

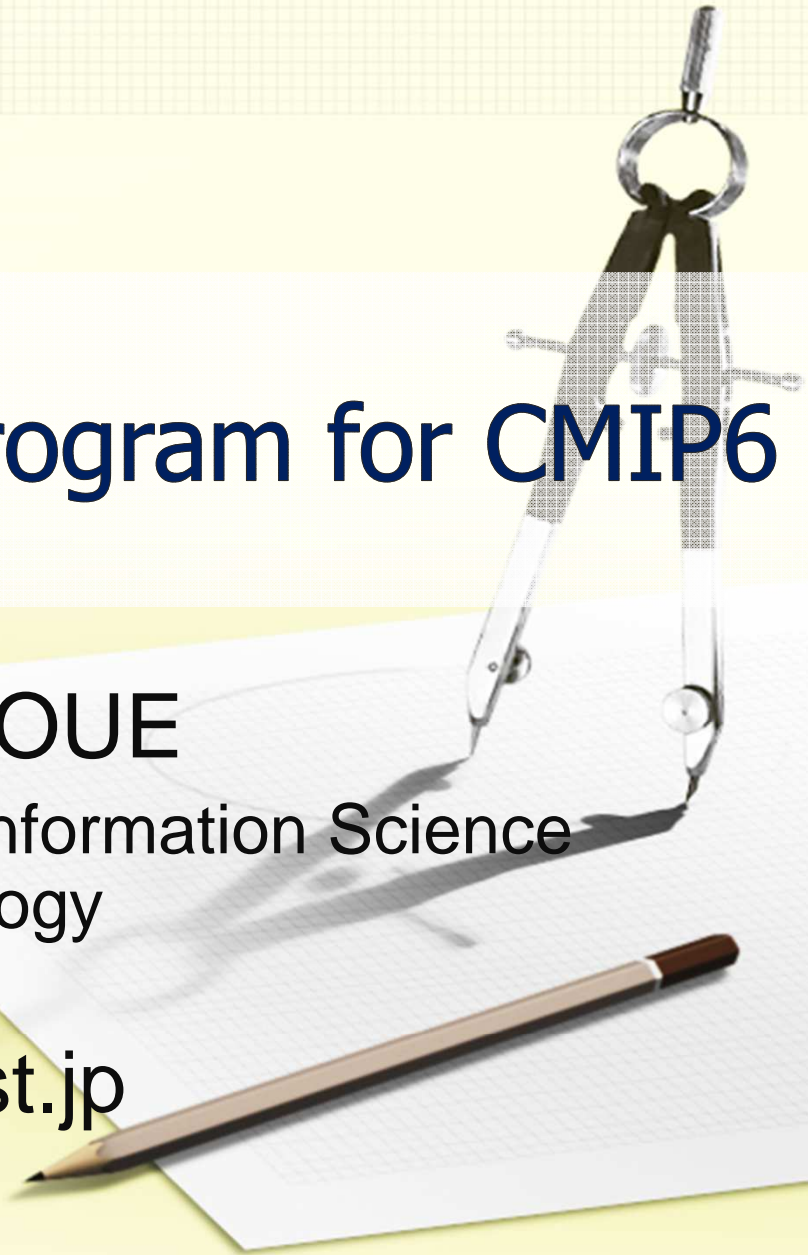


Japanese "TOUGOU" program for CMIP6

Takahiro INOUE

Research organization for Information Science
and Technology
(RIST)

tinoue@rist.jp



Requests for this panel;

- ▶ We're looking for topics related to CMIP6 data analysis and "onsite" capabilities.
- ▶ In particular, we're looking for any experience you have at JAMSTEC related to serving big data (CMIP6) to researchers.
- ▶ A few questions we'll be looking at:
 1. Do you provide analysis capability "next to" your CMIP6 data holdings?
 2. What types of analysis workflows are you using?
 3. Do you use public or private cloud technologies?
 4. What are your biggest challenges in serving CMIP6 data to your use community?
 5. What lessons have you learned?

First of all ...

- ▶ The “**TOUGOU**” program (explains later) plays main part of CMIP6 related activities in Japan.
- ▶ **JAMSTEC** is the leading institute of one of the four sub themes in TOUGOU.
- ▶ **DIAS** is the research project and built platform system for various and vast environmental data, along with the analysis capability.
- ▶ **DIAS** operates a data node for ESGF, and serves as a principal data provider for Japanese climate models, such as MIROC and NICAM.
- ▶ On behalf of the TOUGOU program members, I will answer the questions, from both **JAMSTEC**'s and **DIAS**'s point of view.



Japan Agency for
Marine-Earth Science
and Technology



Data Integration and
Analysis System

Thanks to

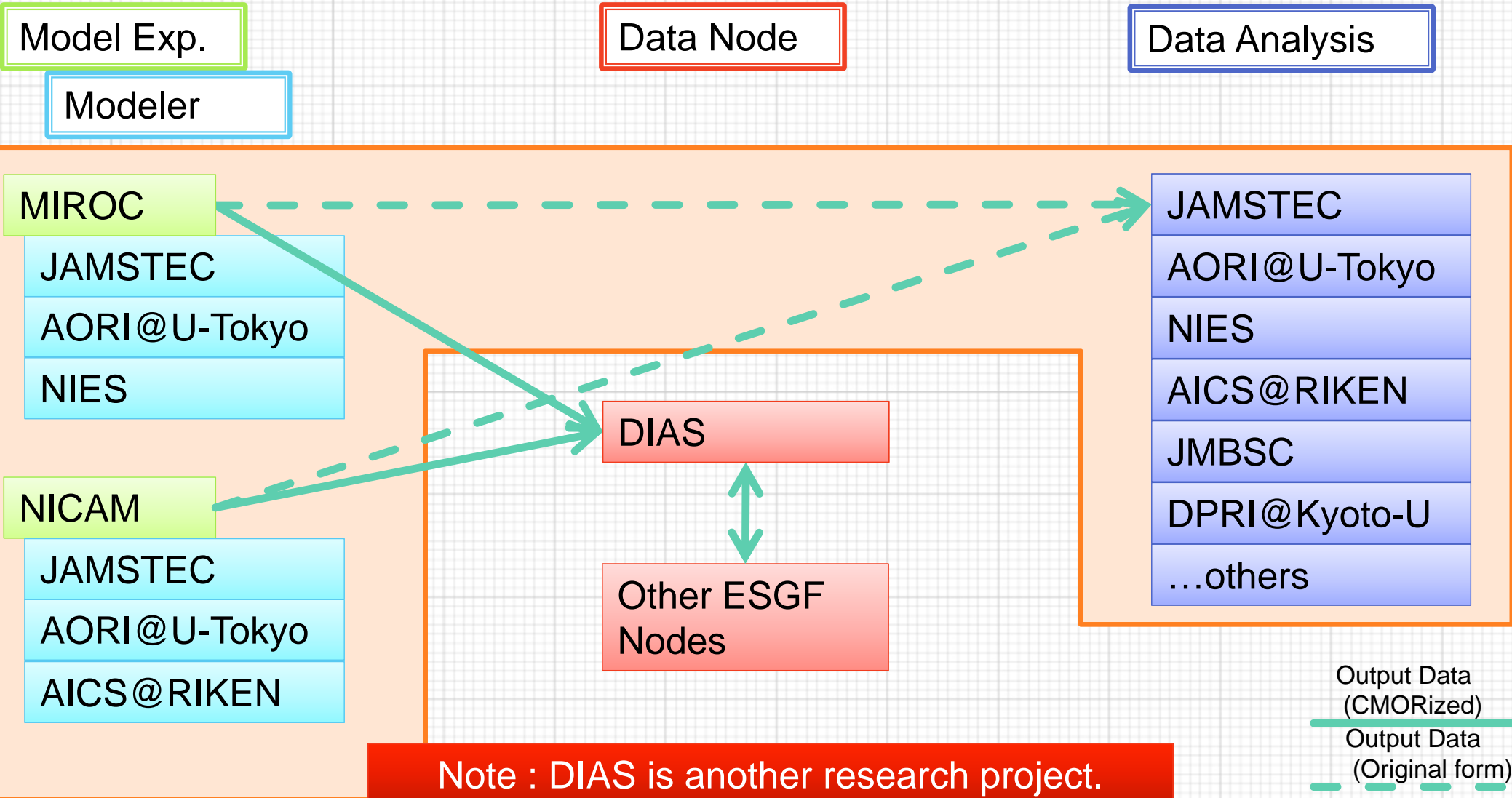
- ▶ Michio Kawamiya(JAMSTEC, MIROC)
- ▶ Manabu Abe(JAMSTEC, MIROC)
- ▶ Hiroaki Tatebe(JAMSTEC, MIROC)
- ▶ Chihiro Kodama(JAMSTEC, NICAM)
- ▶ Hisashi Yashiro(AICS@RIKEN, NICAM)
- ▶ Toshihiro Nemoto[DIAS]
- ▶ Yukari Takayabu(AORI@U-Tokyo, Analysis)
- ▶ Nagio Hirota(NIES, Analysis)
- ▶ Naota Hanasaki(NIES, Analysis)

and many others.

Japanese "TOUGOU" program for CMIP6

TOUGOU Program :
Integrated Research Program for Advancing Climate Models

“統合”



Note : DIAS is another research project.

Output Data (CMORized) →
Output Data (Original form) →

Do you provide analysis capability "next to" your CMIP6 data holdings?

▶ [JAMSTEC] Yes, but limited.

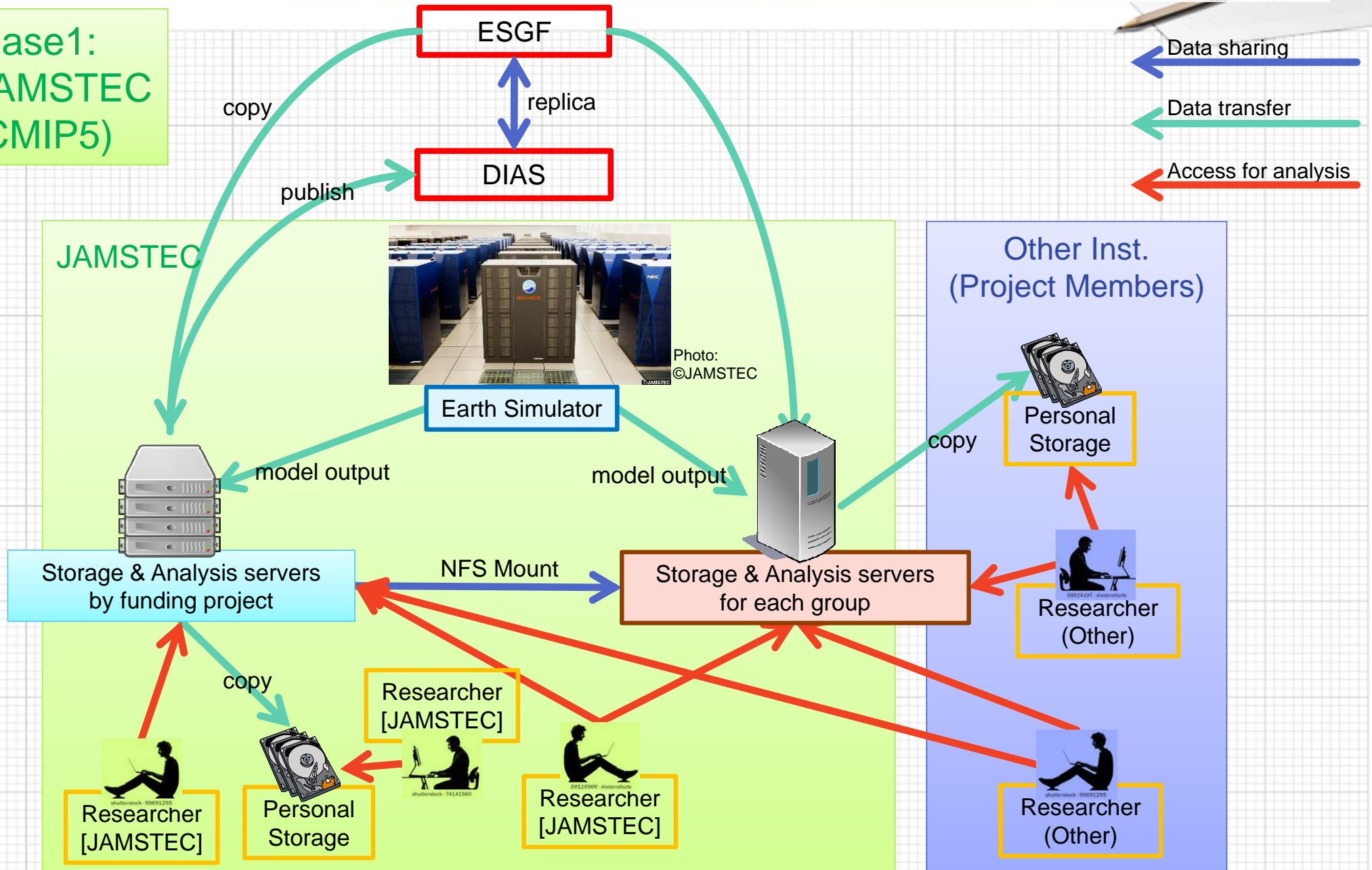
- ▶ In JAMSTEC there are several servers, some of them are for this project, the other are for each research groups. They are dedicated to researchers in JAMSTEC and the project members in related institutes.
- ▶ See also a next slide.

▶ [DIAS] Yes, but limited.

- ▶ DIAS provides large analysis clusters to the registered users.
- ▶ Also provides several web applications, including "CMIP5 Data Analysis System".
- ▶ These are original, not a part of the ESGF software.

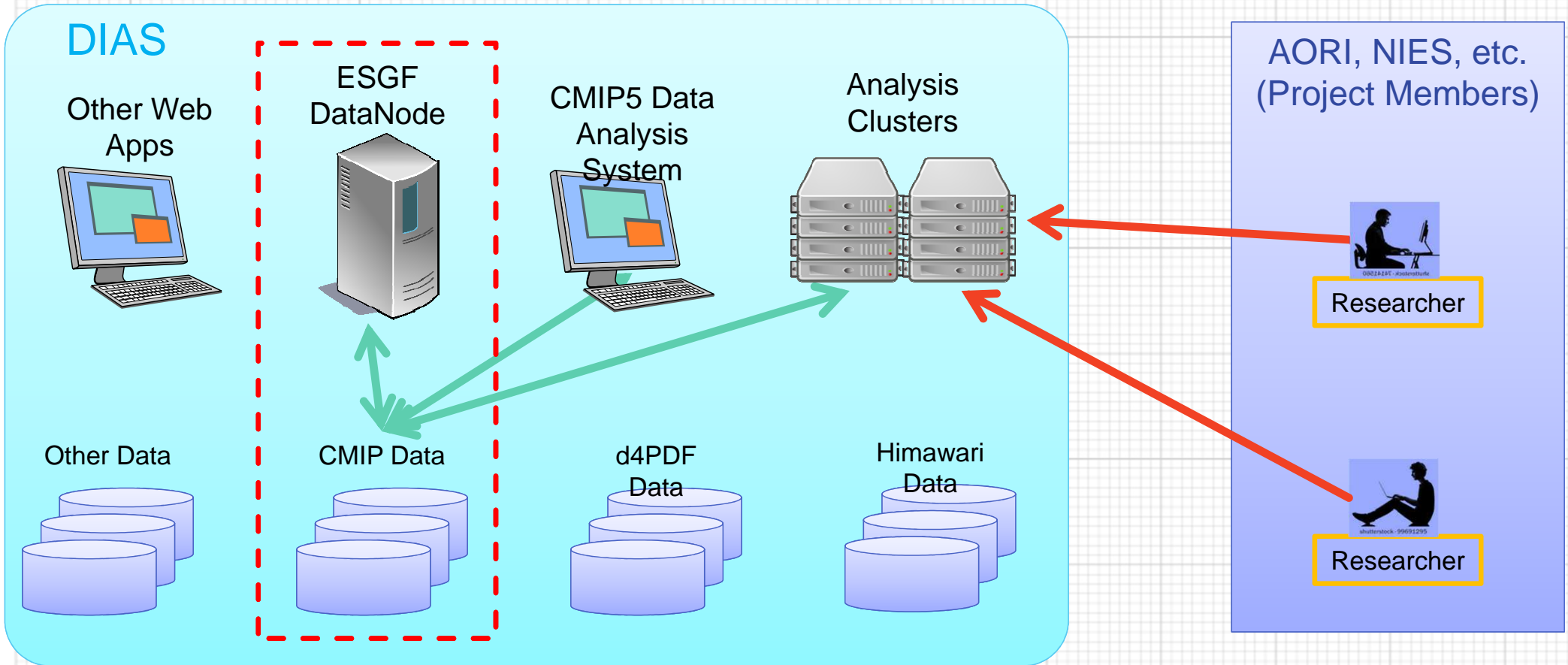
What types of analysis workflows are you using?

Case1:
in JAMSTEC
(CMIP5)



What types of analysis workflows are you using?

Case2: Using DIAS's analysis cluster for CMIP5 data



Some researchers at AORI, NIES etc. used analysis clusters as main computer resources for data analysis,

- to access CMIP multi model data directly, and
- to avoid download/copy large volume data.

Do you use public or private cloud technologies?

▶ [JAMSTEC] No.

▶ [DIAS] Yes.

▶ DIAS is also a cloud platform.

▶ Various machine instances, such as ESGF data node, analysis clusters, are running as virtual machines.

▶ DIAS is preparing various API's to implement (mainly) new web applications.



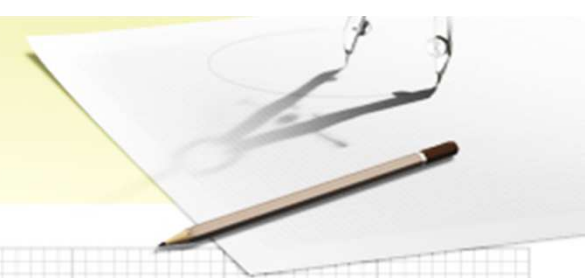
What are your biggest challenges in serving CMIP6 data to your use community?

- ▶ **[DIAS]** Constructing original (DIAS specific) analysis tools/environments for DIAS users.
- ▶ For DIAS, it is not challenging to maintain CMIP6 data as one of the ESGF data node.

What lessons have you learned?

IMHO

- ▶ “Analysis” differs for each researcher, though, most of them have to do the same things, such as;
 - ▶ acquiring data of necessary models/scenarios/variables,
 - ▶ extracting and re-gridding them to necessary spatial/temporal range/resolution.
 - ▶ different models use different resolution.
- ▶ On-Demand processing by WPS API to do these will be helpful.
 - ▶ Is it possible ?
- ▶ And/Or, it should be able for researchers to do analysis close to the data storage (like DIAS).
 - ▶ I know some ESGF Tier1 nodes prepare this as “compute node”.
- ▶ Serving data should not be a funded project but a operational service of the institute.
 - ▶ Particular projects do not last forever.



Thank you.