International Journal of Database Theory and Application
Vol. 2, No. 1, March, 2009

Development of Metadata and Ontology Standardization Strategy for Korean Earth Observation Data

Bu Young Ahn, Min Su Joh, Hun Ju Myung
Korea Institute of Science and Technology Information
{ahnyoung, msjoh, hjmyung}@kisti.re.kr

Abstract

The frequency and scale of natural disasters such as typhoons, floods, earthquakes, and tidal waves from earthquakes have increased recently. Several nations have recognized that earth observation is essential for protecting the Earth's environment. Therefore, 50 nations around the world have agreed upon the construction of GEO (Group on Earth Observation), which is to be in charge of the earth observation in an effort to gain understanding of earth system changes as well as to monitor and predict these changes. However, the data format from earth observation varies depending on the areas, institutions, and countries. This presents a challenge when attempting to share and exchange data among different organizations. Thus, a metadata scheme and system of ontology suitable for domestic situations is developed in this study to facilitate exchanges of data regarding areas that benefit society with reference to the principles of data sharing and exchange regarding GEO. Also established is the KGEO metadata and ontology required to identify the metadata situation of earth observation data that is used for nine societal benefit areas of GEOSS (Global Earth Observation System of Systems).

1. Introduction

Against this backdrop, 50 nations around the world came to share an understanding that the increasing frequency and scale of natural disasters are closely related to changes in the earth environment such as global warming, and that the rapid changes in the earth environment could the survival of mankind. Based upon this consensus, the countries have established Group of Global Observation (GEO) which is to be in charge of earth observation in an effort to gain understanding of earth system changes as well as to monitor and predict these changes. Meanwhile, Global Earth Observation System of Systems (GEOSS), developed and operated by GEO, is designed to perform comprehensive, continued, coordinated observation of weather, climate, oceanography, land and ecology of the earth system, analyze the obtained data, make predictions, and then delivery useful final information to those who need it in a timely manner.

Because issues concerning the earth environment is influenced by domestic as well as international situations, Korea must lay a foundation for data collection, generation, storage, analysis and exchange to enhance the nexus between different systems operated by different governmental organizations or areas, and further increase the level of data utilization, in order to effectively respond to the GEOSS international cooperative project at the national level and to successfully carry out the GEOSS establishment project in Korea. In addition, in order to develop an integrated operation system of earth observation data systems developed and operated by different areas and organizations, the principles of data sharing and exchange discussed in GEO must be taken into consideration to ensure integrated data design that reflects domestic situations and to make specific plans for standardizing metadata and developing ontology among data to be exchanged between areas that benefit the society.

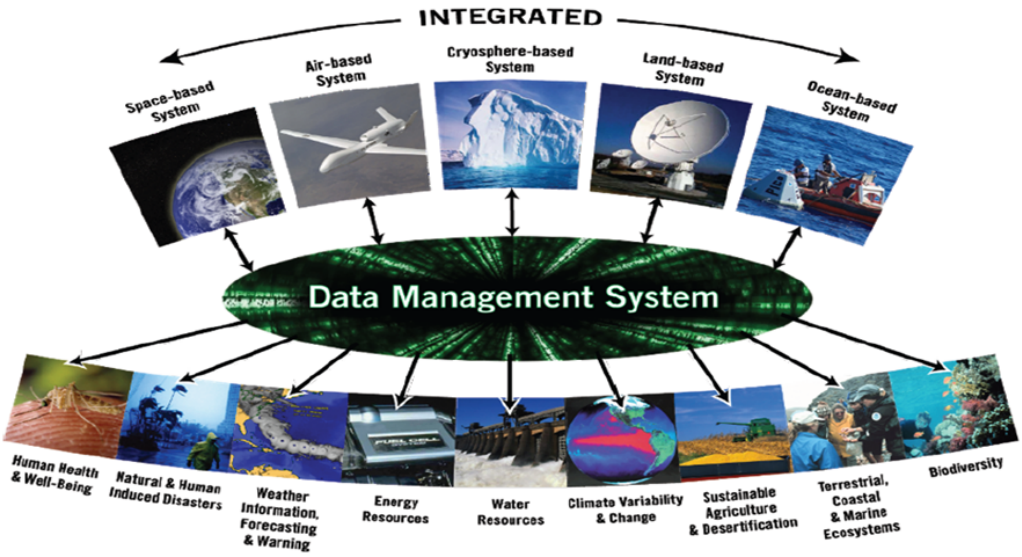

Figure 1. GEOSS Construct [1]

Therefore, in this study, the status of earth observation data used in nine societal benefit areas of GEOSS are examined and measures to design a system for standardizing metadata and ontology that is required to integrate area-specific, various earth observation data are suggested in order to develop an efficient, integrated operation system to utilize domestic earth observation data.

2. Survey and analysis of relevant data home and abroad

2.1 Survey of domestic earth observation data

In Korea, the office of KGEO was set up to effectively respond to GEOSS science and technology international cooperation at the national level as well as to successfully carry out the project to develop Korea's own earth observation system. To build a more constructive national earth observation system, a survey on current conditions of nine GEOSS benefit areas was conducted as part of this study.

The survey was carried out in two ways. First, a survey sheet containing questions concerning five major areas – standardization, operation and management of information system, status of existing raw data and metadata, status of infrastructure and other matters – was used for an on-site, face-to-face interview with 12 experts at nine different agencies that produce earth observation data. Second, data of various fields including meteorology, agriculture, ecosystem, biodiversity, disasters, water resources, health, climate, energy, forestry, the ocean and geographical information that were available on the web were investigated, collected, classified and analyzed.

The survey found that domestic organizations of each area are producing and utilizing a variety of data related to earth observation and that geographical and meteorological information is jointly utilized around the world based on established international standards. Metadata have been built as shown in Table 1, and the survey also indicated that a considerable amount of budget and human resources have been committed to building effective information systems such as meteorological/climate information system maintained by Korea Meteorological Administration, WAMIS and WINS in the water resources field, KBIF for the biodiversity field, and the NGIS in the area of spatial information. In addition, information systems for the areas of oceanography, disasters, ecosystem, and geological resources have been developed and in service. The findings of this survey will have great use as a basis for building a national GEOSS in the future.

Table 1. Status of metadata developed by each organization

Area	Name	Name of Agency
Biodiversity	- KBIF metadata	KISTI
Ecosystem	- Ecological nature status metadata	National Institute of Environmental Research
Health	- Contagious disease patient outbreak report statistics	Korea Centers for Disease Control and Prevention
Agriculture	- Soil map metadata	National Institute of Agricultural Science and Technology
Meteorology	- Meteorological satellite metadata - Satellite image metadata - Aerological observation metadata, etc.	Korea Meteorological Administration
Climate	- O_3 metadata	Korea Meteorological Administration
Water resources	- Basin survey metadata	Han River Flood Control Office
GIS	- Satellite image metadata - Digital topographic map & aerial photography metadata - GPS constant observation data, etc	National Geographic Information Institute
Oceanography	- Marine geographic information metadata - TOIS metadata	National Oceanographic Research Institute

2.2 Survey of international standard for earth observation metadata

Survey on international standards is essential in the effort to standardize domestic earth observation metadata. As such, items contained in ISO/TC211, ISO 19115 metadata standards and WMO Core Metadata Profile were examined.

International Standard Organization Technical Committee (ISO/TC 211) is an international standard organization specific to the geographical information area which came into being in June 1994 under the International Standard Organization (ISO) to establish information standard specifications regarding objects and phenomena that are either directly or indirectly related to the earth's geographical location. The organization's purpose is to enhance accessibility to and integration of geographical information by providing an environment that facilitates easy utilization of geographical information while realizing interoperation between computer systems dealing with geographical information. To expand ISO/TC211's work areas from the current vector-oriented fields to grid-centric geographical information in the future, Working Group 6 (WG6) handles imagery-related work as a sub-working group under ISO/TC211.

ISO 19115, which started with the Working Draft version 1.0 in March 1996, was established as the Committee Draft version 1.0 in July 1998. In June 2000, the Committee Draft version 3.0 was introduced to provide expandability of metadata by invalidating the then existing compliance levels and applying an object model of Unified Modeling Language (UML) instead, which is currently in use. Moreover, since the Committee Draft version 3.0, the composition of metadata sections has undergone drastic changes as standardization of areas related to geographical information took place in addition to metadata and as sections such as description list information, metadata expansion information and applied schema were added [2]. In September 2001, the Draft International Standard (DIS) version 1.0 was introduced, and in DIS, the data quality elements within the Data Quality Information section of the Committee Draft versions were subdivided in accordance with the data quality principles of ISO 19113. In December 2002, the Final Draft International Standard (FDIS) version was published, which was officially approved in March 2003.

The WMO Information System (WIS) led by the World Meteorological Congress (WMO) is an information system that will play a pivotal role in managing and transmitting weather, water resources and climate-related data around the world, and its design is based on the existing Global Telecommunication System (GTS). Heterogeneous data management systems of WMO programs failed to support compatibility among different data, and the environment of non-coordination has led to severe inefficiency requiring redundant efforts and costs, which eventually undermined exchanges of data and information among international user groups. In light of such problems, the Commission for Basic System (CBS) under WMO came to conceive WMO Information System (WIS). Major functional elements of WIS include National Centre (NC), Data Collection and Production Centre (DCPC), Global Information System Centre (GISC), and data networks that connect these elements. NC collects observation data by nation and transmits the collected information to DCPCs or GISCs connected to the nation involved. DCPC is responsible for producing and storing observation data, forecast results and processed information among WMO programs. GISCs, whose number will be 10 or smaller across the globe, will manage the list of metadata of managed data within WIS while providing tools for data search. As for the metadata schema used in WIS, it has been developed up to the WMO Core Version 1.0, and it comprises a total of 44 sections based on the metadata schema under ISO 19115 [3].

Table 2. WMO Core metadata Profile v.1.0

1	Metadata entity	16	Scope	31	Medium
2	Identification	17	Maintenance	32	Standard order process
3	Browse graphic	18	Scope description	33	Extended element
4	Keywords	19	Spatial representation	34	Application schema
5	Representative fraction	20	Dimension	35	Extent
6	Resolution	21	Geometric objects	36	Station
7	Usage	22	Reference System	37	Citation
8	Aggregation	23	Identifier	38	Responsible party
9	Constraint	24	Content Information	39	Address
10	Data quality	25	Range dimension	40	Contact
11	Lineage	26	Portrayal catalogue	41	Date
12	Process step	27	Distribution	42	Online resources
13	Source	28	Digital transfer options	43	Series
14	Data quality element	29	Distributor information	44	Telephone
15	Result	30	Format		

3. Measures to Develop Domestic Earth Observation Metadata

In order for Korea to effectively respond to international cooperative initiatives regarding GEOSS science and technology at the national level and to accomplish the development of Korea's own GEOSS successfully, it is imperative to build a foundation for data collection, production, storage, analysis and exchange designed to raise interoperation between systems that are being operated in each different area while further enhancing the level of data utilization. To that end, an organization like a national GEOSS integrated operation center should be founded to handle exchange of data related to GEOSS home and abroad while operating domestic earth observation systems in an integrated manner. Also needed are more specific measures to design an integrated operation system and to standardize metadata of data to be exchanged between societal benefit areas.

For standardization, OpenGIS catalog service specifications were surveyed and analyzed. The OpenGIS catalog service specifications are widely used in the communities related to geographical information, and are often cited as a promising candidate for GEO protocol and metadata specs. As such, it was believed that using these specifications to design KGEO's data exchange protocol and metadata contents would be a reasonable choice considering their compatibility with the GEO communities. Among OpenGIS catalog services, core catalog schema and message contents as well as GEOSS core architecture were examined centering on metadata [4].

GEOSS ADC Core Architecture comprises GEO Web portal, Clearinghouse and Registry, and it includes a process that allows registration, identification and use of services that can be accessed through GEOSS Interoperability Arrangements.

Meanwhile, GEOSS Core Architecture Implementation Report offers information on GEOSS Core Architecture testing, with its focus on 3 to 4 pilot-level GEO Web portals, Clearinghouse and Registry to provide evaluation details performed by IOC.

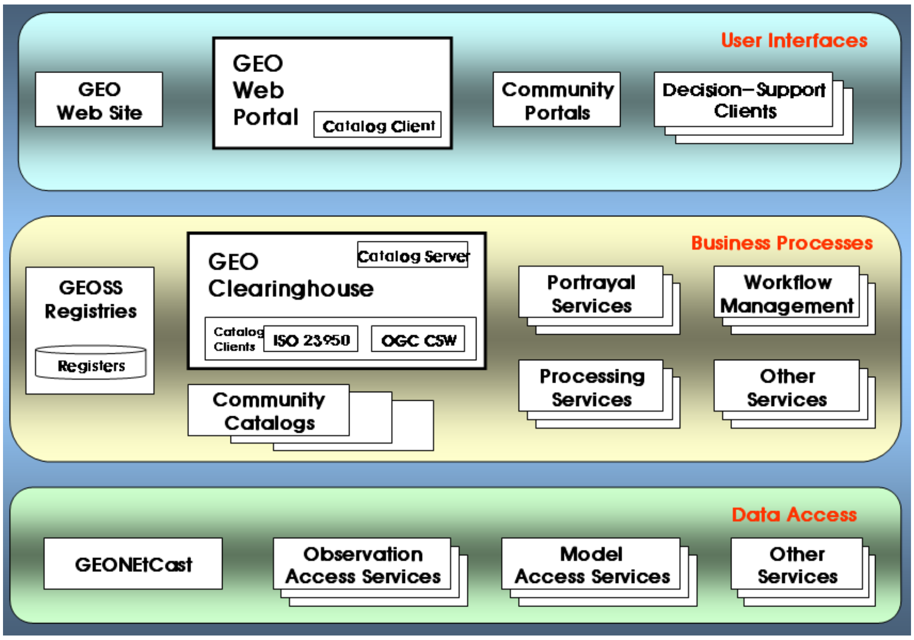

Fig. 2. GEO ADC(Architecture & Data Committee) Architecture - Engineering

The observations introduced in this report can be used as reference for the types of systems that should be included when KGEO designs metadata standards or system components, or when it cooperates with relevant domestic organizations in the future.

The following should be taken into consideration when establishing the KGEO metadata standard, tentatively named KGEO Profile, which is compatible with GEO metadata. First, KGEO Profile must support interoperation with metadata profiles home and abroad in order to be utilized in conjunction with data that are used in domestic as well as international communities. Considering the OpenGIS catalog service specifications and the activities by GEO architecture data committee, KGEO Profile must include additional elements specific to Korean situations based on CSW-Record. Second, KGEO Profile must support the domestic environment for metadata exchange including Korean language elements, in addition to the English environment which is used as a de facto universal language. In this paper, it was decided to use different names for English and Korean elements to facilitate the use of KGEO data both home and abroad. Third, KGEO Profile must be designed with its expansion in mind so that when the number of organizations participating in the KGEO community increases in the future, the users' needs can be reflected with ease. KGEO metadata standard draft (or KGEO Profile) was designed in consideration of the above requirements as shown in Table 3.

Table 3. KGEO metadata v.1.0

Element Name	Description
dc:title	Topic on resource's content
dc:creator	Major entities responsible for creating resource's content
dc:subject	Topic on resource's content
dct:abstract	Description on resource's content
dc:publisher	Major entities responsible for enabling use of resource
dc:contributor	Entities that contributed to resource's content
dct:modified	Date of creation or update of catalogued records
dc:type	Nature and genre of resource's content
dc:format	Physical or digital format of resource
dc:identifier	Sole reference symbol to identify record on catalog
dc:source	Information on the resource from which the current resource derived
dc:language	Language for intellectual record of catalogued record
dc:relation	Name of relation that exist among the resource described by code and relevant resources referenced by using *Source* or *dc:source* property
ows:BoundingBox	Name of relation that exist among the resource described by record and relevant resources referenced by using *Source* or *dc:source* property
dc:rights	Information on rights that belong to resource
kgeo:title	Korean description of dc:title
kgeo:creator	Korean description of dc:creator
kgeo:subject	Korean description of dc:subject
kgeo:abstract	Korean description of dct:abstract
kgeo:publisher	Korean description of dc:publisher
kgeo:contributor	Korean description of dc:contributor
kgeo:format	Korean description of dc:format
kgeo:source	Korean description of dc:source
kgeo:rights	Korean description of dc:rights

4. Measures to Build Domestic Earth Observation Ontology

The ontology of the geographic information area has various types of databases that were developed along with commercial growth of GIS, and a lot of information has been compiled by different countries and regions. As such, ontology database in the area of geographic information should be developed in such a way to enable conversion and storage of existing information into a type that references predefined ontology. Also, as studies are being carried out to express and store existing information through other types of ontology, pre-established information should be made easily available through mapping of other standard ontology.

The area of meteorology/climate requires expansion toward an ontology appropriate for WMO Core Profile standards. Observation data ontology of meteorology/climate data should be designed for query environment and access that can provide an integrated view from a unified, distributed environment Meteorological Centre (V-

GISC) for meteorological information, and it is desirable to conduct ontology modeling by expanding WMO Core Profile defined by OWL/RDF. Also, for integrated exploration and access with meteorological/climate data and other relevant data, measures to map with ontology that is highly relevant to meteorological and climate data should be studied and applied.

While information on the area of biodiversity has been compiled by many organizations, it either includes information for specific and finite purposes such as animals or crops, or has a tendency of being established separately as ecosystem-related information for relevant regions or nations. Therefore, ontology for the area of biodiversity should be organized in such a way to ensure easy integration with biodiversity information that is built in different forms from one another. To that end, ontology at a superior level is required to express diversity-related information of different fields and regions.

Because it is desirable to expand the area of water resources by re-using existing ontology and utilizing well-defined ontology, it is desirable to expand and further develop existing Upper Hydrologic Ontology. In addition, individual ontology for core areas of water resources including hydro-meteorology, basic, river, dam, underground water, mud, water supply, eco-environment, natural disaster and geo-space should be developed in addition to integrated ontology for water resources through migration and mapping of concepts of each core area, attributes and related information. Furthermore, other existing hydrologic unit ontology and mapping methodologies should be examined to ensure integrated scanning of and access to water resources management data through mapping and connection with hydrologic ontology.

Given the nature of agriculture, the ontology of this area must ensure sharing of region-specific information including geography and meteorology to support information sharing between nations, and it must also include a massive amount of genetic information to cover information on seed improvements researched and generated by different research institutes, while making it flexible so that it can address further improvements. Therefore, it is a better idea to develop a multiple number of distributed ontology for different areas and organizations, and map them as needed, instead of building single integrated ontology. To that end, a system to generate ontology which can extract, convert and store individual ontology elements from existing soil information should be developed.

5. Measures to Build Domestic Earth Observation Ontology

In order to develop an integrated operation system for earth observation data which are generated and serviced in/by different areas and organizations, specific measures must be put in place to develop a data integration design that meets the domestic needs, to standardize metadata to be exchanged between societal benefit areas and to establish ontology, by referencing the principles for data sharing and exchange being discussed at GEO.

To that end, the status of earth observation data being utilized by 9 societal benefit areas under GEOSS was examined and measures to standardize metadata and ontology needed to integrate diverse earth observation data of each area were suggested, in order to establish an efficient integrated operation system for domestic earth observation data.

For metadata standardization to build national GEOSS, it is first and foremost needed to organize KGEO metadata profile that offers interoperability with overseas metadata profile through designing and expanding components of KGEO integrated metadata. Secondly, a committee to be in charge of integrated metadata standardization must be set up and operated to facilitate consultation and version management on core metadata elements. Thirdly, a strategy to develop and distribute KGEO profile adaptor software must be developed in order to support exchange of data operated by several domestic communities in the KGEO profile format. In addition, it is also necessary to investigate existing databases and develop measures to convert them, and to build new databases as well as integrated metadata databases and networks.

In order to develop ontology for the national GEOSS, ontology models for individual areas such as geographical information, biodiversity, agriculture, meteorology, climate and water resources must be surveyed. Second, it is desirable to use pre-established ontology databases in the areas of geographical information, biodiversity and agriculture as reference. Third, for areas such as meteorology, climate and water resources, ontology development support tools, integrated management tools and data language for mass ontology handling (addition, deletion, refreshing and inquiry) should be developed to establish ways to develop new ontology databases.

It is anticipated that the measures to standardize metadata and develop ontology suggested in this study will serve as the most fundamental core element of the infrastructure for earth observation to facilitate coordination, data sharing and exchange amongst different ministries, organizations and areas, while addressing the time and space gap in obtaining new data and ensuring continuity of earth observation data, which will lead to creation of new synergy.

10. References

[1]Open Geospartial Consortium Network, GEOSS Architecture and Open Interoperrability, http://www.ogcnetwork.net/node/289#session_4.

[2]ISO 19115, http://www.isotc211.org .

[3]WMO Core Metadata Profile version 3.0 of the ISO Metadata Standard, 2006. http://www.wmo.ch/pages/prog/www/WIS-Web/Metadata.html

[4]OpenGIS® Catalogue Services Specification, OGC Implementation Specification, Editors: Douglas Nebert, Arliss Whiteside, Panagiotis (Peter) Vretanos, Date: 2007-02-23, OGC Document Number 07-006r1, Version 2.0.2, Corrigendum 2, http://www.opengeospatial.org/standards/cat

[5]Foster, I., Kesselman, C., Nick, J., Tuecke, S.: The Physiology of the Grid: an Open Grid Services Architecture for Distributed Systems Integration. Technical report, Global Grid Forum (2002)

[6]Open Geospartial Consortium Network, GEOSS Architecture and Open Interoperability, http://www.ogcnetwork.net/node/289#session_4

[7]WMO Information System(WIS). 2007. WMO WIS Coordination Office. http://www.wmo.int/pages/prog/www/WIS-Web/ What is WIS.html