

# 지구 관측자료 공유를 위한 메타데이터 연구

## A Study on Metadata for Sharing the Information of Earth Observation

이 혜 영(Hye-Young Lee)\*

곽 승 진(Seung-Jin Kwak)\*\*

### 목 차

- |                         |                        |
|-------------------------|------------------------|
| 1. 서론                   | 3.3 메타데이터 요소 도출과정      |
| 2. 이론적 배경               | 4. 자료공유를 위한 KGEO 메타데이터 |
| 2.1 메타데이터와 메타-메타데이터     | 4.1 메타데이터 스키마          |
| 2.2 메타데이터 상호운용성         | 4.2 메타데이터 요소           |
| 2.3 선행연구                | 4.3 메타데이터 검증 및 확장방안    |
| 3. KGEO 메타데이터 설계        | 4.4 메타데이터 합의를 과정       |
| 3.1 국내 지구 관측자료 메타데이터 현황 | 4.5 분야별 메타데이터 통합방안     |
| 3.2 메타데이터 설계방향          | 5. 결론                  |

### 초 록

본 연구는 지구 관측자료 공유 및 효율적 활용을 위한 통합 메타데이터를 설계하는 것이다. 최근 정부 간 조직으로 구축된 GEO(Group on Earth Observations)는 지구전반에 걸쳐 지속적으로 관측되는 자료를 활용하여 인류의 이익을 최대 창출하기 위해 노력하고 있다. 이러한 세계적 흐름에 따라 국내에서 관측되는 지구 관측자료 즉, 재해, 보건의료, 에너지자원, 기후, 수자원, 기상, 생태계/생물다양성, 농업, 산림, 해양, 우주, 국토공간정보 등 12개 분야의 관측자료를 통합 및 활용하기 위한 인프라 구축이 반드시 필요하다. 현재 지구관측자료는 관측분야별로 개별적인 메타데이터를 관리하고 있을 뿐 메타데이터를 통합관리하지 못하고 있다. 국내 지구 관측자료의 관측분야와 관측요소별 다양하고 상이한 메타데이터로 인한 통합메타데이터의 설계 어려움을 메타-메타데이터를 이용하여 개발하였다.

### ABSTRACT

The purpose of this study is to design the metadata for sharing and efficiently using the informations of Earth Observations in Korea. Recently the GEO(Group on Earth Observations) has been starting to manage all nation's informations in a way that benefits the environment as well as humanity by taking a pulse of the plants. For that reason, we have to construct the infra-system to manage and integrate the 12 social benefit areas in Korea: disaster, health, energy, climate, water, biodiversity, agriculture, forest, ocean, space information and geographical information system(GIS). But these informations have only been managed with varied metadatas in each area of observation, and then it is difficult to integrate varied informations. In this study, we solve the problem with meta-metadata.

키워드: 지구관측자료, 메타데이터, 모즈, 메타-메타데이터, 상호운용성, 크로스워크, 데이터 통합  
Earth Observation Information, Metadata, MODS, Meta-metadata, Interoperability, Crosswalk, Data Integration

\* 한국과학기술원 과학도서관 기술원(hye@kaist.ac.kr)

\*\* 충남대학교 문헌정보학과 조교수(sjkwak@cnu.ac.kr)

논문접수일자 2007년 5월 17일

게재확정일자 2007년 6월 16일

## 1. 서론

최근 정부 간 조직으로 구축된 GEO(Group on Earth Observations)은 지구 전반에 걸친 모든 분야의 포괄적인 관측과 더불어 장기간에 걸쳐 지속적으로 생산되는 관측 자료를 활용하여 인류의 최대 이익을 제공코자 노력하고 있다. 이것은 개별 관측분야의 역량을 합하여 부분적인 역량의 합보다 큰 시스템 개발로 이어지고 있다.

우리나라 또한 GEO 회원국으로서 이러한 세계적 흐름에 동참하기 위한 인프라구축이 시급한 실정이다. 이것은 국내 전 지역에서 관측되는 자료를 통합하고 공유하고 공동 활용하기 위한 메타데이터 설계를 포함한다. Global Earth Observation System of Systems(GEOSS) 10년 실행계획은 데이터 라이프사이클 관리, 데이터 통합과 정보 융합, 데이터 마이닝 등을 중요한 연구영역으로 제시하고 있다. 또한 향후 2년 이내의 GEOSS 역량구축 항목으로서 개별 국가의 데이터 접근을 용이하게 할 것을 권고하고 있다.

GEO에서 언급하는 지구 관측분야는 12개 분야로 재해, 보건의료, 에너지자원, 기후, 수자원, 기상, 생태계/생물다양성, 농업, 산림, 해양, 우주, 국토공간정보 관측자료이다. 그러나 현재 국내 지구관측 자료는 관측분야, 관측요소별로 다양한 메타데이터를 통해 관리되거나 메타데이터 없이 자료를 유지 관리하고 있어 자료 통합 및 활용을 저해하고 있다. 따라서 국내 지구 관측자료의 효율적인 통합운영체제 구축을 위한 기본 방안을 수립위한 한 부분으로써 매우 다양한 메타데이터로 이루어진 국내 지구

관측자료 통합을 위한 통합메타데이터 설계가 필요하다.

본 연구는 12개 지구관측분야의 관측 자료 현황 및 메타데이터 구축 현황을 조사하여 다양한 성격의 자료를 상호 공유하고 공동 활용하기 위한 통합 메타데이터를 설계하는 것이 목적으로 연구방법 및 절차는 다음과 같다.

첫째, 지구 관측자료를 공유하기 위한 메타데이터 설계는 국내 관측분야별 관측자료 메타데이터 현황 조사를 거쳐 이들 자원을 공유하기 위한 메타데이터 요소를 설계하였다. 관측분야별 관측자료 메타데이터 현황 조사는 지구 관측 자료를 생산 또는 구축하는 기관의 협조를 구하여 관측 자료와 메타데이터 구축현황을 조사하였다.

둘째, 자료 조사는 12개 관측분야에 해당하는 정보를 생산하거나 또는 이와 같은 자료를 구축하는 관련 기관의 협조를 통하여 통합 메타데이터 설계에 필요한 정보와 메타데이터 현황을 수집하였다. 자료 조사 기관은 기상청, 한국지질자원연구원, 건설교통부, 환경부, 농업과학기술원, 농촌진흥원, 산림청 국립수산과학원, 국립해양조사원, 국토지리정보원 등 12개 기관이다. 수집된 정보는 관측요소, 관측주기, 관측형태, 자료형식, 관측기간, 보유기관, 관측주기별 자료파일 용량, 메타데이터 유무, 자료공유가능여부, 서비스 시스템, 관측분야 관측요소별 메타데이터 스키마 등이다.

셋째, 자료공유를 위한 메타데이터 설계와 데이터 요소 선정에 있어서 국내 지구관측 자료를 생산하거나 데이터 구축에 관련된 기관 담당자, 연구자, 정보관리 전문가의 의견을 수렴하여 진행하였다. 다양한 성격의 데이터 호

환성을 유지하고, 네트워크 자원의 기술에 필요한 일련의 데이터 요소를 규정하여, 이들 자원을 신속하게 검색할 수 있는 메타데이터를 설계하였다.

넷째, 메타데이터 설계는 일반적으로 널리 사용되는 메타데이터 가운데 범용성을 지닌 메타데이터 요소를 중심으로 설계하여 다양한 지구 관측자료의 일반적인 공통된 데이터 요소 모두가 반영되도록 하였다. 관측자료의 일반적인 데이터 항목을 중심으로 관측자료의 다양한 메타데이터 수집 및 관리가 가능하도록 하였다.

다섯째, 관측분야별 관측 자료의 개별 메타데이터의 요소 성격을 충분히 파악하여 공통된 데이터 요소를 도출하였으며, 도출된 공통의 데이터 요소는 메타데이터 통합과 공유를 위한 자격 여건을 평가하여 실제적인 통합 메타데이터의 데이터 요소로서 선정하였다.

## 2. 이론적 배경

### 2.1 메타데이터와 메타-메타데이터

메타데이터(metadata)는 보통 데이터에 관한 데이터(data about data)로 설명되며 데이터를 탐색, 식별, 전달, 접근 그리고 활용할 수 있게 하는 것으로 때로는 방대한 자료로 인한 혼돈을 말끔히 해결해 줄 수 있는 마법상자로 여겨지기도 한다. 메타데이터는 인쇄 출판물을 포함한 다양한 디지털 정보를 기술하며, 어떤 정보자원을 구조화하여 표현한다. 또한 메타데이터는 구조화된 표현 구조를 통해 정보자원에 관한 다양한 정보를 제공하여 다양한 용도로

자원을 활용할 수 있게 한다.

메타데이터는 데이터와 다른 것으로 실제적인 정보 자원을 말한다. 예를 들어 디지털 영화 파일의 경우 우리가 보는 실제적인 그림은 데이터이지만 디지털 영화 파일에 관한 영화이름, 제작일, 프레임 개수 등 어떤 방법에서 데이터를 설명하고 기술하는 것은 메타데이터이다. 따라서 메타데이터는 여러 측면으로 사용할 수 있는데 데이터를 기술할 수도 있지만, 영화파일을 저장하고 있는 컨테이너를 기술할 수도 있다.

Gilliland(2007)는 메타데이터를 용도에 따라 관리(administrative), 서술(descriptive), 보존(preservation), 기술(technical), 이용(use) 메타데이터로 구분하였다. 최근은 데이터 통합(data integration) 또는 데이터 웨어하우스(data warehouse)를 위한 메타데이터 설계 및 관리에 많은 관심이 집중되고 있는 실정이다.

Adrienne(2002)는 메타데이터를 일반적으로 데이터에 관한 데이터로 인식하는 잘못된 해석을 다시 분명히 정의해야 할 필요가 있다고 주장하고 있다. 그는 인스턴스 데이터(instance data)를 사용되는 틀, 어플리케이션, 데이터베이스 또는 간단한 처리 엔진에 입력되는 것이라고 할 때, 메타데이터는 인스턴스 데이터에 관한 자세한 기술 즉 메타데이터 사용자의 역할에 따른 인스턴스와 가치 있는 데이터에 대한 형태 및 특징이라고 기술하고 있다. 이것은 우리가 가지고 있는 데이터는 무엇이며, 이것은 무엇을 의미하고, 어디에 있으며, 어떻게 어긋나게 되었고, 어떻게 그것을 획득할 수 있는지에 대한 질문에 따라 메타데이터의 가치를 조망하고자 한 것이다. 하지만 그 또한 메타테

이터는 정보의 위치식별, 정보 기술, 데이터 통합을 위해 필요하다는 의견에는 같이하고 있다.

메타-메타데이터(meta-metadata)는 '메타 데이터에 대한 상세한 기술로써 메타데이터를 이해하기 위한 기본적인 속성으로 틀에 근거한 처리와 접근 허용을 위한 메타데이터 질과 위치'로써 정의하고 있다. 그는 메타-메타데이터 유형에 대하여 다음과 같이 언급하고 있다 (Adrienne 2002).

- 속성 메타-메타데이터(generic meta-metadata): 인스턴스에 근간하여 개별적으로 분리되어 있는 메타데이터를 통합
- 처리용 메타-메타데이터(processing meta-metadata): 속성 메타데이터와 관련된 처리를 표준화
- 접근용 메타-메타데이터(access meta-metadata): 메타데이터의 속성 유형으로 인한 인터페이스와 검색을 표준화
- 보안용 메타-메타데이터(security meta-metadata): 메타데이터의 속성에 대한 접근 및 처리를 제한 또는 허용

메타-메타데이터 역시 메타데이터이므로 구조화된 형태로 자원을 기술한다. 따라서 메타데이터의 특징을 그대로 가질 뿐 아니라 메타데이터로서 메타데이터의 속성을 기술하고 메타데이터의 접근과 처리에 관련한 기능을 가진다.

## 2.2 메타데이터 상호운용성

메타데이터가 증가됨에 따라 다양한 메타데이터를 때로는 다양한 도메인의 메타데이터를 상호운용하기 위한 노력이 증대되고 있다. 메타데

이터 상호운용성(interoperability)은 다양한 정보 자원 접근을 가능하게 하고, 다른 유형의 도메인에 걸쳐 데이터를 생성하고 공유하는 것이다. 메타데이터를 상호운용하기 위한 연구방법들로 MARC 기반 도서관 종합목록, 메타데이터 레지스트리, 메타검색, 크로스워크(crosswalk) 등이 있다.

전통적인 MARC 기반의 도서관 종합목록은 다양한 정보 자원 출처로부터 메타데이터를 위한 중앙 데이터베이스를 구축하는 것으로 단일한 구조의 종합 색인만 구축할 수도 있다. 이것은 다양한 정보 출처에서 MARC이라는 표준화된 공통의 포맷과 공통의 편목 규칙을 사용하기 때문에 많은 노력 없이 비교적 좋은 효과를 얻을 수 있지만, 상이한 메타데이터 스키마로 구축되는 많은 정보 출처가 있을 경우 단순한 메타데이터 수집만으로는 의미적 상호운용을 기대하기 어렵다.

또 하나의 정보공유 방법인 메타데이터 레지스트리는 메타데이터의 등록과 인증을 통하여 표준화된 메타데이터를 유지 및 관리하고, 메타데이터의 명세와 의미의 공유를 목적으로 하는 것으로 데이터의 표준화를 지원하고 데이터 공유를 가능하게 한다. 그러나 이것은 특정한 메타데이터 데이터 그룹을 검색하고 접근할 수 있도록 하는 것으로 지구 관측자료를 통합 검색하고 접근하는 관점에서는 한계점이 있다.

메타데이터 공유를 위하여 정보교환프로토콜을 이용하기도 하는데, 메타데이터의 의미적 상호운용측면에서는 효용성이 떨어진다. OAI-PMH (Open Archive Initiative-Protocol for Metadata Harvesting)는 다양한 정보원의 다른 유형의 메타데이터를 능동적으로 수집할 수 있는

HTTP와 HTML 기반의 어떤 시스템 환경에서도 도입 가능한 간단하지만 효과적인 시스템 모델을 제공한다. Z39.50 또는 웹 서비스는 또 다른 메타데이터 상호운용 접근방법이다. 이것은 정보를 제공하는 시스템 개별을 분산 검색하는 방법으로 메타검색이라고 불린다. 브로드캐스트 탐색을 하여 정보원의 메타데이터를 접근하는 것이다.

의미적인 상호운용 방법으로는 온톨로지(ontology)와 크로스워크를 들 수 있다. 온톨로지는 일반적으로 개념화를 위한 명세로서 정의되는 것으로, 정보시스템 개발의 또 다른 측면에서 중요한 역할을 하고 있다. 여러 형태의 메타데이터의 상호운용을 위해 각 메타데이터에 포함하고 있는 메타데이터 요소들의 의미론적 의미를 정의한 온톨로지를 사용하여 상호운용을 해결하기 위한 방법이다.

RDFS(Resource Description Framework Schema)는 이름 붙여진 속성과 값의 차원에서 정보 자원 간의 상호 관계성을 기술한, 간단한 모델의 RDF를 보완하여 어떤 정보원에서 사용하는 용어의 해석을 정의하고 제한할 수 있는 구조를 제공한다. 이러한 견지에서 RDFS는 특별한 프로파일을 이용한 메타데이터 요소의 온톨로지라 할 수 있다. 지구관측 자료 가운데 하나인 GIS 분야에서도 온톨로지를 이용한 데이터 공유 사례 또한 확인 할 수 있다(Pundt 2002).

메타데이터 간의 상호운용으로써 크로스워크는 더블린코어 메타데이터 용어사전에 따르면 '2개 이상의 메타데이터 스키마 간의 관계와 동일한 것을 관계지어 놓은 테이블로써 이질적인 데이터베이스를 걸쳐 효율적으로 검색할 수 있도록 검색엔진을 지원한다'라고 정의하고 있

다. 크로스워크는 어느 한 메타데이터 스키마에서 다른 메타데이터 스키마로 일반적으로 이루어지는 매핑이다. 그래서 메타데이터 스키마를 정의하는 표준이 없을 경우 서로 다른 메타데이터 스키마의 요소들은 서로 다른 것을 명시하거나 또는 동일한 속성을 다른 이름으로 명시하여 사용할 수 있기 때문에 매우 복잡해 질 수 있다. 그러나 메타데이터가 표준화 되어 있는 정보영역 즉 도메인에서는 많은 경우, 크로스워크를 통해 메타데이터 상호운용을 해결하고 있다.

본 연구의 지구 관측자료 또한 관측분야와 관측요소별 메타데이터를 사용하고 있어 이질적인 메타데이터를 통합하고 효율적인 검색과 접근을 위해 크로스워크를 이용한 메타데이터 공유 및 통합이 적합한 것으로 분석된다. 그러나 12개 관측분야의 다양한 메타데이터를 하나의 메타데이터로 완전히 매핑하여 통합한다는 것은 메타데이터의 데이터 요소를 수없이 증가시키는 결과를 초래할 것이다. 따라서 메타-메타데이터를 지구 관측자료를 통합하기 위한 메타데이터 설계모형으로 참조하였다.

### 2.3 선행연구

국내 지구 관측 자료를 생산하는 각 관측분야에서는 이미 데이터를 통합하고 체계적으로 관리하고자 하는 연구들이 진행되고 있다. 한종엽(2003)은 '해양과학기술 R&D 결과정보 데이터베이스 구축 연구'에서 문헌정보뿐만 아니라 사실정보로써 해양조사자료의 효율적인 데이터베이스를 구축하였다. 더블린 코어 메타데이터를 준용하여 XML 형식의 메타데이터베이스를 구축하고 전문의 통합관리 및 검색 구현방법을

효율적으로 개선하였다. 해양조사자료는 국가 지식정보관리 차원에서 해양물리자료, 해양화학자료, 해양생물자료, 해양지질 및 지구물리자료, 해양기상자료 등을 포괄하고 있다.

안부영 등(2005)은 생물다양성 데이터교환을 위한 메타데이터 스키마 설계에 관한 연구에서 생물다양성 메타데이터 스키마를 설계하여 발표하였다. 이것은 다양한 생물개체의 공통된 특성을 추출하여 생물다양성 데이터의 접근이 용이하도록 정의된 표준화된 데이터 형식인 Darwin Core를 확장하여 국내에서 활용 가능하도록 한 것이다.

지리정보 관련 지구 관측자료 메타데이터 연구에서도 활발한 연구활동을 볼 수 있다. 송용철(2004) '그리드데이터 메타데이터 내용모델에 관한 연구'에 기록된 국가지리정보체계(NGIS)의 자료이력서(메타데이터) 잠정표준에 따르면, 한국전산원을 주축으로 1999년 발표된 NGIS 메타데이터 잠정표준은 국가 지리정보에 대한 공동 활용과 원활한 유통을 지원한다. 이것은 ISO19115 CD버전1을 기준으로 하여 메타데이터의 구성요소와 내용에 대한 공통된 기준을 제시한다. 이와 같이 각 지구 관측분야에서 그들의 자료를 통합하고 공동 활용하기 위해 자료 특성에 적합하고 국내의 메타데이터 표준 기준에 따른 메타데이터 표준화 작업이 현재에도 계속 이루어지고 있다.

관측 자료의 상호운용성 측면에서도 관련 연구들을 확인할 수 있다. Nogueras(2004) 등의 논문은 크로스워크를 이용한 GIS 분야 데이터 공유를 성취하고자 노력하였다.

국내 지구관측 자료 메타데이터 현황조사에 따르면, 지구관측 자료는 다소 메타데이터화

되어 있지 않은 것을 알 수 있다. 그러나 국내외 연구사례를 통하여 GIS의 지리정보뿐 아니라 전 지구의 많은 관측분야에서 그들의 도메인상의 관측 자료를 통합하고, 메타데이터를 표준화하고자 노력하는 지구관측분야별 연구 노력을 확인 할 수 있다.

### 3. KGEO 메타데이터 설계

#### 3.1 국내 지구 관측자료 메타데이터 현황

지구 관측자료 통합 메타데이터 설계하기 위하여 GEO에서 포괄하고자 하는 국내 지구 관측 자료를 조사하였다. 해당기관에 의뢰하여 2006년 9월 15일부터 10월 5일까지 약 20일간 수집한 자료를 중심으로 관측분야별 메타데이터 구축현황과 관측실태를 분석하였다. 우주관측분야 데이터는 자료 수집과정에서의 어려움으로 국내 지구 관측자료 메타데이터 현황 조사에서 제외되었다. 우주관측분야를 제외한 보건의료, 에너지자원, 기후, 수자원, 기상, 생태계/생물다양성, 농업, 산림, 해양, 국토공간정보, 재해 관측분야에 대한 국내 메타데이터 현황을 조사한 결과는 <표 1>과 같다.

국내 지구 관측자료 11개 분야의 메타데이터를 분석하였을 때, 매우 다양한 관측요소가 존재할 뿐만 아니라 관측자료의 자료형식과 관측 주기도 매우 다양하다는 것을 확인 할 수 있었다. 11개 관측분야 가운데 메타데이터를 구축하고 있는 관측분야는 8개 관측분야(72%)이며, 이 가운데 2개 관측분야는 DB형태의 메타데이터를 구축하고 있는 것으로 조사되었다.

〈표 1〉 국내 지구관측 분야별 메타데이터 현황 분석

관측분야	관측요소 (데이터그룹)	관측주기	자료형식	보유기관	메타 데이터	자료공유 가능여부
보건의료	1	실시간	DBF	질병관리본부	DB형태	가능
에너지자원	3	-	dx shp	한국지질자원연구원	DB형태	가능
기 후	3	월, 년	텍스트	기상청	있음(1)	가능
수자원	17	월~연간	텍스트	건설교통부	있음	가능
기 상	27	30분~12시	GRIB BUFR	기상청	없음	가능
생태계/ 생물다양성	13	10년	GIS-DB	환경부	있음	일부제외
농 업	5	5~40년	GIS파일 Execl..	농업과학기술원(2) 농촌진흥원(3)	있음(2)*	일부제외
산 립	2	10,100년	GIS파일	산림청	있음	가능
해 양	32	1초~1년	DB Seg-Y..	국립수산과학원(17) 국립해양조사원(15)	있음(20)	일부제외
국토공간 정보	11	5~10년	벡터 라스터...	국토지리정보원	있음(6)	일부제외
재해	2	-	Excel	기상청	없음	가능

자료공유 가능여부에 있어서 생태계/생물다양성, 농업, 해양, 국토공간정보 4개 분야의 일부를 제외하고 대부분 자료를 공유할 수 있는 것으로 나타났다. 자료를 보유한 기관에서 '자료를 공유할 수 있는가'라는 물음은 통합 메타데이터를 설계하고 자료를 공유하기에 앞서 선결되어야 할 문제이다. 11개 지구 관측분야에서도 〈표 1〉에서 확인 할 수 있듯이 데이터 조사결과, 자료공유가 허가된 관측 자료는 그리 많지 않다. 그러나 자료 공개와 같은 문제는 자료보유기관의 정책 또는 보안과 관련된 사안으로 통합 메타데이터를 설계하는 것과는 구분하여 별개의 과제로서 해결해야 할 것이다.

관측자료 메타데이터 스키마들은 관측 자료의 성격과 특징 그리고 자료보유기관의 사용용도에 따라 대체로 적절한 메타데이터 요소를 선별하여 사용하고 있었다. 예를 들어, 관측분

야 '기후'의 경우 CO2, 산성비, 오존 등 3개의 관측요소로 구분된다. 다른 관측분야도 마찬가지로이지만 관측분야 '기후' 또한 관측분야에 포함되어 있는 관측요소 모두를 위한 메타데이터가 존재하지 않고 오직 관측요소 '오존'만 메타데이터가 존재한다. 메타데이터 '오존'의 데이터 요소를 분류해보면, 관측대상 content, 관측자 DATA\_generation, 관측장소 Platform, 관측시간 Timestamp, 관측주기 Daily, 관측방법 Instrument 으로 나눌 수 있다. 또한 이들을 메타데이터 요소 그룹으로 임의적으로 구분하면, 관측대상은 식별정보로 구분할 수 있고 관측자, 관측장소, 관측시간, 관측주기, 관측방법은 발행사항으로 구분 할 수 있다.

지구관측 자료를 공유하기 위한 메타데이터 설계에 반영되어야하는 메타데이터는 관측분야 관측요소에 따라 구분되어 약 110여개의 메

〈표 2〉 관측분야 “기후” 데이터 현황

관측요소	관측 주기	관측 형태	자료형식	관측기간	보유기관	관측주기 별 용량	메타데이터 유무	자료 공유
CO2, CH4, N2O, CFC, UV-B	월	자동	문숫자 (텍스트파일)	1999년~2005년	기상청	-	무	가능
산성비	년	자동	문숫자 (텍스트파일)	1999년~2005년	기상청	-	무	가능
오존	월	자동	문숫자 (텍스트파일)	1999년~2005년	기상청	-	유	가능

〈표 3〉 관측분야 “기후”의 관측요소 “오존” 메타데이터 예제

항 목	내 용
Content	ozone
DATA_Generation	기상청
Platform	포항, 한국
Instrument	Brewer
Location	36,033, 129,383
Timestamp	UTC
Daily	Date, Monthly

타데이터 유형이 존재하는 것으로 조사되었다. 이렇게 많은 메타데이터 유형은 메타데이터의 상호교환 및 검색, 자료 공유와 활용을 어렵게 한다. 국내 관측 자료의 데이터 구축현황을 조사한 결과, 지구관측 자료는 단지 “지구 관측자료” 라는 울타리로 둘러싸여져 있을 뿐이다. 이들 자료는 서로 공통된 데이터 요소가 매우 적다. 이 때문에 수많은 관측자료 메타데이터를 단일한 하나의 통합 메타데이터 안에 통합 메타데이터 요소 안에 모두 수용한다는 것은 불가능한 것으로 보인다. 하지만 메타-메타데이터, 메타데이터 상호운용방법 그리고 선행연구들을 기반으로 하여 메타데이터 통합 방안을 모색할 수 있다.

분명한 것은 앞으로 설계될 통합 메타데이터는 수많은 메타데이터 스키마에 따라 바다, 육

지, 하늘, 생태계 등 지구 전반에 걸쳐 실시간으로 측정되고 축적되는 방대한 데이터의 메타데이터를 통합할 것이다. 그리고 이러한 지구 관측자료를 공유하기 위한 통합 메타데이터는 관측분야별 메타데이터 통합, 검색, 자료 발견과 획득, 원시자료(source data) 접근을 용이하게 할 것이다.

### 3.2 메타데이터 설계방향

사실상 지구상의 방대하고 다양한 관측 자료를 통합, 관리함에 있어 메타데이터는 필수적이다. 하지만 지구 관측 자료는 자료의 양만큼이나 메타데이터 스키마도 다양하여 국내 지구 관측자료 메타데이터 현황 조사에서 언급한 것처럼 지구 관측분야, 관측요소에 따른 110개 이



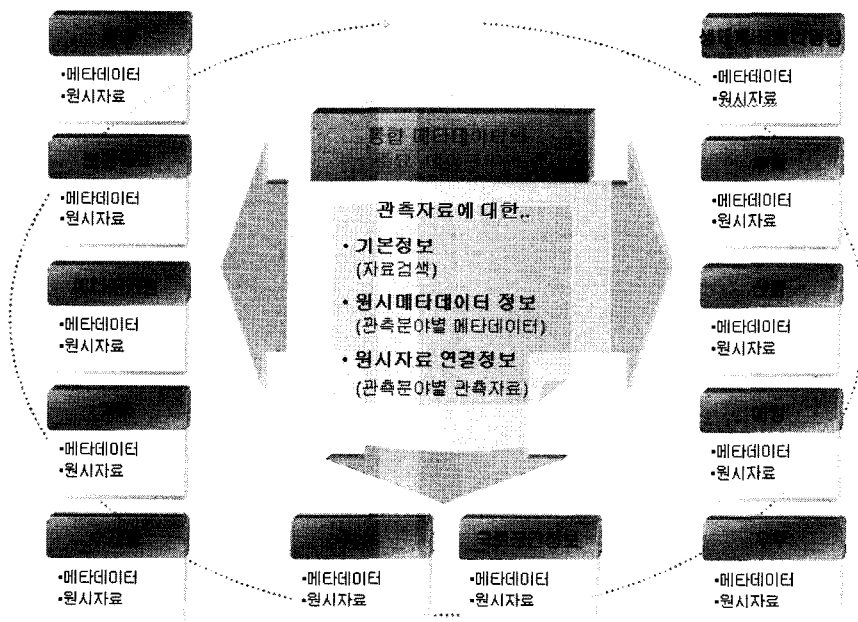
상의 메타데이터 스키마가 존재한다. 이러한 이유로 메타-메타데이터와 메타데이터 상호운용방법에 주안점을 두어 통합 메타데이터를 설계하였다. KGEO 메타데이터를 통한 자료 접근 개념은 <그림 1>과 같다.

관측자료의 일반적이고 공통적인 데이터 항목을 추출하여 메타데이터 필수요소를 선정하고, 모든 지구 관측자료를 검색 및 접근하기 위한 메타-메타데이터를 작성하는 것이다. 기존의 지구 관측자료 메타데이터는 크로스워크를 통하여 통합메타데이터 즉 메타-메타데이터로 전환될 수 있다. 메타데이터 매핑 테이블을 유지함으로 의미적인 메타데이터 상호운용성 즉 자료 공유기능을 제공하였다.

일반적으로 메타데이터를 설계할 때에는 설계기준이 되는 메타데이터를 선정하여 메타데

이터의 기본적인 구조적, 기능적 조건을 보장하여야 한다. 본 연구에서도 또한 메타데이터 설계기준이 될 수 있는 모범적인 메타데이터 스키마를 선정하여 메타데이터 기본 조건을 갖추고자 한다. 데이터 통합대상인 각 지구관측분야 메타데이터의 표준을 검토하여 메타데이터 설계기준을 선정할 수도 있다. 이미 상당히 많은 관측분야에 한하여 메타데이터가 표준화되어 있기 때문이다.

FGDC 표준에 따른 생물분야 메타데이터 National Biological Information Infrastructure scheme, 생태계 분야의 EML(Ecological Metadata Language), 생물 종의 표본 메타데이터 Darwin Core 메타데이터 등이 바로 그것이다. 이들 메타데이터 표준은 특정분야의 자료만을 기술하는 것으로 관측분야만 12개, 관측요소는



<그림 1> KGEO 메타데이터를 통한 자료 접근 개념

110여개가 되는 다방면에 걸친 관측 자료를 어느 한 특정 분야의 메타데이터로 일괄적 통합한다는 것은 문제가 있다. 따라서 본 연구에서는 12개 지구관측분야 데이터들의 공통적인 데이터 요소를 중심으로 통합 메타데이터를 설계하고자 한다. 이러한 이유로 범용 메타데이터를 채용하였다.

범용 메타데이터는 일반화된 메타데이터 스키마로서 다양한 자료에 적용할 수 있어, 우리의 모든 관측 자료에 적용하고 메타데이터 요소 간 상호 매핑에 적용할 것이다. 범용 메타데이터로서 이미 검증된 통합 메타데이터를 설계할 필요가 있다. 이들 표준화된 많은 범용 메타데이터 가운데 MODS(Metadata Object Description Schema)를 지구 관측자료 통합을 위한 메타데이터 설계 기준으로 선정하였다.

MODS는 세계적으로 통용되는 메타데이터와의 매핑에서 알 수 있듯이(표 4 참조) 기본적인 모든 데이터요소를 갖추고 있고 이미 더블린코어 등 많은 메타데이터와 크로스워크를 위한 매핑테이블이 작성되어 있어 메타데이터 상호운용성이 입증된 메타데이터이다. MODS는 전자 자원 관리를 위한 범용 메타데이터로서 XML 기반의 빠른 데이터 구축과 작성이 용이하다. 전자정보의 계층 간 연계정보도 관리할 수 있어 12개 지구관측 자료를 연계하기 위한 메타데이터 확장도 가능하게 한다. 이것은 이미 해양전문정보센터의 멀티미디어 데이터베이스와 한국과학기술정보연구원의 생명과학 오픈아카이브 구축에서도 채용한 바 있다. 이러한 적용사례는 통합 메타데이터 구축을 위한 12개 지구 관측자료 분야에 포함되는 관측분야이기도 하다.

〈표 4〉 MODS, 더블린코어, MARC 참조테이블

요 소	MODS 상위요소	더블린코어요소	MARC 필드
서명사항	titleInfo	Title	245, 210, 242, ...
저자사항	name	Creator, Contributor	1XX, 7XX
자료유형	typeOfResource	Type	008/06,07
장르	genre		
발행사항	originInfo	Publisher, Date	260, 008/15-17, ...
언어	language	Language	
형태사항	physicalDescription	Format	008/23,007/13, 856, ...
초록	abstract	Description	520
목차	tableOfContents	Description	505
이용대상	targetAudience	Audience	008/22
주기	note	Description	5XX
주제	subject	Subject	6XX
분류	classification	Subject	050, 08X
관련정보	relatedItem	Relation	76X, 78X, 534, ...
식별기호	identifier	Identifier	020, 022, 024, ...
소장위치	location		852, 856
이용제한	accessCondition	Rights	506, 540
로컬정보	extention		9XX
레코드정보	recordInfo		040, 001

지구 관측자료는 의미상으로 유사한 분야의 자료만을 수집하는 것이 아니라 지구 전반에 걸친 다양한 정보군의 자료를 통합하는 것을 목적으로 하기 때문에 메타데이터 상호운용방법 가운데 온톨로지는 적용하기 어려운 점이 있다. 메타데이터 상호운용을 위한 크로스워크 또한 관측자료 메타데이터 스키마의 요소들의 상당부분이 서로 상충되는 요소들이기 때문에 문제점을 지니고 있다. 이러한 이유로 모든 관측 자료를 기술하기 위한 기본적이고 공통적인 데이터 요소를 도출하여 이들 데이터 요소만을 크로스워크하는 데이터 통합방식을 적용하였다.

그러나 이미 구축되어진 지구 관측자료가 방대하므로 지구 관측분야 모든 메타데이터들을 상호운용하기 위한 메타데이터 참조데이터를 완전하게 적용할 수가 없다. 12개 지구 관측분야, 110 여개의 관측요소의 메타데이터 요소 모두를 포괄하는 메타데이터를 설계하는 것은 의미가 없다. 매우 다양한 메타데이터 유형은 메타데이터 요소간의 매핑률이 현저히 떨어지고 이로 인해 메타데이터 요소는 방대해 질 수밖에 없다. 상호 매핑되어진 데이터 요소일지라도 다양한 분야에서 데이터 요소를 사용하기 때문에 데이터 요소의 의미가 일반화되어 질수 있어 고유한 데이터의 성격과 의미가 상실될 것이다. 또한 방대한 메타데이터 요소는 매핑 테이블에 의한 메타데이터 매핑 작업에 있어서 상당한 시간과 노력을 요구 할 뿐 아니라, 각 관측분야의 메타데이터 스키마가 변경되었을 경우 메타데이터를 수정 보완해야하는 어려움에 봉착하게 된다.

본 연구에서는 이러한 이유로 인해 관측 자

료의 공통 데이터 요소를 우선 추출하고 이들 데이터 요소를 크로스워크하는 방법으로 지구 관측 자료의 통합 메타데이터를 설계하였다.

### 3.3 메타데이터 요소 도출과정

국내 12개 지구 관측분야 메타데이터를 통합하기 위한 메타데이터의 설계는 다음의 내용에 중점을 두었다. 첫째, 관측분야별 자료별 데이터들의 자료 공유와 공동 활용을 극대화한다. 자료 통합 검색 및 조회, 원시자료 접근뿐 아니라 메타데이터 통합에 따른 정보서비스 시너지 효과를 기대한다. 이것은 GEOSS 향후 10년 수행계획에서 언급한 내용으로 모든 데이터, 메타데이터, 생산물 접근이 용이해야 한다. 둘째, 국내 12개 관측분야의 자료 공유를 위한 메타데이터는 이용자의 자료 검색 효율성과 데이터 제공자의 메타데이터 작성의 편의성 및 관리의 효율성을 고려한다. 지구 관측자료의 통합, 공유, 공동 활용하는 이용자의 편리성, 효율성을 지원해야 한다. 셋째, 관측분야와 자료별 데이터의 고유한 데이터 요소는 보존한다. 분야별 메타데이터의 데이터의 특성 및 차별성을 박탈하지 않는다. 이것은 원시자료 및 원시메타데이터를 유지하여 GEOSS의 데이터 공유 원칙에 따를 뿐 아니라 관측자료 특성에 따른 정보 활용도를 높이고자 함이다.

국내 12개 지구 관측분야 메타데이터를 통합하기 위한 메타데이터의 설계는 관측분야 자료별 데이터들의 자료 공유와 공동 활용을 극대화한다. 다음의 과정을 통해 지구 관측 자료 공유를 위한 메타데이터의 데이터 요소를 도출한다.

- 국내외 지구 관측자료 관련 메타데이터 데이터 요소 세트 및 사례 조사
- 국내 12개 지구 관측분야별 메타데이터 데이터 요소 세트 조사 및 비교 분석
- 다양한 자료를 공유하기 위한 메타데이터 데이터 요소를 도출하기 위해 활용할 메타데이터 세트 선별
- 선별된 메타데이터 세트를 구성하는 데이터 요소와 속성을 나열하고 상호 매핑하여 공통 요소를 추출
- 추출된 메타데이터 공통 데이터 요소를 기준으로 12개 지구 관측분야 통합데이터의 데이터 필수요소 선별
- 공통 데이터 요소 이외 데이터 통합에 반드시 필요한 데이터 요소를 선별하고 추가
- 완성된 메타데이터 요소와 12개 지구 관측분야의 고유 메타데이터 데이터 요소를 상호 매핑하여 통합 메타데이터 데이터 요소 검증

## 4. 자료공유를 위한 KGEO 메타데이터

### 4.1 메타데이터 스키마

국내 지구 관측 자료 공유를 위해 MODS 메타데이터를 기반으로 설계한 “KGEO 메타데이터”는 메타데이터로서의 기능을 충실히 하면서 지구 관측자료를 원활히 통합할 수 있도록 데이터 요소를 선정하였다. 우선 지구 관측분야들의 다양한 메타데이터 스키마로부터 공통된 데이터 요소를 추출하고, 메타데이터로서의 고유 기

능을 유지할 수 있도록 메타데이터의 자료 정보, 기술 및 구조정보, 이용권한, 관리측면의 다방면을 기준으로 데이터 요소를 구분하여 메타데이터로서의 자격을 유지하도록 하였다.

KGEO 메타데이터는 관측분야별 다양한 메타데이터를 통합하기 위해 메타데이터를 기술하기 위한 메타데이터라는 관점에서 ‘메타-메타데이터’를 기반으로 메타데이터를 설계하였다. 따라서 데이터로서 메타데이터 정보를 기술하기 위한 데이터 요소 추가가 불가피하여, 관측분야의 메타데이터를 식별, 접근하기 위한 데이터 요소를 MODS 데이터 요소에 추가하였다.

### 4.2 메타데이터 요소

KGEO 메타데이터의 데이터 요소로서 도출된 요소는 전체 42개이다. 데이터 요소 중 8개는 MODS에서 제공하지 않는 것으로 원시메타데이터 기술을 위해 추가된 데이터 요소이다. KGEO 메타데이터는 관측자료로서 기본정보를 필수데이터 요소로 채택하였다. 기본적인 데이터 요소로서 관측자료 기준계(공간, 시간)와 식별정보를 포함하여 관측자료 분류(관측분야, 관측요소) 그리고 관측자료 원시메타데이터 관련 데이터 요소가 바로 그것이다.

지구 관측자료로서 자료가 관측된 위치 및 시간정보를 필수 데이터요소로 출발하여, 자료 출처정보, 관측분야 및 관측요소를 포함한 관측자료 분류정보, 메타데이터 식별정보를 필수 요소로 하였다. 이들 필수요소는 <표 4>에서도 보여주듯 범용 메타데이터들이 일반적으로 채택하는 데이터 요소를 근간으로 데이터 항목을 상호 매핑하여 선택한 것이다. 또한 국내 관측

자료 메타데이터 구조를 모두 검토한 결과 이들 데이터 요소를 모두 포함하고 있었다.

KGEO 메타데이터의 특징으로 원시메타데이터에 관한 정보를 필수 데이터요소로 선정하는 이유는 본 연구에서 설계하고자 하는 통합 메타데이터는 메타-메타데이터를 근간으로 하므로 원시 메타데이터를 접근하기 위한 정보를 제공할 필요가 있기 때문이다. KGEO 메타데이터는 메타데이터로서 관리적 기능을 충족시키고자 메타데이터를 관리하기 위한 관리정보 또한 필수요소로 지정하였다. MODS 메타데이터에서 필수요소로 선정되지 않은 데이터 요소는 지구 관측자료 메타데이터 특성에 따라 선택적으로 사용할 수 있도록 하였다.

KGEO 메타데이터는 세트는 <표 5>와 같다. 국내 모든 관측분야별 관측요소의 메타데이터에서 공통적으로 사용되는 데이터 요소로서의 미적으로 일치되도록 MODS 데이터 요소를 재구성한 것이다. KGEO 메타데이터는 MODS 메타데이터 계층구조를 활용하여 3계층으로 구조화하였다. titleInfo를 포함한 상위요소 16개, title을 포함한 하위요소 26개이다. 데이터 상위요소의 용도는 다음과 같다.

- TitleInfo: 표제정보로서 특정 관측분야 메타데이터의 제목에 해당하는 요소로 관측자료 내 포함하고 있는 정보 또는 그것을 명명하기 위한 일반적인 표현의 용어, 문구, 문자 또는 문장이다.
- Name: 관측자 정보로서 관측 자료와 어떤 방식으로든 연관되어 있는 인명, 조직 및 행사명, 자동 관측기기, 관측자유형 즉 사람, 학교, 연구기관, 공공기관 등의 역할 정보를 메타데이터 속성 type에 표기할

수 있다.

- Subject: 주제어 정보로서 관측자료에 관측 자료에 관한 주요한 중심내용을 기술하기 위한 단어 또는 어구를 기입한다. MARC21과 같이 서지정보 또는 문헌정보에서는 LSCH를 이용하여 주제어를 기술할 수 있는데, 기술하는 관측자료에 이러한 통제어가 있다면 메타데이터 속성 type에 기술할 수 있다.
- Abstract: 요약정보로서 관측 자료정보를 내용 요약한다. 자료 검색을 위하여 관측분야별 메타데이터의 주요 데이터요소를 발췌하여 데이터를 삽입할 수도 있다. 요약정보의 출처 및 역할은 메타데이터 속성 type에 기술할 수 있다.
- OriginInfo: 관측자료 출처정보로서 관측자료가 기원 또는 생성된 장소, 관측지, 관측일을 포함한 관측자료의 출현에 관련된 정보를 기입한다.
- Genre: 관측분야로서 국내 12개 지구 관측분야 중 자료에 해당하는 관측분야를 기입한다. '재해', '보건의료', '에너지자원', '기후', '수자원', '기상', '생태계/생물다양성', '농업', '산림', '해양', '우주', '국토공간 정보' 중 택일해야 한다.
- TypeOfResource: 관측요소로 관측자료의 일반적인 내용유형을 기술하는 데이터 요소로서 관측분야별 컬렉션 또는 관측요소를 기입한다. "생태계/생물다양성" 분야의 경우 '식생분포', '포유류 분포', '염생 식물' 등을 기술한다.
- Classification: 자료분류로서 관측분야별 또는 관측요소별 데이터의 자료 분류이다.

<표 5> 지구 관측자료 공유를 위한 KGEO 메타데이터 요소

그룹	상위 요소	하위요소	필수여부	내용
식별정보	TitleInfo	title	필수	제목
		name *		부제목
	Name *	namePart *		관측자 정보
	Subject *	topic *	권고	주제
	Abstract *			내용요약
발행사항	OriginInfo *	place	필수	관측지역 - 행정구역
		Publisher *		관측소
		dateIssued	필수	관측일시
		issuance		관측연속성
		frequency		관측주기
	geographic ^			관측지역 - 위도, 경도
자료분야	Genre		필수	관측분야
	typeOfResource		필수	관측요소
	Classification *			자료 분류
기술사항	Language	languageTerm		언어
	Identifier *			인식기호
이용권한	AccessCondition *		필수	사용권한
	TargetAudience			이용대상자
원시 메타데이터정보	Extension (관측분야별 메타데이터 연결정보)	metaContentSource ^	필수	원시메타데이터 작성기관
		metaRepository ^	필수	원시 메타데이터 레포지토리
		metaIdentifier ^	필수	원시 메타데이터 식별번호
		metaAccess ^	필수	원시 메타데이터 연결정보
		metaType ^	권고	원시 메타데이터 표준
		metaCreationDate ^	권고	원시 메타데이터 최종입력일
		metaChangeDate ^	필수	원시 메타데이터 최종입력일
원시자료 연결정보	PhysicalDescription*	form		원시 자료 자료형태
		note		원시 자료 설명
	Location *	physicalLocation	권고	원시 자료 소장처
		url		원시 자료 연결정보
레코드 정보	RecordInfo *	recordContentSource	필수	레코드 작성기관
		recordCreationDate	필수	레코드 최초입력일
		recordChangeDate	필수	레코드 최종입력일
		recordIdentifier	필수	레코드 식별번호

\*: 반복 사용가능 ^: MODS에 추가된 데이터 요소

주제영역에 따른 특성별 관측자료를 코드화 및 조직화하기 위한 정형화된 시스템을 적용하였을 때 관측자료 주제에 해당하는 정보 부류를 기술한다. 정형화된 표준 분류표에 의한 자료 분류를 정의해야

하는데 자료에 적용된 분류표는 메타데이터 속성 authority 에 기술한다.

- Language: 언어로서 자료 내용을 기술하는 위해 사용된 언어이다. 상위요소 Language는 하위요소가 1개이지만 MODS의 데이

터요소 설계에 따라 상위요소 Language 가 아닌 하위요소 languageTerm에 통합 메타데이터에 기술된 정보의 언어를 기술한다.

- Identifier: 인식기호로서 관측자료 개별을 구분하기 위해 유일한 표준번호나 코드를 기술한다. 보건의료 분야의 경우 관측요소 전염병의 경우 질병코드가 이에 해당한다. 문헌정보의 경우 인식기호가 발달하였는데 ISBN, ISSN, LCCN 등을 기술한다. 인식기호의 유형을 기술하고자 할 경우 메타데이터 속성 type을 사용한다.
- AccessCondition: 사용권한로서 관측자료 접근에 요구되는 제약조건으로써 접근권한 및 사용범위 등의 정보이다. 다양한 접근유형을 기술하고자 할 경우 메타데이터 속성 type을 사용한다.
- TargetAudience: 이용대상자로서 관측자료를 접근 및 이용 가능한 이용자유형 또는 그룹을 기술한다.
- Extension: 관측분야별 메타데이터 정보로 MODS의 데이터 요소 확장을 대비하여 설계된 것으로 지구 관측자료 통합 메타데이터에서는 관측분야별 메타데이터 관련정보를 간략히 기술하기 위해 사용한다. 통합 메타데이터를 메타-메타데이터라 지칭할 때 그 설명하고자 하는 대상 메타데이터의 형태, 속성 등의 설명적인 요소를 기술한다.
- PhysicalDescription: 관측자료 즉, 원시자료(Source data)에 관한정보로서 원시자료에 대한 정보를 기입한다. 하위요소 form은 제작형태를 관리한다. 생태계/생

물다양성 분야의 경우 GIS-DB이고, 국토 공간정보 분야의 경우 벡터파일, 라스터파일, Geotiff, Tiff, 텍스트 등이 기입된다.

- Location: 관측자료 출처정보로서 관측자료 즉 원시자료의 소장처 및 접근정보를 기술한다.
- RecordInfo: 레코드 정보로서 지구 관측자료 통합 메타데이터에 관한 정보를 기술한다.

#### 4.3 메타데이터 검증 및 확장방안

설계한 통합 메타데이터의 타당성을 검증하기 위하여 지구 관측자료 공유를 위해 개발된 통합 메타데이터 요소와 관측분야 12개 분야의 메타데이터 요소를 비교하였다. 관측분야별 메타데이터의 요소와 의미적 측면에 입각하여 데이터 요소를 비교 검증하였기 때문에 실제 관측분야별 메타데이터 실제 데이터 레코드와 비교 검토했을 때와 차이가 있을 수 있다. 그러나 관측분야별 메타데이터 레코드를 확보하는 어려움으로 인하여 지구관측 자료의 통합 설계에 실 데이터 검증은 포함되지 않았다.

메타데이터 데이터 요소 비교검증에서 상위요소 RecordInfo는 통합 메타데이터 관리를 위한 데이터요소이기 때문에 비교 검토에서 제외하였다. <표 6>은 12개 관측분야 가운데 4개 관측분야를 선별하여 주요 메타데이터 요소를 통합 메타데이터와 크로스 매핑하기 위한 참조테이블을 작성한 예제이다.

지구 관측자료 공유를 위한 메타데이터의 설계 근간이 된 MODS 메타데이터는 데이터요소 extent를 제공하여 메타데이터 확장성을 고

려하여 설계되었기 때문에 이들 데이터 요소를 활용한 메타데이터 확장이 가능하다. 그러나 메타데이터 데이터요소를 수정, 변경, 추가하는 행위는 매우 신중하게 진행해야하는 중대한 사항이다. 본 연구에서 설계한 KGEO 메타데이터는 관련한 모든 시스템의 설계 및 구축에 민감한 영향을 끼치게 됨으로 시스템 공학적 논점을 빌리지 않더라도 메타데이터 설계 및 감리는 매우 중요하다. 따라서 통합 메타데이터의 데이터 요소 변경 또는 추가 행위가 불가피할 경우 정보전문가, 시스템 설계자, 각 기관의 담당자 등 많은 관련자와의 긴밀한 협의를 통하여 진행해야 할 것이다.

#### 4.4 메타데이터 합의 과정

KGEO 메타데이터는 국내 전 지구 관측자료를 통합하는 메타데이터로 관측자료 및 관련 메타데이터를 생산 또는 구축하는 실무자 및 관련 전문가들과의 협의과정을 통해 확장할 필요가 있다. 메타데이터의 합의 과정은 다음과 같이 제시할 수 있다.

첫째, 앞서 설계된 KGEO 메타데이터를 실제 활용하기에 앞서 전문가들로 구성된 위원회를 조직하여 2~3차례 메타데이터 재심의를 거쳐 지구 관측자료 공유를 위한 메타데이터로 활용할 수 있다.

둘째, 통합 메타데이터에서 제시하는 메타데이터 필수요소가 각각의 관측분야 메타데이터에서 제공하지 않는 요소가 상당히 발견되기 때문에 해결방안 마련이 요구된다.

셋째, 자료 공유 및 교환상의 데이터의 의미적 불일치를 예방하기 위하여 관측분야별 자료

의 분류체계 및 코드체계를 재확인해야 한다.

넷째, 메타데이터 심의를 거친 후 다양한 관측분야, 관측요소의 관측 자료의 일정량을 샘플링하여 실제 본 메타데이터로 매핑하여 그 실효성과 적합성을 면밀히 검토해야 한다.

#### 4.5 분야별 메타데이터 통합방안

지구 관측자료의 분야별 메타데이터 통합 방안은 다음과 같이 제시될 수 있다. 첫째, 지구 관측자료 관련 기관 부처 및 정보전문가를 포함한 많은 관련자 및 실무담당자들의 충분한 의견 수렴과 적절한 적합성 검토를 통하여 12개 지구 관측분야의 다양한 관측 자료의 통합 방안을 모색하여야 할 것이다.

둘째, 국내 지구관측 자료의 메타데이터 현황에서 관측 자료를 분야별 관측요소별 구분하였을 때 관측 자료의 약 21%가 메타데이터를 정의되어있지 않으므로 이러한 자료에 대한 대책이 필요하다.

셋째, 앞서 기술한 관측자료 공유를 위한 메타데이터는 다양한 분야의 관측 자료의 기본적인 데이터 요소와 통합 관리에 필요한 데이터 요소를 정의하였기에 관측분야별, 관측요소별 메타데이터 구축의 필요성이 없다고 판단되는 관측 자료의 경우 이곳에 설계된 메타데이터를 활용 할 수도 있을 것이다.

넷째, 특정 관측분야별 관측요소별 자료의 메타데이터가 정의되어 있다고 하더라도 메타데이터를 정형화 및 표준화 노력을 기울여 고정된 정보 조직체계 구축이 필요시 된다.

다섯째, 관측분야별 자료 분류에 대한 일정한 분류체계를 정립함으로써 자료 공유 및 교환



<표 6> K GEO 메타데이터와 관측분야별 메타데이터 요소 매핑 예제

그룹	통합 메타데이터		보건의료	에너지자원	수자원	해 양
	필수	요소				
식별 사항	필수	제목	질병명	제목title	개인제공 정보명	DataSet제목
		부제목		시리즈 series		
		관측자정보			데이터셋 제작 기관명	
	권고	주제	신중여부	키워드 keyword	주제핵심어	
		내용요약		요약 abst	요약설명	DataSet요약
발행 사항	필수	관측지역-행정구역	시도명, 구군명, 우편번호	공간적범위 spatExt	서쪽경계좌표, 동쪽경계좌표	해역명
		관측소	보건소 요양기호		"건설교통부"	"국립해양조사원 .."
	필수	관측일시	진단일	시간간범위 temExt		개시일, 종료일
		관측연속성			시간주기	
		관측주기			시간주기	
	관측지역-위도..			서쪽경계좌표, 동쪽경계좌표	남단위도, 북단위도	
자료 분야	필수	관측분야	"보건의료분야"	"에너지자원분야"	"수자원분야"	"해양분야"
	필수	관측요소	"전염병"	"전체"	"전체"	DataSetID
		자료 분류	질명코드	주제분류 tpCat		
기술 사항		언어		메타데이터언어mLang		DataSet언어
		인식기호	질명코드			
이용 권한	필수	사용권한	"공개"	접근제약, 사용제약	"공개"	접근제한, 사용제한
		이용대상자		접근제약, 사용제약		접근제한, 사용제한
원시 메타 데이터	필수	분야별메타데이터 작성기관	"질병관리본부"	"한국지질자원연구원"	"건설교통부"	"국립해양조사원.."
	필수	분야별메타데이터 레포지토리	<임의부여>	데이터셋식별자 dataId	데이터셋 이름	<임의부여>
	필수	분야별메타데이터 식별번호	<임의부여>	식별 idInfo	식별테이블 외부키	metadata ID
	필수	분야별메타데이터 연결정보				
	권고	분야별메타데이터 표준		메타데이터표준명 mStName		
	권고	분야별메타데이터 최초입력일		메타데이터생성일자	메타데이터 작성일자	생성날짜
	필수	분야별메타데이터 최종입력일			메타데이터 검토일자	최종개정날짜
원시 자료		원 자료 자료형태			원시자료매체	
		원 자료 설명			데이터셋 원시자료 부여 정보	배포매체
	권고	원 자원 소장처				
		원 자원 연결정보				URL주소

상의 데이터의 의미적 불일치를 예방해야 한다.

여섯째, 12개 관측분야 116개 관측요소의 관측 자료는 자료의 양이 방대할 뿐 아니라 자료 형태, 자료유형이 매우 다양함으로 관측 자료를 생산 및 구축하는 기관에서 메타데이터 매핑 테이블을 통하여 자료를 공유하고 메타데이터를 데이터 교환 프로토콜을 통하여 공유 또는 통합하는 형태의 시스템 설계가 가장 적절할 것으로 사료된다.

## 5. 결론

국내 전 지역의 포괄적이고 지속적인 지구 관측자료는 내국민의 이익 창출뿐 아니라 전지구의 일원으로서 인류의 이익 창출을 가져올 것이다. 그러나 관측 분야별로 독자적으로 생산 또는 구축되는 관측 자료는 그 효용가치를 제한한다. 본 연구에서는 12개 지구 관측분야의 자료를 통합, 공유하기 위한 KGEO 메타데이터를 설계함으로써 분산된 관측 자료를 검색, 교환, 획득을 위한 기반을 마련할 수 있게 되었다.

KGEO 메타데이터는 MODS 메타데이터 계층구조를 활용하여 3계층으로 구조화하였다.

titleInfo를 포함한 상위요소 16개, title을 포함한 하위요소 26개이다. 지구 관측자료로서 자료가 관측된 위치 및 시간정보를 필수 데이터 요소로 출발하여, 자료 출처정보, 관측분야 및 관측요소를 포함한 관측자료 분류정보, 메타데이터 식별정보를 필수요소로 하였다. 또한 원시메타데이터 정보와 레코드 정보도 필수요소이며, MODS 메타데이터에서 필수요소로 선정되지 않은 데이터 요소는 지구 관측자료 메타데이터 특성에 따라 선택적으로 사용할 수 있도록 하였다.

메타데이터는 구조화된 정보로서 자원의 소재 파악, 식별, 기술, 접근과 관리를 지원하여 디지털 자원의 이용과 공유를 증진시키는 효과적인 도구이다. 따라서 국내 지구관측 자료의 효율적인 통합운영체계 구축을 통하여 자원의 공유와 활용, 재사용 등의 시너지 효과를 얻을 수 있다.

본 연구의 KGEO 메타데이터를 기반으로 구축될 자료는 지구 관측을 위한 인프라 구축의 가장 기본적인 핵심요소로 작용하여 기관 간, 학제 간 조정 및 자료 공유를 활성화하고 자료획득에서의 시간적, 공간적 격차를 해소하고 관측자료의 연속성을 보장하여 새로운 이익을 창출할 수 있을 것으로 기대된다.

## 참 고 문 헌

곽승진. 2004. 청소년 과학정보 메타데이터 요소 및 데이터베이스 구축 연구. 『한국문헌정보학회지』, 38(1): 263-279.  
안부영, 조희형, 안성수, 박형선. 2005. 생물다양

성 데이터교환을 위한 메타데이터 스키마 설계. 『정보과학회 학술대회논문집』.  
이수상. 2004. MODS(Metadata Object Description Schema - 디지털도서관의 서지정

- 보 표준 메타데이터. 『2004년도 KISTI 메타데이터 설계 및 적용사례 세미나 발표자료』.
- 송용철 2004. 『그리드데이터 메타데이터 내용모델에 관한 연구』. 석사학위논문, 인하대학교 대학원, 지리정보공학과.
- 조천호 등. 2005. 『전자구관측시스템 국가대응체제 구축 연구: 과학기술국제화사업 최종보고서』. 과학기술부.
- 한종엽, 최영준. 2004. 해양전문정보센터의 멀티미디어 메타데이터베이스 및 디지털 도서관 통합정보시스템 구현에 관한 연구. 『정보관리학회지』, 21(4): 5-26.
- Anne J. Gilliland. 2007. Introduction to Metadata-Pathways to Digital Information. Setting the Stage. [cited 2006.9.22]. <[http://www.getty.edu/research/conducting\\_research/standards/intrometadata/setting.html](http://www.getty.edu/research/conducting_research/standards/intrometadata/setting.html)>.
- Arun Sen. 2004. "Metadata Management: Past, Present and Future." *Decision Support Systems*, 37: 151-173.
- Caplan, Priscilla. 2003. *Metadata fundamentals for All Librarians*. Chicago: American Library Association.
- Caplan, Priscilla. 2004. 『메타데이터의 이해』. 오동근 역. 대구: 태일사.
- Christophe Blanchi and Jason Petrone. 2001. "Distributed Interoperable Metadata Registry." *D-Lib Magazine*, 7(12). [cited 2006.09.28]. <<http://www.dlib.org/dlib/december01/blanchi/12blanchi.html>>.
- Cliff Wootton, 2006. Developing Quality Metadata: Building Innovative Tools and Workflow Solutions. Focal Press.
- CODATA Working Group on Biological Collection Data Access Homepage. [cited 2006.10.02]. <<http://bgbm3.bgbm.fu-berlin.de/TD-WG/CODATA/Schema/default.htm>>.
- Content Standard for Digital Geospatial Metadata(CSDGM)-Federal Geographic Data Committee(FGDC). [cited 2006.9.23]. <<http://www.fgdc.gov/metadata/csdgm>>.
- David Marco and Michael Jennings. 2004. Universal Meta Data Models. Indianapolis: Wiley Publishing, Inc.
- Diane J. Hillman and Elaine L. Westbrook. 2004. Metadata in Practice. Chicago: American Library Association.
- DCMI. 2001. Dublin Core Metadata Glossary, Final draft [cited 2007.04.30 ]. <<http://dublincore.org/documents/usageguide/glossary.shtml>>
- Dublin Core Metadata Initiative Homepage. [cited 2007.05.02]. <<http://dublincore.org/>>.
- Guenther, Rebecca and McCallum, Sally. 2003. "New Metadata Standards for Digital Resources: MODS and METS." *Bulletin of the American Society for Information Science and Technology*, 29(2).
- Intner, Sheila S. Lazinger, Susan S. Weihs, Jean, 2005. Metadata and Its Impact

- on Libraries. Libraries Unltd Inc.
- J. Nogueras-Iso et al. 2004. "Metadata Standard Interoperability: Application in the Geographic Information Domain." *Computers, Environment and Urban Systems*,(28): 611-634.
- MODS Homepage. [cited 2006.12.15].  
<<http://www.loc.gov/standards/mods/v3/mods-userguide-3-0.html>>
- Open Archives Initiative Protocol for Metadata Harvesting. [cited 2006.10.15].  
<<http://www.openarchives.org/OAI/openarchivesprotocol.html>>
- Pundt, H. et al. 2002. "Domain Ontologies for Data Sharing. An Example from Environmental Monitoring Using Field GIS." *Computer & Geosciences*, (28): 95-102.
- Stephanie C. Haas et al. 2003. "Darwin And MARC: A Voyage of Metadata Discovery," *Library Collections, Acquisitions, & Technocal Service*, (27): 291-304.
- Tannenbaum, Adrienne. 2002. *Metadata Solutions: Using Metamodels, Repositories, XML, and Enterprise Portals to Generate Information on Demand*. Addison-Wesley.
- The Global Earth Observation System of Systems(GEOSS) 10-Year Implementation plan. [cited 2007.05.17].  
<<http://www.earthobservations.org/docs/10-Year%20Implementation%20Plan.pdf>>