Registration and retrieval of document centric metadata for accelerating integrated use of earth environmental data

Hiroko Kinutani[1*], Toshiyuki Shimizu[2], Masatoshi Yishikawa[2], Masafumi Ono[1],
Kazuyo Fukuda[3], Mika Ichino[3], Haruko Kawamoto[3]

[1]The University of Tokyo, [2]Kyoto University, [3]JAMSTEC

DIAS (Data Integration and Analysis System) is one of GEOSS activities in Japan. It is also a leading part of the GEOSS task with the same name defined in GEOSS Ten Year Implementation Plan. The main mission of DIAS is to construct data infrastructure that can effectively integrate earth environmental data such as observation data, numerical model outputs, and socio-economic data provided from the fields of climate, water cycle, ecosystem, ocean, biodiversity and agriculture. Some of DIAS's data products are available at the following web site of http://www.jamstec.go.jp/e/medid/dias.

Most of earth environmental data commonly have spatial and temporal attributes such as the covering geographic scope or the created date. The metadata standards including these common attributes are published by the geographic information technical committee (TC211) in ISO (the International Organization for Standardization) as specifications of ISO 19115:2003 and 19139:2007. Accordingly, DIAS metadata is developed with basing on ISO/TC211 metadata standards. From the viewpoint of data users, metadata is useful not only for data retrieval and analysis but also for interoperability and information sharing among experts, beginners and nonprofessionals.

On the other hand, from the viewpoint of data providers, two problems were pointed out after discussions. One is that data providers prefer to minimize another tasks and spending time for creating metadata. Another is that data providers want to manage and publish documents to explain their data sets more comprehensively. Because of solving these problems, we have been developing a document centric metadata registration tool.

The features of our tool are that the generated documents are available instantly and there is no extra cost for data providers to generate metadata. Also, this tool is developed as a Web application. So, this tool does not demand any software for data providers if they have a web-browser. The interface of the tool provides the section titles of the documents and by filling out the content of each section, the documents for the data sets are automatically published in PDF and HTML format. Furthermore, the metadata XML file which is compliant with ISO19115 and ISO19139 is created at the same moment. The generated metadata are managed in the metadata database of the DIAS project, and will be used in various ISO19139 compliant metadata management tools, such as GeoNetwork.

Because the number and kind of datasets stored in DIAS core system increase, we need the means to get a quick overview of the current state of our datasets. As an application of metadata created by our metadata registration tool, we are also developing a dataset search and cataloging tool providing users several matrices tables. We can search datasets based on keywords, spatial conditions, and temporal conditions. We can also access dataset documents through matrices, in which datasets are classified into some categories based on such criteria as GCMD Science

Keywords or GEOSS SBAs. By using this dataset cataloging tool, we can get an overview of all datasets of DIAS. Furthermore, we can dynamically change classification criteria and get various overviews from different viewpoints.

Keywords: DIAS, metadata, data integration and fusion, search, data access matrix