

Open Science Activities in DIAS

Hiroko Kinutani

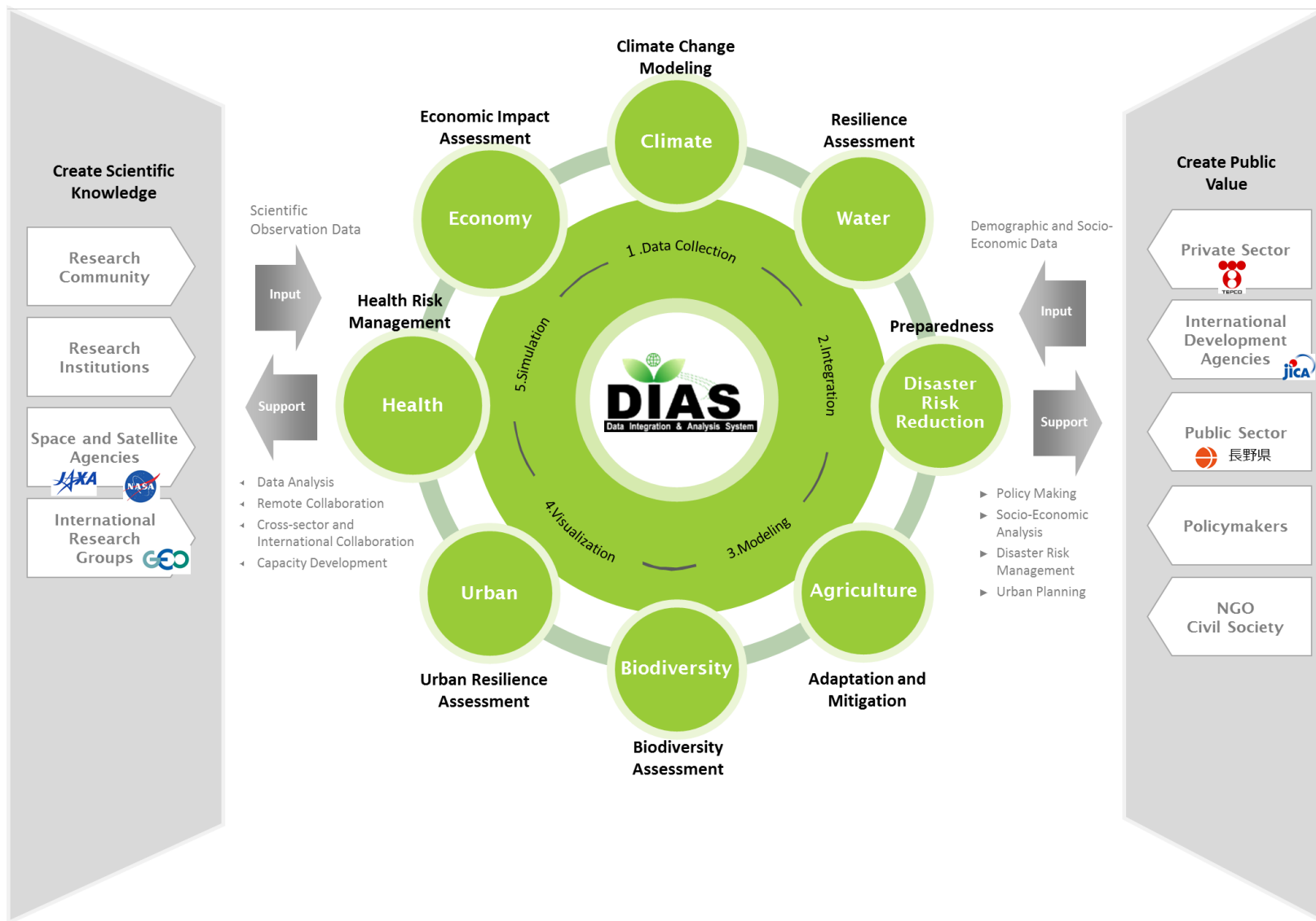
Earth Observation Data Integration & Fusion Research
Initiative (EDITORIA), The University of Tokyo
kinutani@tkl.iis.u-tokyo.ac.jp

Mar. 1st, 2017

Contents

- ❖ DIAS Overview
- ❖ Data Life Cycle in DIAS
- ❖ DIAS Metadata
- ❖ The Workshop for DIAS metadata input
- ❖ DOI registration plan for DIAS data

DIAS: Overview

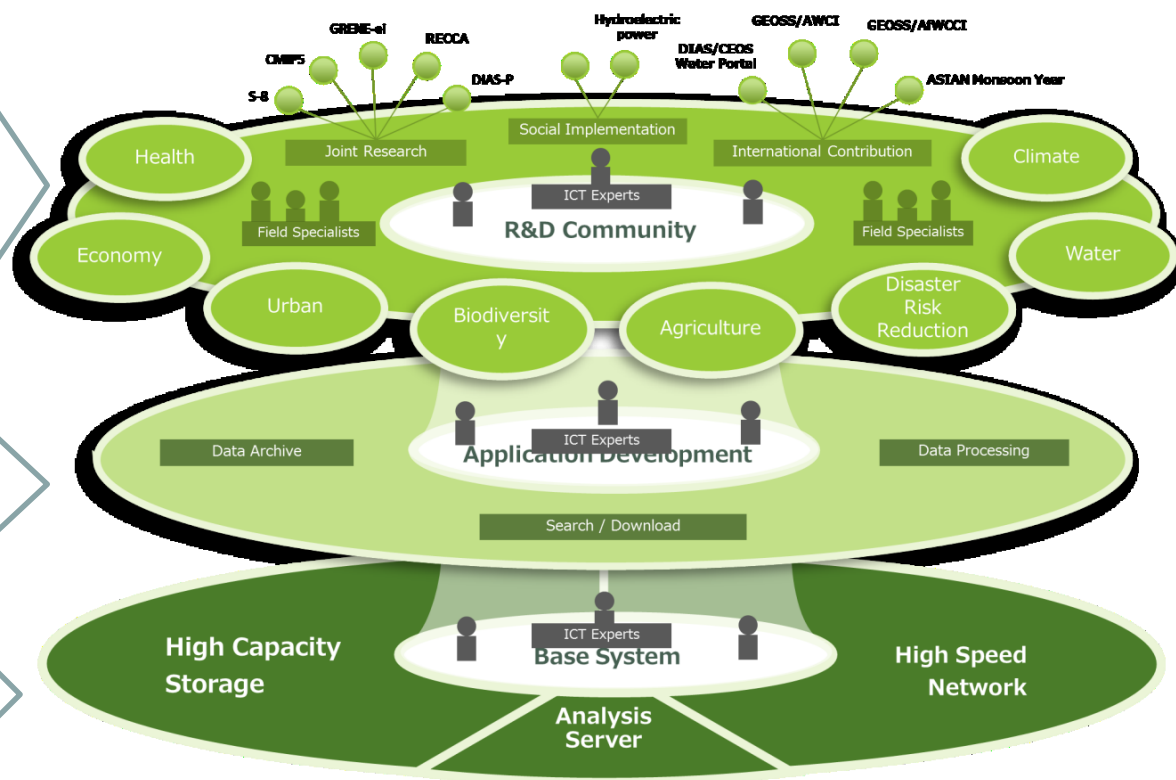


DIAS Phase I and II (2006FY-2015FY)

1. Collect data, Store data, Release data
2. Analysis space
3. Cooperation with various research fields for implementing the outcome of research projects

R&D Community
Climate, Weather, Water,
Agriculture, Biodiversity,
Urban, Economics, Health
Registered 2,000 person

Application Development
for Data Archiving,
Searching,
Processing and Analyzing
Storage, Server, Network
Infrastructure

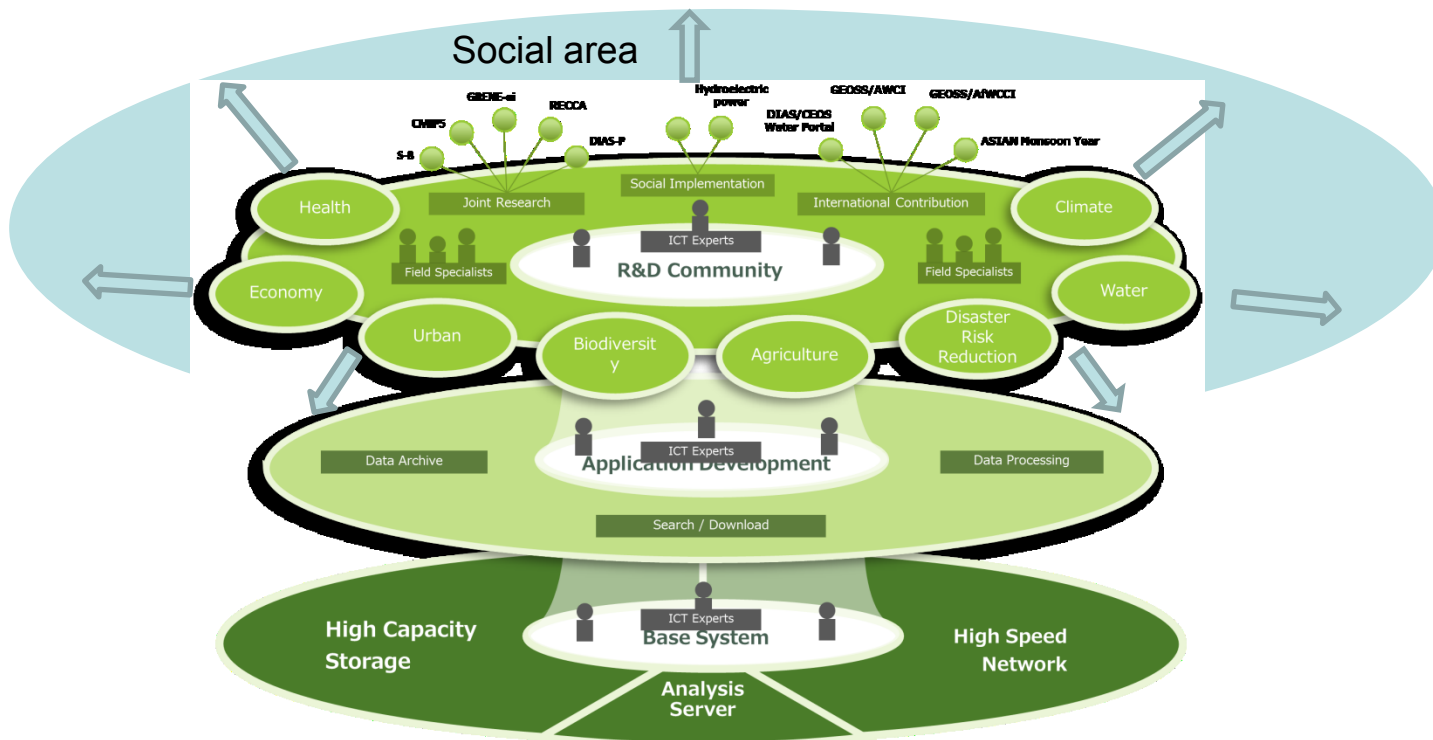


DIAS Phase III (2016FY-2020FY)

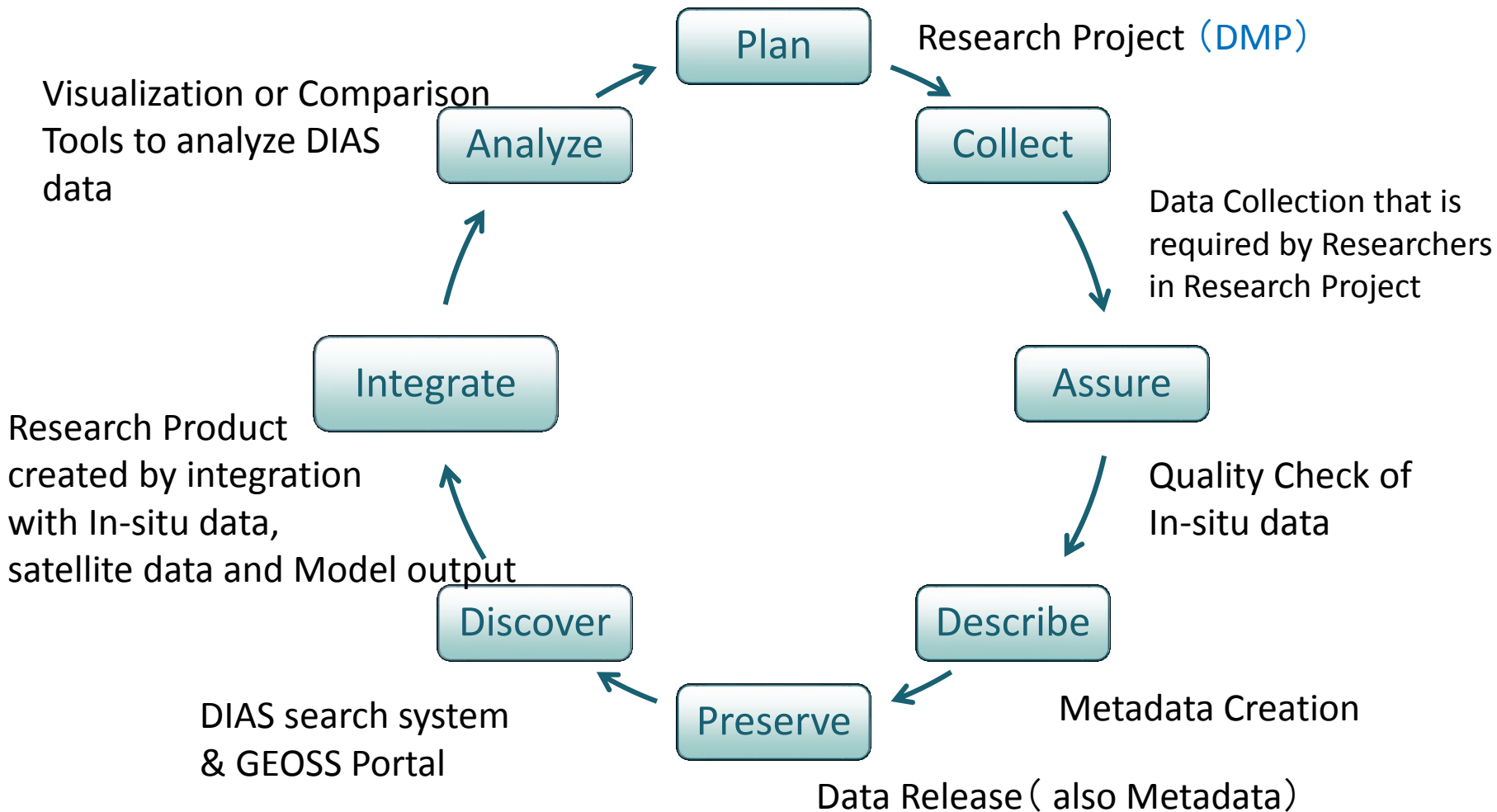
1. Collect data, Store data, Release data
2. Analysis space
3. Cooperation with various research fields for implementing the outcome of research projects



Long-term stable social Platform to help solve various social problems

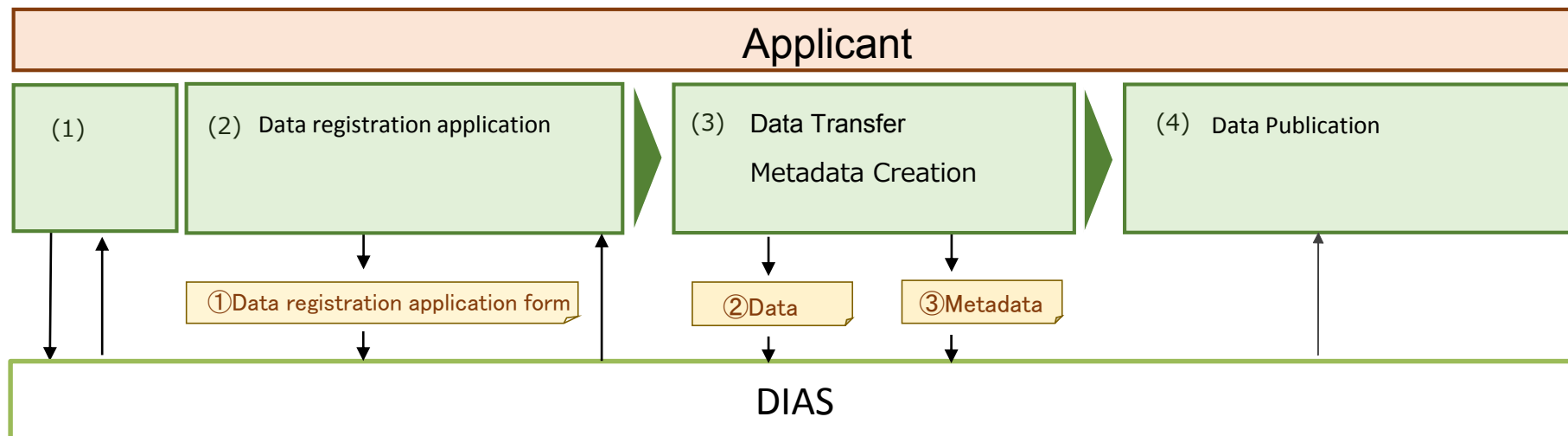


Data Life Cycle in DIAS



Original Figure: DataONE Education Module: Data Entry and Manipulation. DataONE. Retrieved Nov12, 2012. From http://www.dataone.org/sites/all/documents/L07_Metadata.pptx

Flow of DIAS data registration & data publish



(1) Consultation with DIAS office (if required):

dias-office@diasjp.net

(2) Data registration application procedure:

DIAS will decide if this application can accept or not.

(3) Data registration procedure:

Data transfer, Metadata creation

(4) Data publication procedure:

User will able to download data

DIAS Metadata

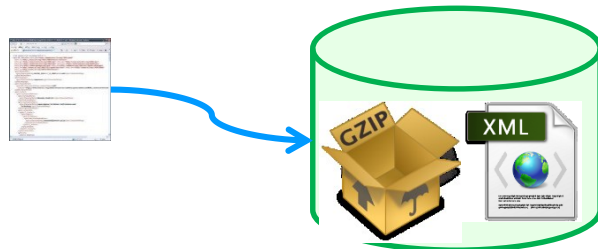
- ❖ We are managing various dataset in DIAS
- ❖ Basic strategy:

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

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

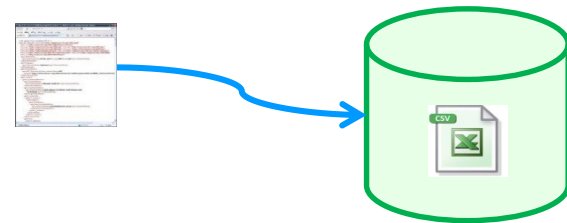
Examples of datasets

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:schema="http://www.isotc211.org/2005/gmd"
http://www.isotc211.org/2005/gmd/gmd.xsd" xmlns:xsi="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:xdt="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_Charac
codeListValue="UTF />
  </gmd:characterSet>
  <- <gmd:contact>
    <gmd:CI_IndividualParty>
      <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 dot go dot
```

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

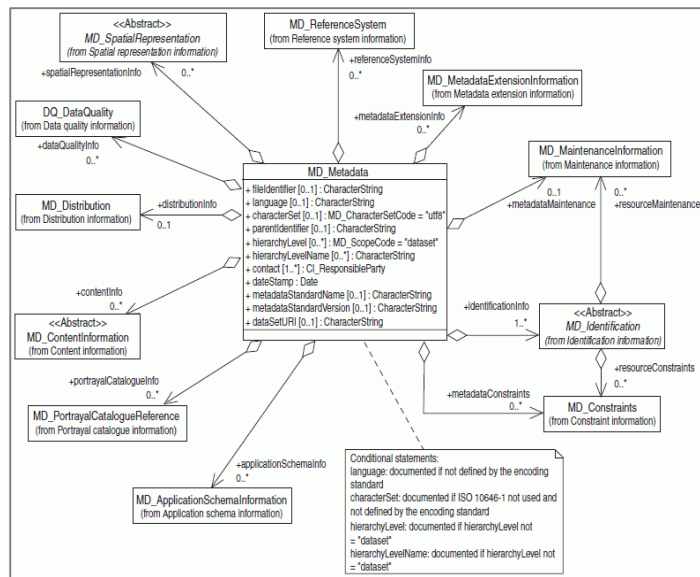
2.2.1 Data Integration and Analysis System

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

PDF document

ISO19115 / ISO19139

- ❖ Most of earth environmental data commonly have spatial and temporal attributes such as the covering geographic scope or the created date.
- ❖ Accordingly, DIAS metadata is developed with basing on ISO/TC211 metadata standards.



ISO19115

```
<?xml version="1.0" encoding="utf-8" ?>
<xs:schema targetNamespace="http://www.isotc211.org/2005/gmd" elementFormDefault="qualified"
version="0.1" xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:gco="http://www.isotc211.org/2005/gco"
xmlns:gmd="http://www.isotc211.org/2005/gmd">
<!-- ===== Abstraction ===== -->
<!-- ===== Annotation ===== -->
<!-- ===== Imports ===== -->
<xs:import namespace="http://www.isotc211.org/2005/gco" schemaLocation="../gco/gco.xsd" />
<xs:include schemaLocation="../gmd/spatialRepresentation.xsd" />
<xs:include schemaLocation="../gmd/metadataExtension.xsd" />
<xs:include schemaLocation="../gmd/metadataApplication.xsd" />
<xs:include schemaLocation="../gmd/applicationSchema.xsd" />
<xs:include schemaLocation="../gmd/portrayalCatalogue.xsd" />
<xs:include schemaLocation="../gmd/dataQuality.xsd" />
<xs:include schemaLocation="../gmd/freeText.xsd" />
<!-- ===== Classes ===== -->
<!-- ===== Complex Type ===== -->
<xs:complexType name="MD_Metadata_Type">
<xs:annotation>
<xs:documentation>Information about the metadata</xs:documentation>
</xs:annotation>
<xs:complexContent>
<xs:extension base="gco:AbstractObject_Type">
<xs:sequence>
<xs:element name="fileIdentifier" type="gco:CharacterString_PropertyType" minOccurs="0" />
<xs:element name="language" type="gco:CharacterString_PropertyType" minOccurs="0" />
<xs:element name="characterSet" type="gmd:MD_CharacterSetCode_PropertyType" />

```

ISO19139

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, 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 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_CTD¹¹









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  , [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

Online Resource

- 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 and the terms and conditions at <http://www.jamstec.go.jp/cruisedata/e/readme.html>

Data Policy for Project

Data Integration and Analysis System

- 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.
- DIAS data sets are to be used only for scientific research or educational purposes [*1].
- Any modification or change of the original DIAS data sets is prohibited.
- 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_CTD¹²

ISO19115 Core Metadata item vs. DIAS Metadata item

	ISO19115 Core Metadata items	DIAS Metadata (represented as document) item
1	Dataset title (M)	1.TITLE
2	Dataset reference date (M)	6.DATE OF DATASET
3	Dataset responsible party (O)	2.CONTACT
4	Geographic location of the dataset (by four coordinates or by geographic identifier) (C)	7.4 Geographic Bounding Box
5	Dataset language (M)	× nothing
6	Dataset character set (C)	× nothing
7	Dataset topic category (M)	7.2 Topic Category(ISO19139)
8	Spatial resolution of the dataset (O)	7.5 Grid
9	Abstract describing the dataset (M)	7.1 Abstract
10	Distribution format (O)	7.10 Distribution Information
11	Additional extent information for the dataset (vertical and temporal) (O)	7.3 Temporal Extent
12	Spatial representation type (O)	× nothing
13	Reference system (O)	× nothing
14	Lineage (O)	8. DATA PROCESSING
15	On-line resource (O)	7.8 Online Resource
16	Metadata file identifier (O)	1.TITLE (Metadata Identifier)
17	Metadata standard name (O)	× nothing
18	Metadata standard version (O)	× nothing
19	Metadata language(C)	× nothing
20	Metadata character set(C)	× nothing
21	Metadata point of contact (M)	3. DOCUMENT AUTHOR
22	Metadata date stamp (M)	5. DATE OF THIS DOCUMENT

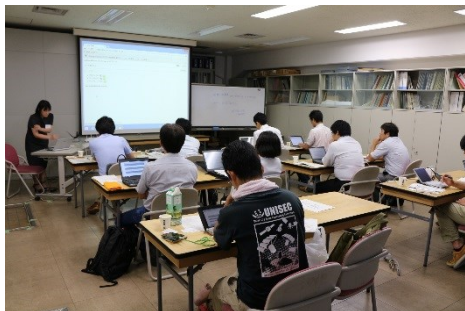
The Workshop for DIAS metadata input

- ❖ The contents of metadata cannot know except the data provider.
- ❖ We have held the workshop about DIAS metadata input in order to understand what kind of information should be inputted for each field of metadata.



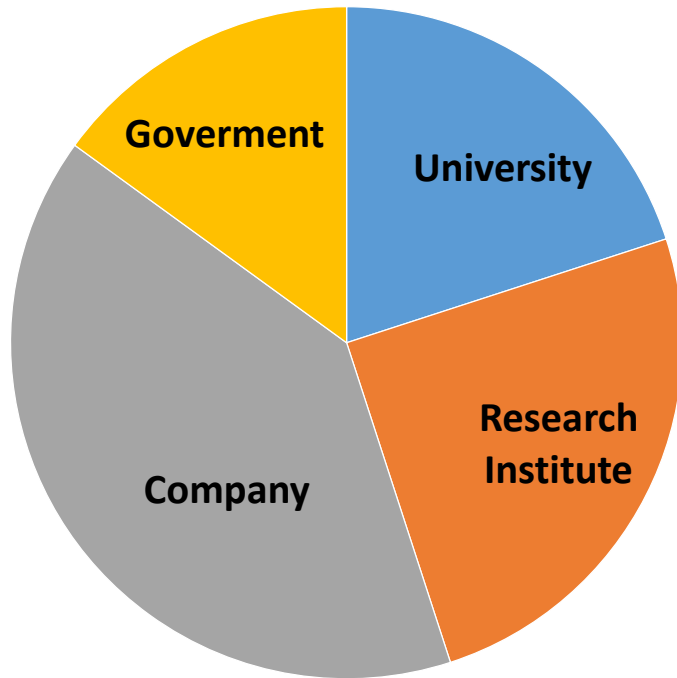
History

- ❖ Sept. 2013 : participated the JaLTER All Scientist Meeting to understand what the Data input camp is.
- ❖ The workshop for metadata input camp
 1. 2014/7/31 1st workshop at U.Tokyo 7 participants
 2. 2014/11/25 2nd workshop at UN univ. 5 participants
 3. 2015/2/25 3rd workshop at Hatsumei Kaikan 6 participants
 4. 2015/10/25 4th workshop at U.Tokyo 12 participants
 5. 2016/3/28 5th workshop at U.Tokyo 9 participants
 6. 2017/2/28 6th workshop at Nikkei Hall 19 participants

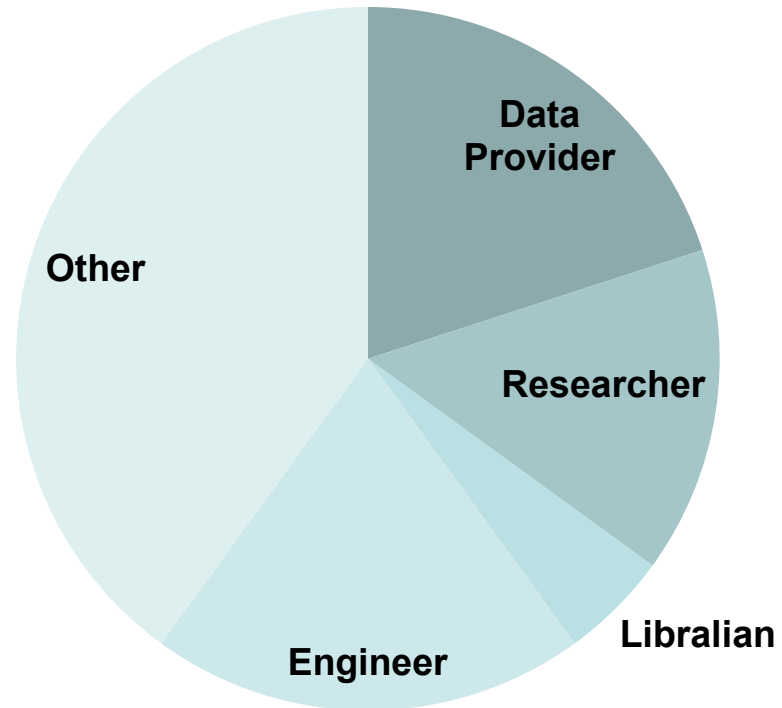


The 6th Workshop Participants

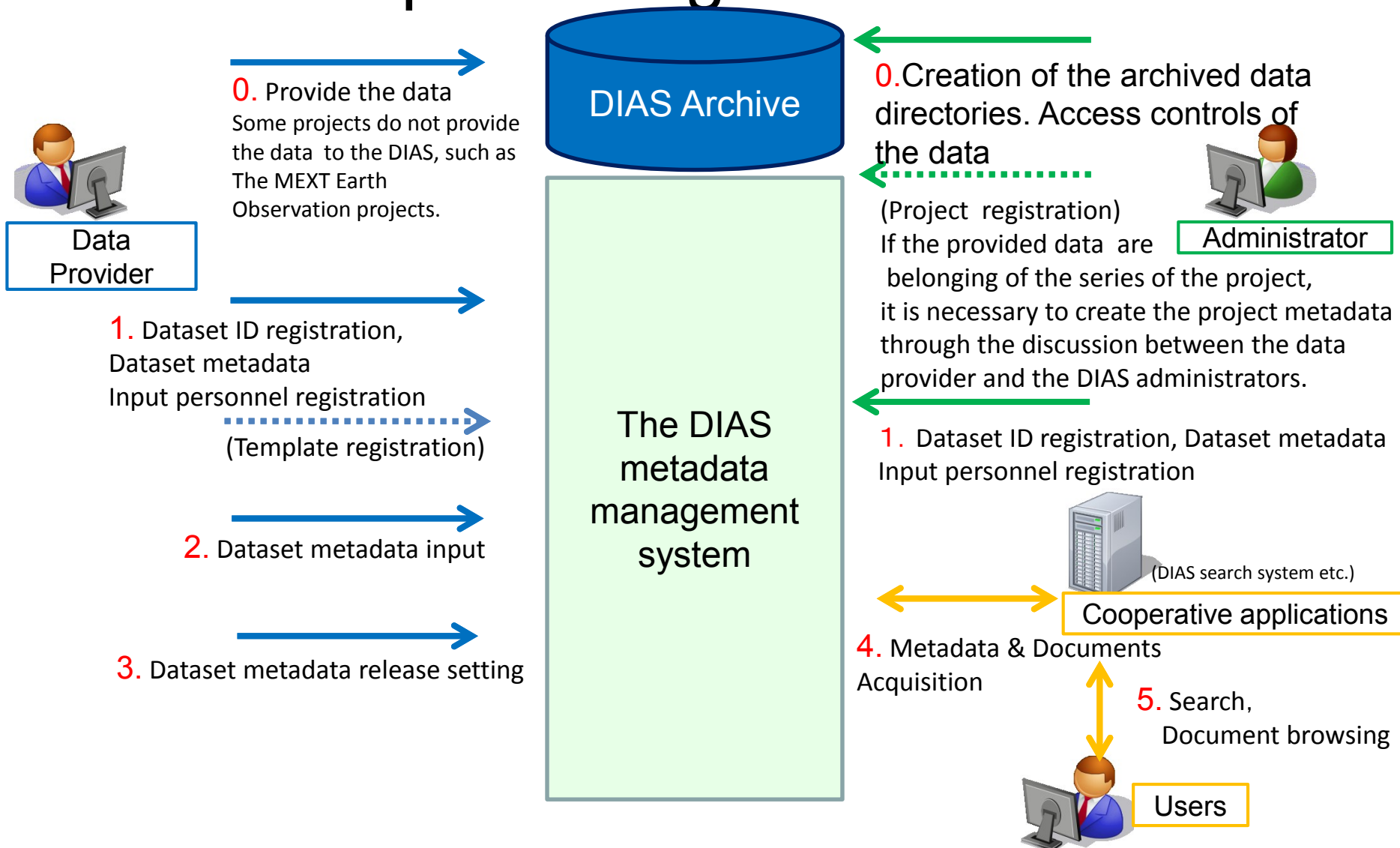
Organization



Category



Flow of the Metadata registration & documents publishing



Metadata creation using the DIAS metadata management system

① Login

② User Menu

③ Dataset List

データセットID	データセット名(日)	データセット名(英)	メモ	メタデータ	関連プロジェクトID	入力者ID
test_dataset_20121221	テストデータセット 2012年12月21日	Test Dataset		Temporary / Temporary		

④ Dataset Selection & Dataset Metadata History

登録日時	メタデータの状態(日)	メタデータの状態(英)
2012/12/21-15:04:42	Temporary	Temporary
2012/12/21-15:02:49	Uninputted	Uninputted

Editing screen

⑤ Metadata Edit

(Title Input screen)

(Keyword Input screen)

キーワード	キーワードタイプ
GEOS5	
GCM science	
Forest Science	Please select
Precipitation	Precipitation Amount
Atmospheric Temperature	Air Temperature
Atmospheric Water Vapor	Humidity
Atmospheric Winds	Surface Winds
Soils	Soil Heat Budget
Atmospheric Chemistry	Carbon and Hydrocarbon Compounds

Dataset Metadata : Keyword

- ❖ **Keyword information is very important** among the metadata.
- ❖ Please enter at least one keyword from each GCMD_science, GCMD_platform and GEOSS and enter keywords as much as possible.
- ❖ If keyword is layered (GCMD_science, GCMD_platform, etc.), you may select the upper layer keyword.

identifier on the system	Name	abstract
GCMD_science	GCMD Science Keyword	Global Change Master Directory, Science Keyword
GCMD_platform	GCMD Platform	GCMD, Keyword on platforms like satellite and so on
AGU	AGU Index Terms	American Geophysical Union
GEOSS	GEOSS SBA	Group on Earth Observations, nine societal benefit areas, disasters, health, energy, climate, water, weather, ecosystems, agriculture, biodiversity
GEO_COP	GEO Critical Observation Parameters	GEO user interface committee report, important Earth Observation items, vocabulary mainly developed by the University of Tokyo team of DIAS
Country	Country Name	Information of country related to data

DOI registration plan for DIAS data

- ❖ Until the end of FY2016, we have a plan to register a few DOIs to JaLC (Japan Link Center).
- ❖ 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
- ❖ Create a new DOI registration system from DIAS to JaLC



<https://japanlinkcenter.org/>

System revision for introducing DOI

- ❖ Metadata Management System
 - ❖ Select dataset to assign doi by DIAS.
doi:10.20783/DIAS.1
 - ❖ Input DOI assigned outside DIAS.(in Future)
- ❖ Select dataset necessary to register/update DOI information to JaLC/DataCite.
 - ❖ Use OAI-PMH function of the Metadata Management System.
- ❖ DOI registration System to JaLC/DataCite (New System)
 - ❖ Register/Update DOI information of DIAS dataset

Conclusion

❖ DIAS Phase III

- ❖ To construct data platform, we try to communicate and collaborate with private company and local government engineers.
- ❖ For sustainable running of DIAS, we will create a new organization structure considering open science.

❖ DIAS Metadata

- ❖ To touch and know about DIAS metadata at DIAS metadata input Camp
- ❖ DOI registration