

## The DIAS data release and its cross-disciplinary usage

KINUTANI, Hiroko<sup>1\*</sup>, Toshiyuki Shimizu<sup>2</sup>, Toshihiro Nemoto<sup>1</sup>, Masatoshi Yoshikawa<sup>2</sup>, Masaru Kitsuregawa<sup>1</sup>, KOIKE, Toshio<sup>1</sup>

<sup>1</sup>The University of Tokyo, <sup>2</sup>Kyoto University

For 5 years from 2006FY, by collaboration of information science technologies and various kinds of earth environmental technologies, we have constructed a data infrastructure of the Data Integration and Analysis System (DIAS). DIAS is one of a project of a global-scale integrated observation and monitoring system (the Earth Observation and Ocean Exploration System), which is a part of national basic 'Earth Observation Promotion Strategy'. In DIAS we have integrated the earth observation data, numerical model outputs, socio-economical data and information in order to create knowledge enabling us to solve earth environment problems. From 2011FY, we started 'a program of earth environment information integration and fusion'. We continue to extend and enhance the DIAS functions. As our purpose of this program, we will construct an information infrastructure by collaboration of multi-disciplinary users integrating of extreme big data in order to create new value.

From October 2010, we have released data of DIAS with 'Document-metadata', describing about dataset in English and Japanese. Anyone can use the DIAS data discovery system by accessing <http://dias-dss.tkl.iis.u-tokyo.ac.jp/ddc/finder>, and can download data files of 176 datasets through the system. User registration and data policy agreement is required before file download.

The data in DIAS is classified into 4 categories:

- 1) Numerical simulation outputs for the purpose of research,
- 2) Satellite data for the purpose of research,
- 3) Datasets created by DIAS researchers,
- 4) Datasets created at related projects supported by DIAS.

The main datasets are 1) model outputs of CEOP, JP10, JRA25 and K-1, 2) CEOP Satellite data, CZCS SeaWiFS Chlorophyll data, 3) Mirai CTD data, Ocean Reanalysis data, Global map of interannual response of normalized difference vegetation index (NDVI), Triton buoy data, 4) CEOP In-situ data, AWCI In-situ data, GPV data, and so on.

The DIAS has provided the computation place enabling seamless usage of these dataset for researchers participating in this project, and they have gained many research outputs combining Satellite data, numerical model outputs and In-situ data.

On the other hand, for 1.5 year data release period, there were 50 person new user registrations and data downloading 300 times. The frequent download datasets are AWCI In-situ data, Global map of interannual response of normalized difference vegetation index (NDVI), Mirai CTD data and JRA25. Interest of released data of DIAS does not always motivate user to download data. Therefore the number of top page viewing of the DIAS data discovery system is about 4,500 times, and the number of document-metadata viewing is about 12,000.

We provide to display the distribution matrix of the dataset in the DIAS data discovery system in order to enhance cross-disciplinary usage of its data by specifying the axis GCMD Science keyword, area social benefits GEOSS, the GCMD data set as a platform. We found that the vast majority of users select the cell matrix instead of input search keywords in order to search the list to find the appropriate data.

Generally, in the catalogue search of the data center, the search system often provide functions to specify, period, region and keywords to users. Users can narrow down (or exclude) the search results using facet search function. In DIAS, as for increasing the number of released datasets in future, for improving our search function, we think we should add another axis of matrix of datasets, add facet search function, and add search results ranking.

We think adding these functions are necessary for cross-disciplinary dataset search. In order to implement these functions, we need to describe many aspects of attributes of datasets. We will plan to enhance our developed document-centric metadata creation system and the data downloading functions.

Keywords: DIAS, Release of Geoscience data, cross-disciplinary usage, Satellite data, Model output data, In-situ data