ACG032-P06   Room:Convention Hall   Time:May 27 16:15-18:45

GeoScience Data Release in DIAS

Hiroko Kinutani[1*], Toshiyuki SHIMIZU[2], Masatoshi YOSHIKAWA[2], Masaru KITSUREGAWA[1], Toshio KOIKE[1]

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

Amount of earth science data use by geoscientist, such as satellite observation data, sensor data or the result of simulation, is rapidly growing in the era of the data intensive science. Our project, DIAS (Data Integration and Analysis System) started in 2006, has a purpose of constructing data infrastructure that can integrate earth observation data, numerical model outputs, and socio-economic data effectively. DIAS also has a purpose to create knowledge enabling us to solve earth environment problems and to generate socio-economic benefits. The number of researchers of DIAS exceeds 150 and they participate from multi-disciplinary research fields; hydrological cycle, weather, ocean, agriculture, biodiversity, ecosystem, information science, etc.

Four years has elapsed from the start of the development of DIAS, and various kinds of geosciences data has been accumulated in DIAS.

We have decided to release these data, for converting them into more useful information and creating additional scientific or social value. And we, data providers, multiple disciplines' researchers, and system developers, have discussed many times for developing the systems for data release, and finally in last October, we released data to public.

The data in DIAS is classified into 4 categories:

1. Datasets created by DIAS researchers,

2. Datasets created at antecedent projects or related projects,

3. Mirror-data (research purpose satellite data, research purpose model output data, and research purpose Metrological data), And

4. Working datasets created by DIAS researchers.

The target categories of data release are 1, 2, and 3.

The applicable fields of datasets are agriculture, biodiversity, climate, disaster, ecosystem, energy, hydrological cycle, weather, etc.

In order to use data in different fields from data providers' research field, it is important to provide not only data itself but also documents written by data providers based on their knowledge.

We have developed a document centric metadata registration system.

We asked data providers to create metadata and document describing the dataset to be released using this system.

We asked to create 2 kinds of metadata and document, Japanese and English. That is, Japanese document about overseas' dataset, or English document about Japanese dataset is available. This is our outreach effort for both Japanese and overseas people.

We have discussed about the following issues carefully:

1. The unit of dataset to manage and retrieve, and

2. Representation in both Japanese and English.

We have decided about "what is the unit of dataset" through discussion between data providers and system developers with actual data one by one.

As a result, there are datasets composed of millions of files, and at the same time there are datasets composed a few files.

However, each file in the same dataset has the same dataset agreement and reference agreement.

And user access right is also managed by the unit of dataset.

We have developed our systems, such as search system and file download system, with both English and Japanese interfaces.

The functions necessary for data release are:

1. Seamless operation among data look down, data discovery and data download.

2. User management including new user registration

3. Access right management for each user and dataset

4. Report of data download history to data provider

To achieve these functions, we have developed the user management system, the DIAS look down and discovery system, and the access management system.

Anyone can use the DIAS look down and discovery system by accessing http://dias-dss.tkl.iis.u-tokyo.ac.jp/ddc/finder, and can download data files through the system. User registration is required before file download.