이터 값의 의한 인스턴스간의 이질적인 문제를 해결해주는 기능을 제공한다. 이 장에서는 XMDR의 구성요소인 온톨로지 시소러스, 표준, 위치메타데이터 정의 및 설계와 XMDR의 구성개념 및 XMDR 기반의 데이터 허브 시스템 구조에 대해 기술한다.

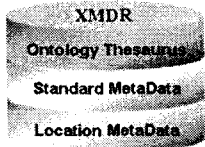


그림3. XMDR 개념도

4.1 온톨로지 시소러스

온톨로지 시소러스는 데이터 값과 데이터 값 사이에 인스턴스 연관 관계를 구조화하여 구체적으로 지식을 표현한 것으로, 그림4는 XMDR 모델 구성에서 정의된 메타데이터간의 스키마 매핑 및 데이터 값 인스턴스의 매핑을 표현한 것이다. 다시 말해 협업할 때 발생하는 이질성을 해결하고자 데이터 값, 스키마의 연관 관계성을 정의한 것이다. 그림4는 자산관리 솔루션 시스템에서 사용되는 스키마와 데이터를 요소화하여 4가지 연관 관계를 도식한 것이다. 이 온톨로지 시소러스의 연관 관계는 다음과 같이 정의된다.

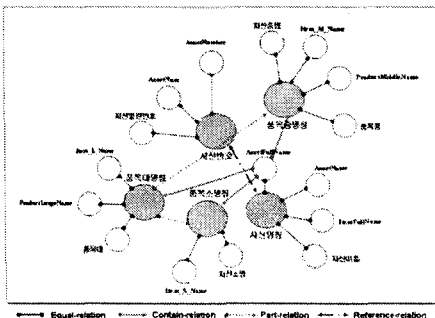


그림4. 온톨로지 시소러스

- 동등관계(Equal-relation) - 데이터의 표현하는 방식은 다르지만 의미상 동등한 의미를 가지는 관계를 표현.
  - ▷ 자산번호에서 자산번호, AssetNumber, 자산일련번호는 동등관계이다.
- 포함관계(Contain-relation) - 데이터의 표현하는 방식이 여러 개의 데이터를 하나의 의미로 표현.
  - ▷ 품목대명칭, 품목중명칭, 품목소명칭은 AssetFullName으로 대표하는 관계이다.
- 부분관계(Part-relation) - 데이터의 상위, 하위 개념을 의미하는 관계를 표현.
  - ▷ 품목대명칭은 품목중명칭 또는 품목소명칭은 부분관계이다.
- 참조관계(Reference-relation) - 데이터간의 참조 역할이 되는 관계를 표현.
  - ▷ 자산번호는 자산명칭과 참조관계이다.

이 온톨로지 시소러스에 4가지 연관 관계를 토대로 레거시 시스템에서 데이터 협업을 하기 위한 메타데이터 스키마, 데이터 값의 인스턴스 관계매핑을 하기 위해 참조되는 사전적인 역할을 한다. 예를 들어, 자산번호를 가지고 스키마 매핑을 할 때 다른 레거시에서 사용하는 스키마가 AssetNumber인 경우 온톨로지 시소러스에 정의된 자산번호 요소가 AssetNumber와 Equal-relation 관계를 참조하여 매핑 함으로써 협업을 할 수 있게 되는 것이다. 이를 통해 데이터 통합 시 발생하는 의미적 상호 운용성의 문제점을 해결할 수 있다. 그림4는 온톨로지 시소러스를 문서교환이 표준인 XML로 표현하면 그림5와 같다.

```

<?xml version="1.0" encoding="utf-8"?>
<Ontology xmlns="http://www.xmdr.org" id="http://www.xmdr.org/ontology/location/ontology.xml">
  <AssetNumber>
    <자산번호 id="9551118001">
      <자산번호 id="9551118001" id="AssetNumber"/>
      <자산번호 id="9551118002" id="AssetNumber"/>
      <자산번호 id="9551118003" id="AssetNumber"/>
    </자산번호>
    <품목대명칭 id="9551118002">
      <자산번호 id="9551118002" id="AssetNumber"/>
      <자산번호 id="9551118002" id="AssetNumber"/>
      <자산번호 id="9551118002" id="AssetNumber"/>
    </품목대명칭>
    <품목중명칭 id="9551118003">
      <자산번호 id="9551118003" id="AssetNumber"/>
      <자산번호 id="9551118003" id="AssetNumber"/>
    </품목중명칭>
    <품목소명칭 id="9551118004">
      <자산번호 id="9551118004" id="AssetNumber"/>
    </품목소명칭>
    <자산명칭 id="9551118005">
      <자산번호 id="9551118005" id="AssetNumber"/>
    </자산명칭>
  </AssetNumber>
  <AssetName>
    <자산명칭 id="9551118001" id="AssetName"/>
    <자산명칭 id="9551118002" id="AssetName"/>
    <자산명칭 id="9551118003" id="AssetName"/>
    <자산명칭 id="9551118004" id="AssetName"/>
    <자산명칭 id="9551118005" id="AssetName"/>
  </AssetName>
  <AssetLocation>
    <자산위치 id="9551118001" id="AssetLocation"/>
    <자산위치 id="9551118002" id="AssetLocation"/>
    <자산위치 id="9551118003" id="AssetLocation"/>
    <자산위치 id="9551118004" id="AssetLocation"/>
    <자산위치 id="9551118005" id="AssetLocation"/>
  </AssetLocation>
  <AssetStatus>
    <자산상태 id="9551118001" id="AssetStatus"/>
    <자산상태 id="9551118002" id="AssetStatus"/>
    <자산상태 id="9551118003" id="AssetStatus"/>
    <자산상태 id="9551118004" id="AssetStatus"/>
    <자산상태 id="9551118005" id="AssetStatus"/>
  </AssetStatus>
  </Ontology>
  </Ontology>
</pre>

```



을 구성하는 레거시 시스템들의 스키마의 구조 변경이나 메타데이터의 의미 변경 없이 협업에 참여할 수 있도록 하였다.

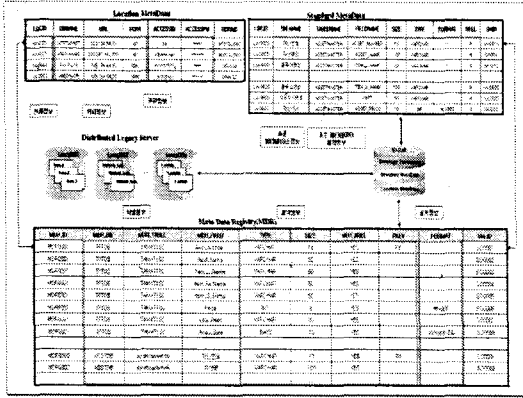


그림8. XMDR의 구성 개념

4.5 XMDR 기반의 데이터 허브 시스템 구조

XMDR 기반의 데이터 허브 시스템은 그림9와 같이 구성되며, 각 계층은 시스템을 구성하기 위한 XMDR 데이터 허브 모듈과 응용 계층과 협업에 참여한 레거시의 데이터베이스 계층으로 이루어져 있다.

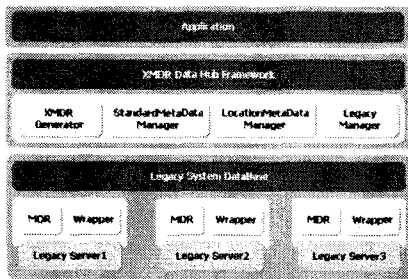


그림9. XMDR 기반의 데이터 허브 시스템 구조

이 시스템에서 협업에 참여한 레거시 시스템들은 참여 등록 및 스키마 매핑되어 있어야 하며, 그렇지 않은 시스템은 XMDR 데이터 허브 시스템을 이용한 레거시 시스템 상호간의 데이터 공유를 할 수 없게 된다. 그림9는 XMDR 기반의 데이터 허브 시스템 구조이다. 각 모듈의 기본적인 기능과 설명은 다음과 같다.

- 애플리케이션 : 단일 뷰어로 통합 정보를 요청하는 End-User에 대한 기본 검색 및 확장 검색이 가능한 애플리케이션을 제공한다.

- XMDR 생성기 : XMDR의 스키마 문서를 생성해주는 모듈이다.
- 표준메타데이터 관리 : 협업 시스템을 구성하기 위해 필요한 표준 메타데이터를 관리하는 모듈이다. 메타데이터 항목에 대해 추가, 수정, 삭제될 때 관리하는 모듈을 제공한다.
- 위치메타데이터 관리 : 레거시 시스템들에 접근하기 위한 메타데이터 위치접근정보, 물리적인 정보를 관리하는 모듈이다. 협업 관계에 있는 레거시 시스템들이 데이터 동기화를 정보를 제공한다.
- 레거시 관리 : 레거시 시스템의 정보 변경, 스키마 자체 내용이 변경되어 표준 메타데이터와 매핑 내용이 변경될 경우 관리하는 모듈이다.
- Wrapper : 협업된 다른 레거시 시스템에 실시간으로 데이터의 동기화를 하는 미들웨어이다.

5. 시스템 적용 및 비교분석

5.1 시스템 적용

제안된 시스템을 적용하기 위한 시스템 환경은 Windows 2003 Server, MS-SQL 2000, Oracle8i, ASP.NET을 이용하였다. 적용 사례는 협업된 통합 자산 관리 시스템을 구현하였다. 그림10의 인터페이스는 관리 중인 레거시 시스템의 메타데이터 스키마 정보를 XMDR 기반의 데이터 허브 시스템에 매핑하여 협업 서비스에 참여시킨다. 스키마 매핑 인터페이스에서 현재 서비스 중인 레거시 스키마를 선택하면, 각 원천서버에서 이미 매핑된 정보가 있을 경우 매핑 정보를 보여주며, 변경이 필요한 경우 매핑 정보를 수정한다. 현재 협업 시스템에 등록된 레거시 시스템의 DB를 선택하고, 테이블 필드 명을 선택한다. 위치 메타데이터와 표준 메타데이터에서 현재 매핑할 레거시 시스템의 스키마 정보와 맞는 메타데이터를 선택하여 매핑을 하게 되면 데이터베이스에 즉시 적재된다.

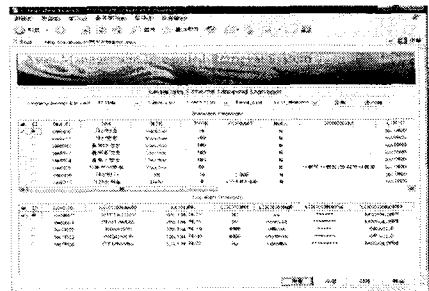


그림 10 스키마 매핑 인터페이스

표 1. 타 프레임워크와의 비교

	Die Ink 프레임워크	o6/64 프레임워크	Powerlink 프레임워크	본 프레임워크
XMDR 지원	일부지원	일부지원	일부지원	지원
데이터 공유 자동화	지원	일부지원	일부지원	지원
범용 명세 스키마	지원 없음	지원	지원	지원
적용 범위	모든 분야	모든 분야	IT/EC 분야	모든 분야
저장소 구조	중앙 집중	분산	분산	분산과 집중
데이터 연관성	부분지원	부분지원	부분지원	지원
실시간 데이터 동기화	일부지원	일부지원	일부지원	지원

그림 11은 시스템에서 사용자에게 제공하는 검색 화면과 검색된 결과이다. 레거시 시스템 측의 사용자는 검색 분류 선택에서 기본, 확장을 선택한다. 기본검색 선택인 경우에는 자체 레거시 시스템에 대한 검색을 원하는 것이고, 확장 검색을 선택한 경우에는 참여된 레거시 서버들의 검색을 요청하는 것이다. 그리고 사용자가 검색을 원하는 서버 선택할 수 있게 하여 보다 정확한 검색을 할 수 있게 하였고, 검색 결과 개수 또한 사용자에게 입력할 수 있게 하였다.

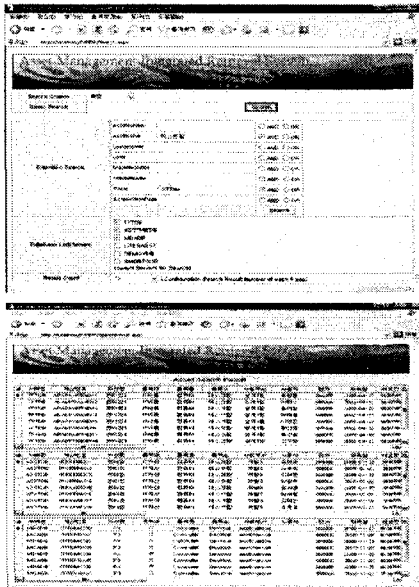


그림 11. 검색 인터페이스 및 검색 결과 화면

### 5.2 타 시스템과 비교

본 논문에서 제안한 시스템은 분산 환경에서 이질성 문제 해결과 데이터 통합 및 공유를 실시간 동기화를 지원하기 위하여 연구된 시스템이다. 따라서 표 1에서와 같이 세 가지 프레임워크를 대상으로 다음 7가지 항목으로 구분하여 비교하였다. 그 항목으로는 XMDR 지원여부, 데이터 공유 자동화여부, 범용 명세 스키마 지원여부, 적용 범위, 저장소 구조, 데이터 연관성, 실시간 데이터 동기화에 대한 부분으로 비교가 된다.

### 6. 결론 및 향후연구

본 논문에서는 데이터 통합에 필요한 레거시 시스템간의 데이터를 협업할 때 실시간으로 변화는 데이터를 일관성 있게 유지한 메커니즘을 제안하였다. 또한 XMDR을 이용하여 협업에 의해 데이터 통합 시 발생하는 의미적 상호 운용성의 문제점을 해결하는 데이터 허브 시스템을 구축하였다. 따라서 본 논문에서 제안한 시스템은 기업간의 정보를 공유하고 데이터를 통합하는 적합하게 적용될 것으로 기대된다. 기대효과는 기술 검증과 이전을 통한 기업의 기술 가치가 향상되고, 새로운 비즈니스 모델의 제시를 통한 시장이 창출되고, 기업 간의 정보를 공유 및 통합함으로써 생산성을 향상시킬 수 있다. 이후는 그리드 환경에 도입할 필요가 있으면, XMDR과 레거시 시스템 사이의 표준 변환을 위한 래퍼의 파싱 기술과 지식 표현 기술에 대한 연구가 더 필요하다.

#### [참고문헌]

- [1] Andrew B. Whinston and Varghese S. Jacob. "Electronic Commerce." Information Technology & Management 1., 2000.
- [2] Christopher C. Shilakes and Julie Tylman, "Enterprise Information Portals," Merrill Lynch, Inc., NY, November 16, 1998
- [3] David Wang, "Automated Semantic Correlation between Multiple Schema Information Exchange", M.I.T, MM, May, 2000.
- [4] ISO, Metadata Registry, ISO-11179, 2000.
- [5] <http://xmdr.org>
- [6] <http://www.oracle.com>