

# Data and Metadata Management at DIAS: Toward More Open Earth Environmental Information Platform

Toshiyuki Shimizu

Graduate School of Informatics, Kyoto University

tshimizu@i.kyoto-u.ac.jp

Dec. 7<sup>th</sup>, 2017

International Workshop on Sharing, Citation and Publication of Scientific Data across Disciplines  
Tachikawa, Tokyo, Japan

# Contents

## ❖ About DIAS

## ❖ Data and Metadata Management

- ❖ Data registration procedure
- ❖ Metadata management

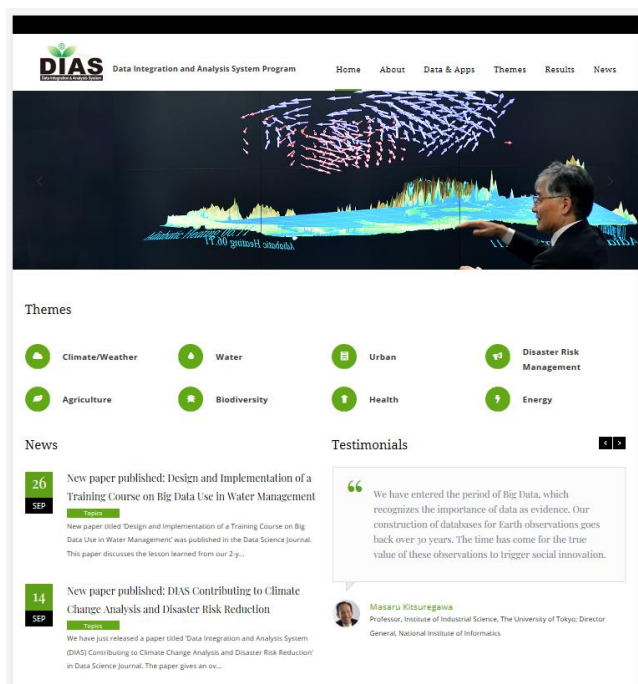
## ❖ Open Science Activities

## ❖ Current and Future Prospects

- ❖ DIAS as a national repository
- ❖ Focusing on metadata quality

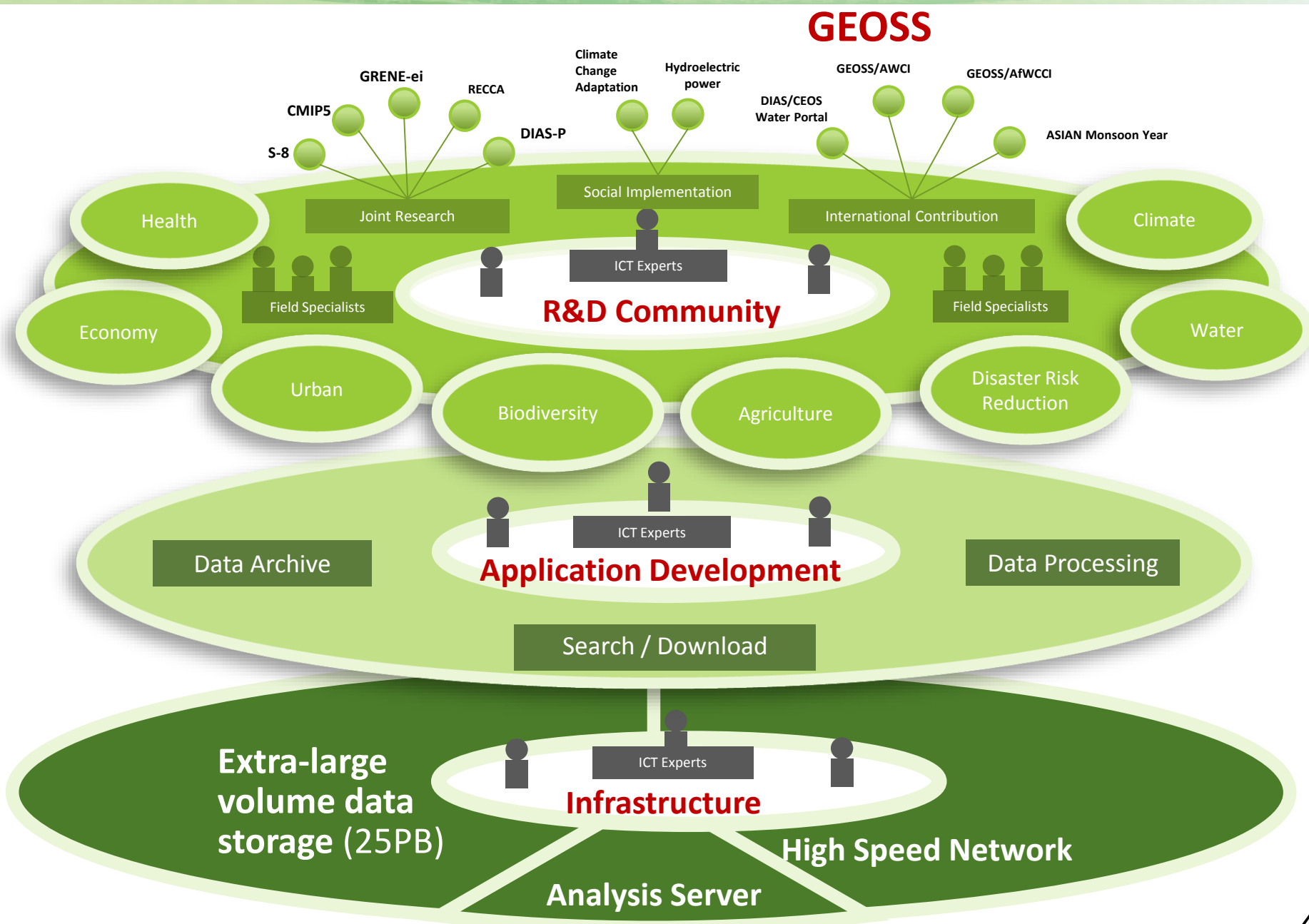
# DIAS (Data Integration and Analysis System)

- ❖ DIAS continuously collected and managed earth observation data.
- ❖ The first phase of DIAS started from 2006, and now we are in the third phase (2016-2020).



## Topics of Datasets available in DIAS

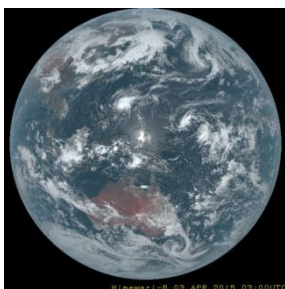
- ❖ Earth Observation Satellites
- ❖ Greenhouse Gases Observations
- ❖ Terrestrial Ecosystems / Carbon Flux Observations
- ❖ Weather Observations
- ❖ Watershed Observations
- ❖ Ocean Observations
- ❖ Reanalysis
- ❖ Prediction
- ❖ Downscaled Data
- ❖ Natural Disasters
- ❖ Land Use
- ❖ Health Hazard



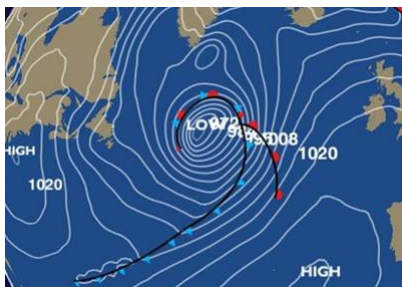
# Various Applications

[http://www.diasjp.net/en/apps\\_search/](http://www.diasjp.net/en/apps_search/)

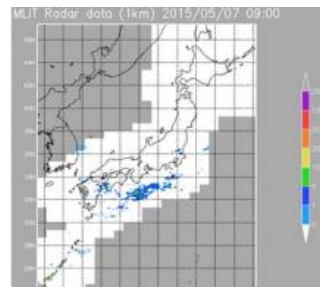
## Data Dissemination



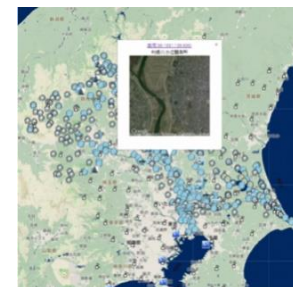
Himawari-8 Satellite



Weather Forecast GPV

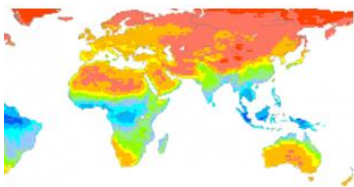


Radar Data

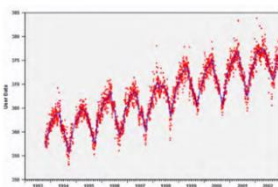


River Telemeters

## 1. Climate

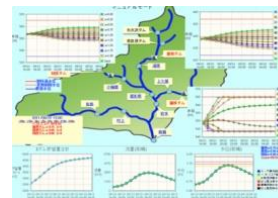


CMIP5 Model



Visualization Tools

## 2. Water



Dam Control



Water Management

## 3. Agriculture

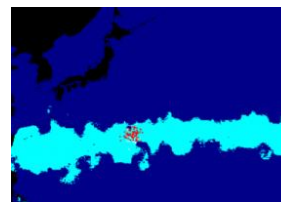


Potential of Rice Crops  
after Climate Change

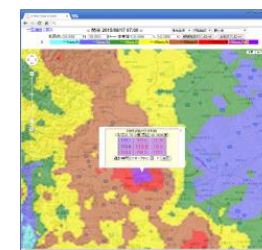
## 4. Biodiversity



Citizen science-based  
observations



Fish eggs and growth  
distribution



Accumulated  
radar rainfall

# Contents

## ❖ About DIAS

## ❖ Data and Metadata Management

- ❖ Data registration procedure
- ❖ Metadata management

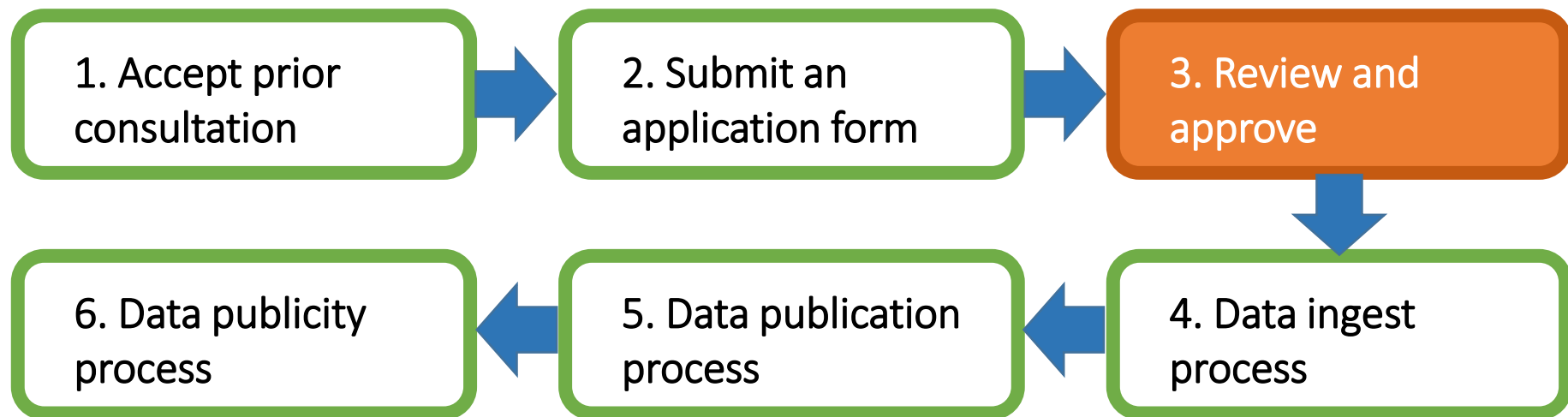
## ❖ Open Science Activities

## ❖ Current and Future Prospects

- ❖ DIAS as a national repository
- ❖ Focusing on metadata quality



# Data Deposit Workflow



- ❖ The applications will be reviewed from the viewpoints of value of the data itself, compatibility with DIAS, etc.
- ❖ You can consult with DIAS Office [dias-office@diasjp.net](mailto:dias-office@diasjp.net) about the data deposit.

# DIAS Metadata

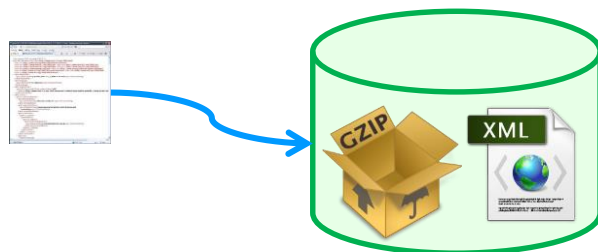
- ❖ We are managing various datasets in DIAS
- ❖ Basic strategy:

Make **dataset-level metadata** in the **common format** for **all datasets** stored in DIAS

- ❖ The granularity of dataset is decided by the data provider

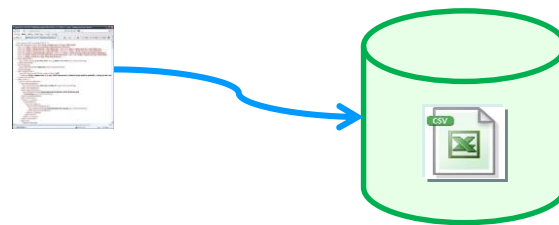
## Examples of datasets

CEOP Satellite Datasets  
(TRMM > PR > 3PRECI)



2,694 files (gz, xml, etc.)

Bombus terrestris and  
native bumblebee monitoring



5 files (csv)



# DIAS Metadata (cont.)

- ❖ Adopt the XML metadata used in geographic information system **ISO19115 (ISO19139)**
- ❖ We have developed web-based metadata registration tool
- ❖ Once metadata is created, documents for the dataset is automatically generated in **HTML** and **PDF** (**document-metadata**)

```
<?xml version="1.0" encoding="UTF-8"?>
<gmd:MD_Metadata xmlns:schemaLocation="http://www.isotc211.org/2005/gmd
http://www.isotc211.org/2005/gmd/gmd.xsd" xmlns:schema="http://www.w3.org/2001/XMLSchema"
xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:fn="http://www.w3.org/2005/02/xpath-
functions" xmlns:xd="http://www.w3.org/2005/02/xpath-datatypes"
xmlns:gml="http://www.opengis.net/gml" xmlns:gco="http://www.isotc211.org/2005/gco"
xmlns:gts="http://www.isotc211.org/2005/gts" xmlns:gss="http://www.isotc211.org/2005/gss"
xmlns:gsr="http://www.isotc211.org/2005/gsr" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xmlns:gmd="http://www.isotc211.org/2005/gmd">
  <gmd:fileIdentifier>
    <gco:CharacterString>MAHAPGP20160707104136-DIAS20160706142617-
    en</gco:CharacterString>
  </gmd:fileIdentifier>
  <gmd:language>
    <gco:CharacterString>English</gco:CharacterString>
  </gmd:language>
  <gmd:characterSet>
    <gmd:MD_CharacterSetCode
      codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml#MD_Charact
      codeListValue="utf8"/>
  </gmd:characterSet>
  <gmd:contact>
    <gmd:CI_Contact>
      <gmd:individualName>
        <gco:CharacterString>Koiti Masuda</gco:CharacterString>
      </gmd:individualName>
      <gmd:organisationName>
        <gco:CharacterString>Japan Agency for Marine-Earth Science and
        Technology</gco:CharacterString>
      </gmd:organisationName>
      <gmd:contactInfo>
        <gmd:CI_Contact>
          <gmd:address>
            <gmd:CI_Address>
              <gmd:electronicMailAddress>
                <gco:CharacterString>masuda@jamstec.go.jp
              </gmd:electronicMailAddress>
            </gmd:CI_Address>
          </gmd:address>
        </gmd:CI_Contact>
      </gmd:contactInfo>
    </gmd:CI_Contact>
  </gmd:contact>
</gmd:MD_Metadata>
```

XML metadata  
(ISO19115 (ISO19139))

DIAS MAHASRI Pathfinder Gridded Precipitation Data of the Asian Region

1. TITLE

|                     |   |
|---------------------|---|
| Name                | MAHASRI Pathfinder Gridded Precipitation Data of the Asian Region |
| Edition             | data version 1.1 - document 1.0en                                 |
| Abbreviation        | MAHAPGP1.1_doc1.0en   |
| Metadata Identifier | MAHAPGP20160707104136-DIAS20160706142617-en                       |

2. CONTACT

2.1 CONTACT on DATASET

|              |   |
|--------------|---|
| Name         | Koiti Masuda  |
| Organization | Japan Agency for Marine-Earth Science and Technology                  |
| Address      | 3173-25 Showa-machi, Kanazawa-ku, Yokohama, Kanagawa, 236-0001, Japan |
| TEL          | +81-45-778-5538   |
| FAX          | +81-45-778-5706   |
| E-mail       | masuda@jamstec.go.jp  |

2.2 CONTACT on PROJECT

2.2.1 Data Integration and Analysis System

|      |             |
|------|-------------|
| Name | DIAS Office |
|------|-------------|

HTML document

DIAS MAHASRI Pathfinder Gridded Precipitation Data of the Asian Region

1. TITLE

|                     |   |
|---------------------|---|
| Name                | MAHASRI Pathfinder Gridded Precipitation Data of the Asian Region |
| Edition             | data version 1.1 - document 1.0en                                 |
| Abbreviation        | MAHAPGP1.1_doc1.0en   |
| Metadata Identifier | MAHAPGP20160707104136-DIAS20160706142617-en                       |

2. CONTACT

2.1 CONTACT on DATASET

|              |   |
|--------------|---|
| Name         | Koiti Masuda  |
| Organization | Japan Agency for Marine-Earth Science and Technology                  |
| Address      | 3173-25 Showa-machi, Kanazawa-ku, Yokohama, Kanagawa, 236-0001, Japan |
| TEL          | +81-45-778-5538   |
| FAX          | +81-45-778-5706   |
| E-mail       | masuda@jamstec.go.jp  |

2.2 CONTACT on PROJECT

2.2.1 Data Integration and Analysis System

|      |             |
|------|-------------|
| Name | DIAS Office |
|------|-------------|

PDF document

# An Example of Metadata

## TITLE

|                     |   |
|---------------------|---|
| Name                | MIRAI CTD dataset                             |
| Metadata Identifier | MIRAI_CTD20160706144049-DIAS20160706142617-en |

## CONTACT

### CONTACT on DATASET

|              |  |
|--------------|--|
| Name         | Data Management Office   |
| Organization | Japan Agency for Marine-Earth Science and Technology                     |
| Address      | 3173-25 Showa-machi, Kanazawa-ku, Yokohama,<br>Kanagawa, 236-0001, Japan |
| E-mail       | dmo at jamstec.go.jp   |

### CONTACT on PROJECT

#### Data Integration and Analysis System

|              |  |
|--------------|--|
| Name         | DIAS Office  |
| Organization | The Earth Observation Data Integration and Fusion<br>Research Initiative (EDITORIA), The University of Tokyo |
| Address      | 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505, Japan  |
| E-mail       | dias-office@diasjp.net   |

## DOCUMENT AUTHOR

|              |  |
|--------------|--|
| Name         | Data Management Office                               |
| Organization | Japan Agency for Marine-Earth Science and Technology |
| E-mail       | dmo at jamstec.go.jp                                 |

## DATASET CREATOR

|              |  |
|--------------|--|
| Name         | Data Management Office                               |
| Organization | Japan Agency for Marine-Earth Science and Technology |
| E-mail       | dmo at jamstec.go.jp                                 |

## DATE OF THIS DOCUMENT

2016-07-06

## DATE OF DATASET

- publication : 2010-01-13

## DATASET OVERVIEW

### Abstract

This data set is vertical profiles of water temperature and salinity observed by CTD on research vessel MIRAI operated by JAMSTEC. This data will be utilized to analyze the characteristics of seawater and current.

### Topic Category(ISO19139)

- oceans

### Temporal Extent

|            |            |
|------------|------------|
| Begin Date | 1998-11-03 |
| End Date   | 2008-02-10 |

“MIRAI CTD dataset”

[http://search.diasjp.net/en/dataset/MIRAI\\_CTD](http://search.diasjp.net/en/dataset/MIRAI_CTD)









# An Example of Metadata (cont.)

## Geographic Bounding Box

|                      |      |
|----------------------|------|
| North bound latitude | 80   |
| West bound longitude | -180 |
| Eastbound longitude  | 180  |
| South bound latitude | -60  |

## Keywords

### Keywords on Dataset

| Keyword Type | Keyword  | Keyword thesaurus Name |
|--------------|--|------------------------|
| theme        | [Oceans  ] > [Ocean Temperature  ,<br>[Oceans  ] > [Salinity/Density  ] ><br>[Salinity  , [Oceans  ] > [Ocean<br>Chemistry  ] > [Oxygen  | GCMD_science           |
| theme        | OCEANOGRAPHY PHYSICAL  | AGU                    |
| theme        | Climate  | GEOSS                  |
| place        | INDIAN OCEAN,PACIFIC OCEAN,WESTERN<br>PACIFIC OCEAN  | others                 |

- Data Research System for Whole Cruise Information (DARWIN) :  
<http://www.godac.jamstec.go.jp/darwin/e>
- file download : <http://dias-dss.tkl.iis.u-tokyo.ac.jp/dl/storages/filelist/dataset:5>

## Distribution Information

| name | version        | specification |
|------|----------------|---------------|
| CSV  | no information |               |

## DATA REMARKS

Quality levels of data are described on  
<http://www.godac.jamstec.go.jp/darwin/explain/1/e#valueLevel>.

## USE CONSTRAINTS

### Data Policy by Data Provider

Please refer to the data policy at  
[http://www.jamstec.go.jp/e/database/data\\_policy.html](http://www.jamstec.go.jp/e/database/data_policy.html) and the terms and conditions at <http://www.jamstec.go.jp/cruisedata/e/readme.html>

### Data Policy for Project

#### Data Integration and Analysis System

1. With regard to Data policy, if there is any data policy indicated by the data provider, that policy always has priority over this DIAS data policy.
2. DIAS data sets are to be used only for scientific research or educational purposes [\*1].
3. Any modification or change of the original DIAS data sets is prohibited.
4. Any Re-export or transfer of the original data sets to a third party is

“MIRAI CTD dataset”

[http://search.diasjp.net/en/dataset/MIRAI\\_CTD](http://search.diasjp.net/en/dataset/MIRAI_CTD)

# DIAS Metadata Management System

- ❖ A Web Application.  
The system manage the registered metadata at the server side.
- ❖ Metadata input person using this system does not need to be aware of the XML.
- ❖ There are minimum required fields specified by the metadata schema, and recommended fields by the DIAS.

The screenshot displays three overlapping windows of the DIAS Metadata Management System.

**Dataset List Window:** Shows a table of datasets. The 'Assigned' column indicates the user assigned to the dataset. The 'Others' section lists datasets assigned to other users.

| 番号 | Dataset ID                  | Dataset Name(ja)                                |
|----|-----------------------------|---|
| 1  | MIRAI_CTD                   | みらいCTDデータセット                                    |
| 2  | CEOP_Mode1_Grid_BoM         | CEOP Model Output : 3D Gridded BoM data         |
| 3  | CEOP_Mode1_Grid_CPT EC      | CEOP Model Output : 3D Gridded CPTEC data       |
| 4  | CEOP_Mode1_Grid_ECMWF       | CEOP Model Output : 3D Gridded ECMWF data       |
| 5  | CEOP_Mode1_Grid_ECP C_RII36 | CEOP Model Output : 3D Gridded ECP C RII36 data |

**Dataset Metadata History Window:** Shows the history of metadata updates for a specific dataset. The table includes columns for Registered Date and Metadata Status(ja).

| Registered Date     | Metadata Status(ja) |
|---------------------|---------------------|
| 2016/07/06-14:40:49 | Complete            |
| 2016/06/24-14:12:14 | Complete            |
| 2016/06/24-14:07:30 | Complete            |
| 2016/06/24-14:07:30 | Complete            |
| 2016/06/24-13:45:04 | Complete            |
| 2015/09/15-11:49:01 | Complete            |
| 2014/01/21-12:01:54 | Complete            |
| 2013/10/30-15:20:02 | Complete            |
| 2013/08/21-17:48:29 | Complete            |

**Edit Dataset Metadata(MIRAI CTD) Window:** Shows the form for editing dataset metadata. The form includes fields for Title, Contact, Document Author, Dataset Creator, Date of Dataset, Category, Abstract, Temporal Extent, Geographic Bounding Box, Grid, Keywords, Online Resource, Data Environmental Information, Data Distribution, Data Processing, Data Remarks, Data Policy, Acknowledgement, and References. The 'Citation Name' field is highlighted as required.

**Administrator Menu > Dataset List > Dataset Metadata History > Edit Dataset Metadata**

# A Search and Discovery System for DIAS Datasets

<http://search.diasjp.net/en>

- ❖ Overview of entire DIAS datasets
- ❖ Search based on keyword/spatial/temporal conditions
- ❖ Link to the data download system

Login

The collage illustrates the workflow of the DIAS search and discovery system:

- Search Interface:** The main search page shows various filters and a map. A red box highlights the "selection of external metadata portals" section, which includes links to JALTEC, NIPR, and JAMSTEC.
- Axis type selection:** A red box highlights the "Axis type selection" section, which allows users to choose between different data types (e.g., GOMD, GOMD-Plus, GOMD-Plus-Plus).
- Dataset overview:** A red box highlights the "Datasets overview by two axis" section, which displays a table of datasets with columns for name, description, and download link.
- Login:** A screenshot of the login page, showing a "ログイン" (Login) button and a "パスワード" (Password) field.
- Data download:** A red box highlights the "Data download" section, which includes a "データをダウンロード" (Download data) button.
- Metadata download:** A red box highlights the "Metadata download" section, which includes a "メタデータをダウンロード" (Download metadata) button.
- File list:** A screenshot of the "ファイル検索" (File search) page, showing a table of files with columns for name, size, and date.
- Dataset document:** A screenshot of the "データドキュメント" (Data document) page, which provides detailed information about a specific dataset.



# Management of Data Access Privilege

- ❖ Access to and search for **document-metadata** is open to public
- ❖ **Data Access Restrictions:**
  - Login account is required
  - 1. **Free access**
  - 2. **Agreement with data policy is required**
  - 3. **Approval from data administrator is required**
    - ❖ Require manual procedure for approval
    - ❖ Prepare an application form, assist on automatic email and so on.
  - 4. **Others / special treatment**
    - ❖ Contact with data administrator by email or other media. If an application is approved, the user account is granted permission.
    - ❖ The system provides UI for data administrator to change the access privilege for individual user account.



# Architecture of DIAS Metadata Systems

## DIAS Metadata Management System

### Registration of dataset metadata

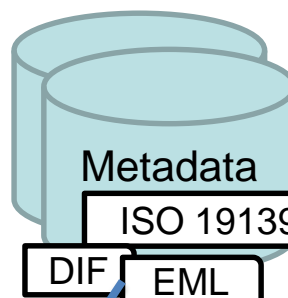


Metadata

ISO 19139

OAI-PMH

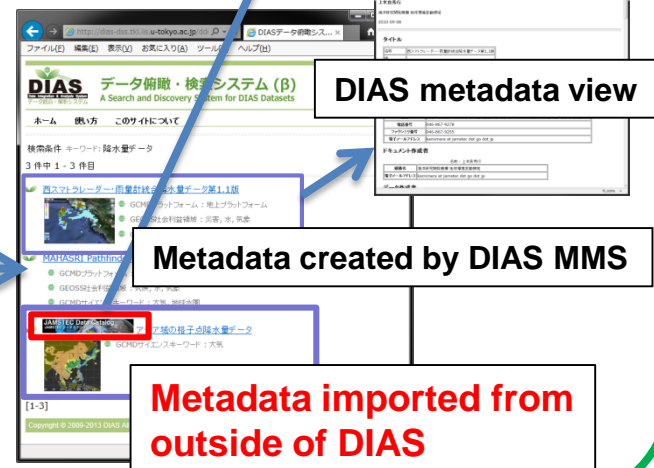
## Systems outside of DIAS



Original metadata page of each system

## DIAS Dataset Search and Discovery System

<http://search.diasjp.net/en>



Metadata created by DIAS MMS

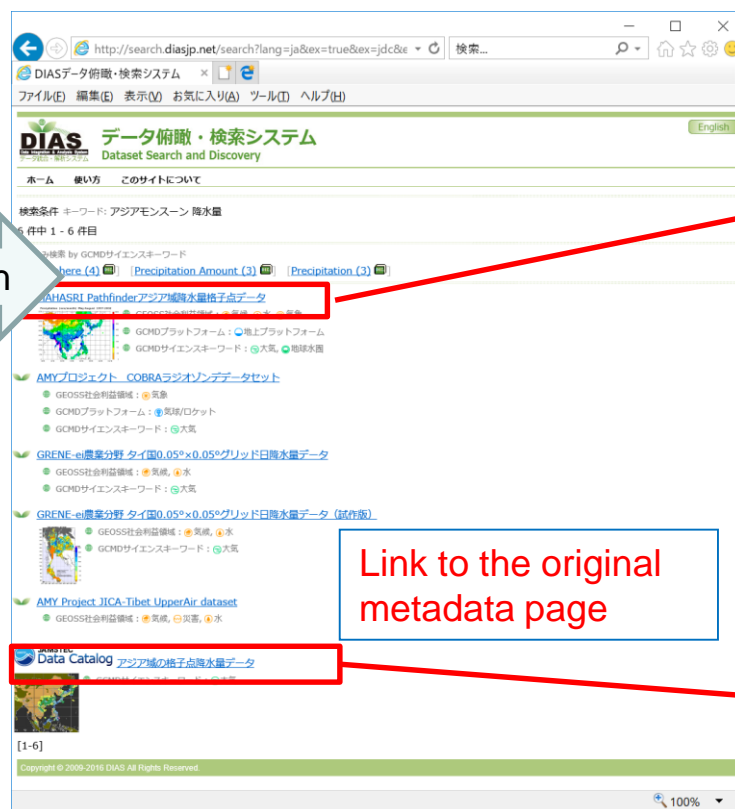
Metadata imported from outside of DIAS

# Metadata Collaboration with Systems outside of DIAS



DIAS metadata

Metadata from outside system(s)



Link to the original metadata page



| System                          | Metadata format | URL   |
|---------------------------------|-----------------|---|
| JAMSTEC Data Catalog            | DIF             | <a href="http://www.godac.jamstec.go.jp/catalog/data_catalog/">http://www.godac.jamstec.go.jp/catalog/data_catalog/</a> |
| JaLTER Data Catalog             | EML             | <a href="http://db.cger.nies.go.jp/JaLTER/">http://db.cger.nies.go.jp/JaLTER/</a>                                       |
| NIPR Science Database           | DIF             | <a href="http://scidbase.nipr.ac.jp/">http://scidbase.nipr.ac.jp/</a>   |
| NIPR Arctic Data archive System | ISO19139, DIF   | <a href="https://ads.nipr.ac.jp/">https://ads.nipr.ac.jp/</a>   |

# Contents

## ❖ About DIAS

## ❖ Data and Metadata Management

- ❖ Data registration procedure
- ❖ Metadata management

## ❖ Open Science Activities

## ❖ Current and Future Prospects

- ❖ DIAS as a national repository
- ❖ Focusing on metadata quality

# DIAS Third Phase and Open Science

1. **DIAS Third Phase (2016-2020)** : from research phases to the operation phase.
2. **Open science** : selected as one of strategic keywords in the national-level science and technology policy.
3. **DIAS Open Science Special Interest Group (SIG)** : planning and implementation to make DIAS ready for open science.
4. **More stakeholders**: variation of openness.

# DOI registration for DIAS data

- ❖ **Digital object identifier (DOI)** : architecture of systems and organizations to make resources findable using a global identifier.
- ❖ DIAS has already started **the assignment of DOI since March 2017**.
  - ❖ We have 26 datasets with DOI assigned in DIAS (Dec. 2017)
- ❖ DOI registration system from DIAS to JaLC and DataCite
  - ❖ Add a new function to DIAS metadata management system to manage DOIs.
  - ❖ Add DOI in each DIAS document-metadata (XML, HTML, PDF)
  - ❖ Convert DIAS metadata XML to JaLC XML to registrate DOI to DataCite through JaLC



# First Assignment of DOI on March 2017

DIAS データ統合・解析システム

Home DIASとは データ・アプリケーション 適用分野 活動紹介 お知らせ

## お知らせ

2017年4月5日

### データ統合・解析システム（DIAS）におけるDOI付与の開始



～地球環境データへのアクセス向上によりデータの公開と利活用を推進～

DIASは公開データに対するDOI（デジタルオブジェクト識別子）の付与を開始しました。DOIは、論文や研究データなど様々な学術情報への永続的なアクセスを支援する識別システムとして、世界的に利用が拡大しています。DIASでもDOIを活用したデータ利活用を促進するために、DOIを付与するための体制を整え、こ

ニュース

- トピックス
- イベント
- データセット

**doi:10.20783/DIAS.496**

DIAS データ俯瞰・検索システム  
Dataset Search and Discovery

日本語

Home How to use About

## GAME Tibet

Data File Download  
with DIAS data download system

IDENTIFICATION INFORMATION

|                     |   |
|---------------------|---|
| Name                | GAME Tibet  |
| DOI                 | <a href="https://doi.org/10.20783/DIAS.496">doi:10.20783/DIAS.496</a> |
| Metadata Identifier | GAME_Tibet20170725133015-DIAS20170725102541-en                        |

## CONTACT

CONTACT on DIAS

|              |   |
|--------------|---|
| Name         | Tamakiwa Icharm   |
| Organization | International Centre of Excellence for Water Hazard and Risk Management |
| Address      | 1-6, Minamihara, Tsukuba, Ibaraki, 305-8516, Japan                      |
| TEL          | 029-879-6779  |
| E-mail       | tamakawa@icharm.org   |

## CONTACT on PROJECT

Data Integration and Analysis System

|              |  |
|--------------|--|
| Name         | DIAS Office  |
| Organization | Remote Sensing Technology Center of Japan  |
| Address      | TOKYU REIT Toranomon Building 2F 3-17-1 Toranomon, Minato-ku, Tokyo, 105-0001, Japan |
| E-mail       | dias-office@diasjp.net   |

[http://www.diasjp.net/infomation/  
press-release-dias-first-doi-registration/](http://www.diasjp.net/infomation/press-release-dias-first-doi-registration/)

[http://search.diasjp.net/en/dataset/GAME\\_Tibet](http://search.diasjp.net/en/dataset/GAME_Tibet)



# Landing Page with Citation Text (under development)

**データ俯瞰・検索システム**  
Dataset Search and Discovery

日本語

Home   How to use   About

## GAME Tibet

Data File Download  
with DIAS data download system



The citation for this dataset is:  
Koike Toshio. (2017, March 19). GAME Tibet. Retrieved December 7, 2017, from <http://doi.org/10.20783/DIAS.496>  
Select citation format:

### IDENTIFICATION INFORMATION

|                     |  |
|---------------------|--|
| Name                | GAME Tibet   |
| DOI                 | <a href="http://doi.org/10.20783/DIAS.496">doi:10.20783/DIAS.496</a> |
| Metadata Identifier | GAME_Tibet20171023144905-DIAS20170725102541-en                       |

# Domain and National Repository

- ❖ DIAS is a domain repository in the areas of earth science and environment.
- ❖ DIAS is a national repository to disseminate research results from Japan.

DIAS can take an important role among the open data policy of Japanese research organizations and funding agencies.

# DIAS as a National Repository

- ❖ Recently, we have accepted some datasets from outside of DIAS.
  - ❖ DIAS can be a candidate for storing large data.
- ❖ DIAS can be used as a repository of **evidence data for research articles**.
- ❖ Data deposited in DIAS can be used for submission to a data journal (e.g. ESSD).
  - ❖ We are discussing on getting official certificates of trustworthy data repositories so that DIAS can be considered as trustworthy from stakeholders.



<https://www.earth-system-science-data.net/>

# Metadata Quality Issues


























- ❖ Some metadata do not contain enough information
  - ❖ due to some reasons, such as metadata specification, usability of systems, motivation of metadata author, etc.
  - ❖ Metadata quality affect the findability of datasets.
  
- ❖ I am especially focusing on keyword information in metadata.

# Keywords in metadata

## Keywords in document-metadata

| Keywords            |   |                        |
|---------------------|---|------------------------|
| Keywords on Dataset |   |                        |
| Keyword Type        | Keyword   | Keyword thesaurus Name |
| theme               | Oceans > Ocean Temperature, Oceans > Salinity/Density > Salinity, Oceans > Ocean Chemistry > Oxygen | GCMD_science           |
| theme               | OCEANOGRAPHY PHYSICAL   | AGU                    |
| theme               | Climate   | GEOSS                  |
| place               | INDIAN OCEAN,PACIFIC OCEAN,WESTERN PACIFIC OCEAN  | others                 |

## Categorization of datasets using keywords

|                       |   | GCMD Platforms  |  |  |  |   |   |  |  |   |
|-----------------------|---|---|--|--|--|---|---|--|--|---|
|                       |   |  Aircraft  |  Balloons / Rockets |  Earth Observation Satellites |  In Situ Land-based Platforms |  In Situ Ocean-based Platforms |  Maps / Charts / Photographs |  Models |  Navigation Platforms |  Undefined |
| GCMD Science Keywords |  Agriculture               |   |  | [ 2 ]  | [ 44 ]   |   |   | [ 27 ]   |  | [ 3 ]   |
|                       |  Atmosphere                | [ 4 ]   | [ 8 ]  | [ 34 ]   | [ 28 ]   | [ 21 ]  |   | [ 47 ]   | [ 3 ]  | [ 100 ]   |
|                       |  Biological Classification |   |  |  | [ 1 ]  | [ 3 ]   |   |  |  | [ 9 ]   |
|                       |  Biosphere                 |   |  | [ 4 ]  | [ 5 ]  | [ 7 ]   |   | [ 1 ]  |  | [ 39 ]  |
|                       |  Climate Indicators        |   |  |  |  |   |   | [ 27 ]   |  | [ 2 ]   |
|                       |  Cryosphere                |   |  | [ 8 ]  | [ 34 ]   |   | [ 1 ]   | [ 27 ]   |  | [ 12 ]  |
|                       |  Human Dimensions          |   |  |  |  |   |   |  |  | [ 2 ]   |
|                       |  Land Surface              |   |  | [ 6 ]  | [ 49 ]   |   | [ 1 ]   | [ 27 ]   |  | [ 35 ]  |
|                       |  Oceans                    |   |  | [ 22 ]   | [ 2 ]  | [ 31 ]  |   | [ 32 ]   |  | [ 61 ]  |
|                       |  Paleoclimate              |   |  |  |  | [ 5 ]   |   |  |  |   |
|                       |  Solid Earth               |   |  | [ 1 ]  | [ 12 ]   | [ 6 ]   |   |  |  | [ 19 ]  |
|                       |  Spectral / Engineering    |   |  | [ 10 ]   | [ 1 ]  |   |   |  |  | [ 19 ]  |
|                       |  Sun-Earth Interactions    |   |  |  |  |   |   |  |  | [ 14 ]  |
|                       |  Terrestrial Hydrosphere   |   |  | [ 6 ]  | [ 2 ]  |   |   | [ 1 ]  |  | [ 1 ]   |
|                       |  Others                    |   |  | [ 2 ]  |  |   |   |  |  |   |
|                       |   |  Undefined | [ 1 ]  |  | [ 1 ]  |   |   |  | [ 9 ]  |   |

e.g. [http://search.diasjp.net/en/dataset/MIRAI\\_CTD](http://search.diasjp.net/en/dataset/MIRAI_CTD)

## Dataset Search and Discovery

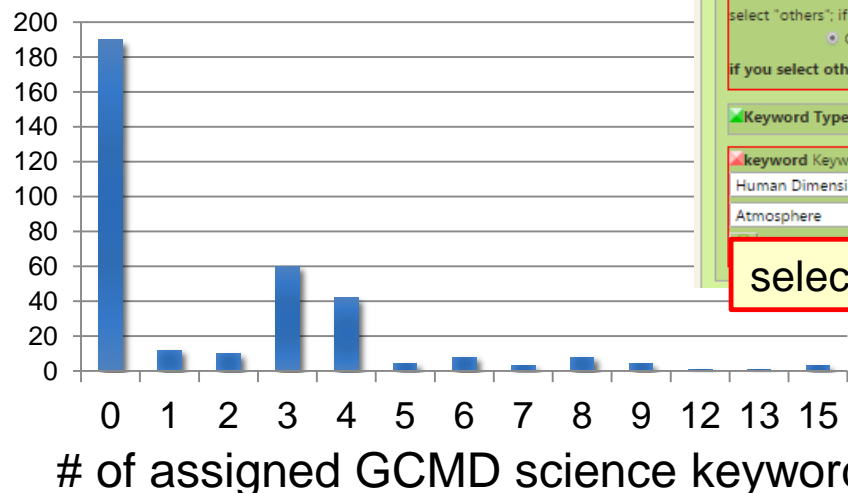
<http://search.diasjp.net/en>

- ❖ We can understand the data through keywords.
- ❖ Keywords are also important for search and categorization of datasets.
- ❖ DIAS manages various datasets.

# Keywords in metadata (cont.)

- ❖ We don't have enough keywords in metadata
  - ❖ The cost of keyword input is high
  - ❖ It is difficult for novice users to input keywords (lack of knowledge)
- ❖ We are now developing keyword recommendation function.

# of datasets in DIAS



Keywords Enter the keyword for the categorization of the data.

Keywords The keyword information is very important for metadata. If possible, please enter at least one dictionary of GCMD\_science, GCMD\_platform and GEOSS.

Keyword Thesaurus Name Keyword thesaurus name selection is required. If using the dictionaries to select "others"; if not using any dictionary, please select "No Dictionary". [Required]

☐ GCMD\_science ☐ GCMD\_platform ☐ AGU ☐ GEOSS ☐ GEO\_COP ☐ Country ☐ other

If you select others, please enter: \_\_\_\_\_

Keyword Type If keyword thesaurus name is selected with "others" or "no dictionary", the selection

keyword Keywords are required. Entering more than one keyword is allowed. Please enter one keyword

Human Dimensions > Economic Resources > Agricultural Economics

Atmosphere > Atmospheric Chemistry/Oxygen Compounds > Ozone

specification of ontologies (GCMD\_science, GCMD\_platform, GEOSS, AGU, Country, others)

selection of hierarchical keywords from menu

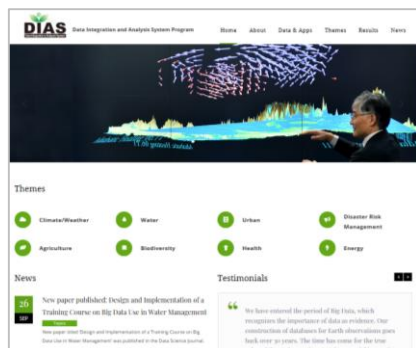
Current interface for keyword input boxes (in the DIAS Metadata Management System)



# Summary

- ❖ DIAS is not only a data repository, but also an **information platform** for data science.
- ❖ We are managing various kinds of datasets through the **metadata**.
- ❖ We will continuously work to make DIAS more **open** platform.
  - ❖ DOIs to datasets
  - ❖ FAIR Data Principle

# Thank you!



DIAS Website  
<http://www.diasjp.net/en/>



You can search DIAS datasets via  
DIAS Dataset Search and Discovery System  
<http://search.diasjp.net/en>

## Acknowledgments

I thank people in the DIAS open science special interest group, Dr. Asanobu Kitamoto, Dr. Masafumi Ono, Dr. Hiroko Kinutani, Dr. Masatoshi Yoshikawa, and Mrs. Yoko Nakahara for helpful discussion.