DATA SCIENCE JOURNAL

Kawasaki, A et al 2017 Data Integration and Analysis System (DIAS) Contributing to Climate Change Analysis and Disaster Risk Reduction. *Data Science Journal,* 16: 41, pp. 1–17, DOI: https://doi.org/10.5334/dsj-2017-041

PRACTICE PAPER

Data Integration and Analysis System (DIAS) Contributing to Climate Change Analysis and Disaster Risk Reduction

Akiyuki Kawasaki[1], Akio Yamamoto[2], Petra Koudelova[1], Ralph Acierto[2], Toshihiro Nemoto[2], Masaru Kitsuregawa[2,3] and Toshio Koike[1,4]

[1] Department of Civil Engineering, The University of Tokyo, Tokyo, JP

[2] Institute of Industrial Science, The University of Tokyo, Tokyo, JP

[3] National Institute of Informatics, Tokyo, JP

[4] International Centre for Water Hazard and Risk Management (ICHARM), Tsukuba, JP

Corresponding author: Akiyuki Kawasaki (kawasaki@hydra.t.u-tokyo.ac.jp)

In 2015, global attempts were made to reconcile the relationship between development and environmental issues. This led to the adoption of key agreements such as the Sustainable Development Goals. In this regard, it is important to identify and evaluate under-recognized disaster risks that hinder sustainable development: measures to mitigate climate change are the same as those that build resilience against climate-related disasters. To do this we need to advance scientific and technical knowledge, build data infrastructure that allows us to predict events with greater accuracy, and develop data archives. For this reason we have developed the Data Integration and Analysis System (DIAS). DIAS incorporates analysis, data and models from many fields and disciplines. It collects and stores data from satellites, ground observation stations and numerical weather prediction models; integrates this data with geographical and socio-economic information; then generates results for crisis management of global environmental issues. This article gives an overview of DIAS and summarizes its application to climate change analysis and disaster risk reduction. As the article shows, DIAS aims to initiate cooperation between different stakeholders, and contribute to the creation of scientific knowledge. DIAS provides a model for sharing transdisciplinary research data that is essential for achieving the goal of sustainable development.

Keywords: Climate change; Disaster risk reduction; Integration; DIAS; CMIP; Myanmar

Introduction

Humans have needs – food, water, energy, health, security – and development is required to provide for these needs. Such development, and a rapid increase in population, have led to improper land use change, disorderly urbanization, and unstable governance, with societal problems the result. Development has also caused environmental issues on a global scale: climate change, desertification, deforestation, and loss of biodiversity. States must consider how to balance these issues with needed development.

In 2015, global attempts were made to reconcile the relationship between development and environmental issues. The Sendai Framework for Disaster Risk Reduction was developed in March that year (Aitsi-Selmi et al., 2016). The Sustainable Development Goals (SDGs) followed in September, the Paris Agreement on Climate Change (Paris Agreement) in December. As the SDGs and the Paris Agreement state, measures to mitigate climate change are the same as those that build resilience against climate-related disasters. In this regard, it is important to identify and evaluate under-recognized disaster risks that hinder sustainable development. This can be done by studying hazards and vulnerabilities that arise from societal and environmental problems and the way these risks change.

To build resilient societies, scientists and engineers must follow two paths in developing and using science and technology (Future Earth, 2013): promote interdisciplinary research between natural/applied sciences and humanities/social sciences and advocate transdisciplinary cooperation to facilitate the implementation of science and technology in social contexts. We need to advance scientific and technical knowledge, build relevant data infrastructures that allow for the anticipation of future events with greater accuracy, develop data archives, and expand our understanding of how changes and disasters unfold across different regions and sectors.

For the above reasons, we developed the Data Integration and Analysis System (DIAS), which collects and stores data from satellites, ground observation stations, and numerical weather prediction models; integrates these data with geographical and socio-economic information; and provides results for the crisis management of threats such as global environmental issues and natural disasters. These results are available in Japan and overseas. This paper gives an overview of DIAS and summarizes its application to climate change analysis.

Overview of DIAS

Development process

In April 2006, relevant departments of the University of Tokyo established a collaborative initiative among researchers from various fields: civil engineering, public policy, global earth observation, information science and technology, disasters, and agriculture. The collective was named the Earth Observation Data Integration & Fusion Research Initiative (EDITORIA).

The Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) supported EDITORIA to develop DIAS as part of the Earth Observation and Ocean Exploration System, which is one of the five National Key Technologies defined by the 3rd Basic Program for Science and Technology (Japan). DIAS, a demonstrative data system that effectively and efficiently integrates global and local observation data and information, has led to research breakthroughs in understanding, forecasting, and adapting to global environmental changes, particularly with regard to changes in the water cycle and the effects on water management systems and societies across the globe. For example, the Powered Visualizer for Earth Environmental Science of DIAS was used to explain the mechanism of moist air flow over the Tibetan Plateau, which plays a key role in the onset and maintenance of the Asian Summer Monsoon (Yasukawa et al., 2008). With the support of DIAS, Seto et al. (2013) analyzed the vertical structure of the atmospheric heating process over the Tibetan Plateau. In addition, DIAS was used to facilitate the acquisition and processing of data in a study by Monichoth et al. (2014), which focused on the determination of current and estimation of future rainfall patterns, optimal planting dates, and rice yields in Cambodia. Moreover, the work of Rasmy et al. (2015) in assessing future water resources in Japan and of Sawada et al. (2015) on land data assimilation for the simultaneous simulation of soil moisture and vegetation dynamics were supported by DIAS data and data integration resources. DIAS also serves as a platform for the operational real-time flood prediction system with a dam-operation-optimization support function for the Upper Tone River in Japan (Shibuo et al., 2016).

During the second stage of development, from 2011, DIAS became the Data Integration and Analysis System-Program. In this period, an operational framework beyond 2016 was established. During 2006–2016, DIAS research and development was fully financed by MEXT. Since 2016, MEXT has continued to support DIAS research and development, and is also highly recommending the infusion of private funds. DIAS is exploring opportunities for collaboration with various private sector organizations. In this context, for promoting private sector involvement, the main project management body of DIAS was transferred from EDITORIA of the University of Tokyo to the Remote Sensing Technology Center of Japan (RESTEC) in 2016. EDITORIA focuses on system development and science implementation.

From an international perspective, the role of DIAS is to contribute to the Global Earth Observation System of Systems (GEOSS). Connection with data centers in countries involved with GEOSS is proceeding. The Belmont Forum, an international group of funding agencies for science and technology, is promoting the development of e-infrastructures and data management for global change research to aid interdisciplinary and transdisciplinary cooperation. DIAS contributes to this activity as an e-infrastructure project (Belmont Forum, 2015).

DIAS framework

DIAS has several goals:

- to coordinate cutting-edge information in communities and among researchers working to address global environmental issues;
- to collect and archive environmental data (global and local) from around the world;
- to build a data infrastructure that allows for the effective integration of earth observation data, numerical model outputs, and socio-economic data;
- to use and analyze earth observation data to support sound decision-making with regard to risk and resource management issues in the global environment, such as climate change and disaster-risk reduction; and
- to develop scientific knowledge that helps solve global environmental issues and generate socio-economic benefits.

The framework of DIAS is shown in **Figure 1**.

DIAS features the development of research activities across the data lifecycle, including the measurement, retrieval, analysis, archiving, and sharing of information related to the global environment. There are a number of similar specialized infrastructure systems in data measurement or management. For example, CoopEUS (Cooperation EU + US), ICSU-WDS (World Data System), and EarthCube (US) cover a wide range of data from space weather prediction, carbon observation, and biodiversity to geosciences, but mainly focus on scientific data. The concept of the Data Observation Network for Earth (DataONE, US) considers the needs of science and society, but focuses on education and outreach. By contrast, DIAS aims to solve social problems using scientific and social data related to the Earth's environment and to address solution development using a broad cross-sectorial approach. To do so, the DIAS platform design comprises an infrastructure system, application development, and a research and development (R&D) community. DIAS provides a unique model for sharing transdisciplinary research data to facilitate social problem-solving. While the other systems mentioned above are designed to provide users with comprehensive data for each field, they do not address solution development with data users.

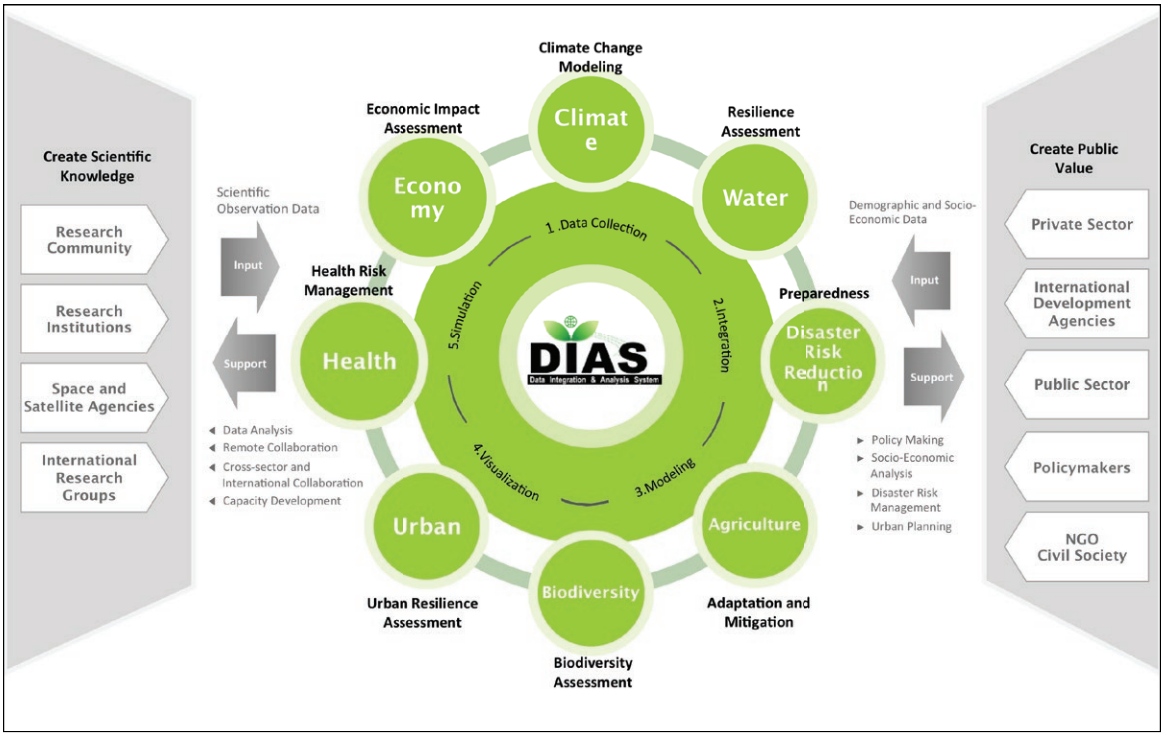

Figure 1: DIAS framework.

Figure 2 shows the three systems that comprise DIAS, detailed as follows.

- Infrastructure system: integrates Earth and local observation data and numerical models, socio-economic data, and other massive datasets relating to the global environment.
- Application development: implements storage, search, analysis, visualization, and other data-related functions on the infrastructure system to provide scientific knowledge and resolve global environmental and societal issues.
- Research and development (R&D) community: DIAS provides a close R&D environment that allows domain scientists who study the environment and IT experts to promote joint research and development through cooperative work, planning, and production.

The DIAS infrastructure system

To archive, analyze, and simulate ultra-high volumes of data, DIAS has a massive data storage/analytical space (25 PB) and an analysis cluster exceeding 16 cores × 120 nodes. To enable high-speed data transfer between data centers and supercomputers at remote organizations, it is connected with the National Institute of Informatics' Science Information Network (SINET). Other features include data processing applications and analysis tools and a vast quantity of multidisciplinary global environmental data, which are stored on the DIAS infrastructure system. Application of such integrated data infrastructure simplifies the super-positioning and related analysis of different datasets, thereby enabling the development of new knowledge and services. Tools developed in pursuit of a given goal can also be re-used for other objectives, thus opening paths for horizontal expansion across disciplines. The structure of the DIAS infrastructure system is such that applications are extended and enhanced in a continuous and expanding cycle.

DIAS application development

DIAS applications and tools consist of software and services that implement data storage/search, analysis, visualization, and other functions developed by domain scientists and IT experts working together to find concrete solutions to global and local environmental and social issues. These applications and tools simplify access to global and regional environment datasets and include functions that allow users to specify a

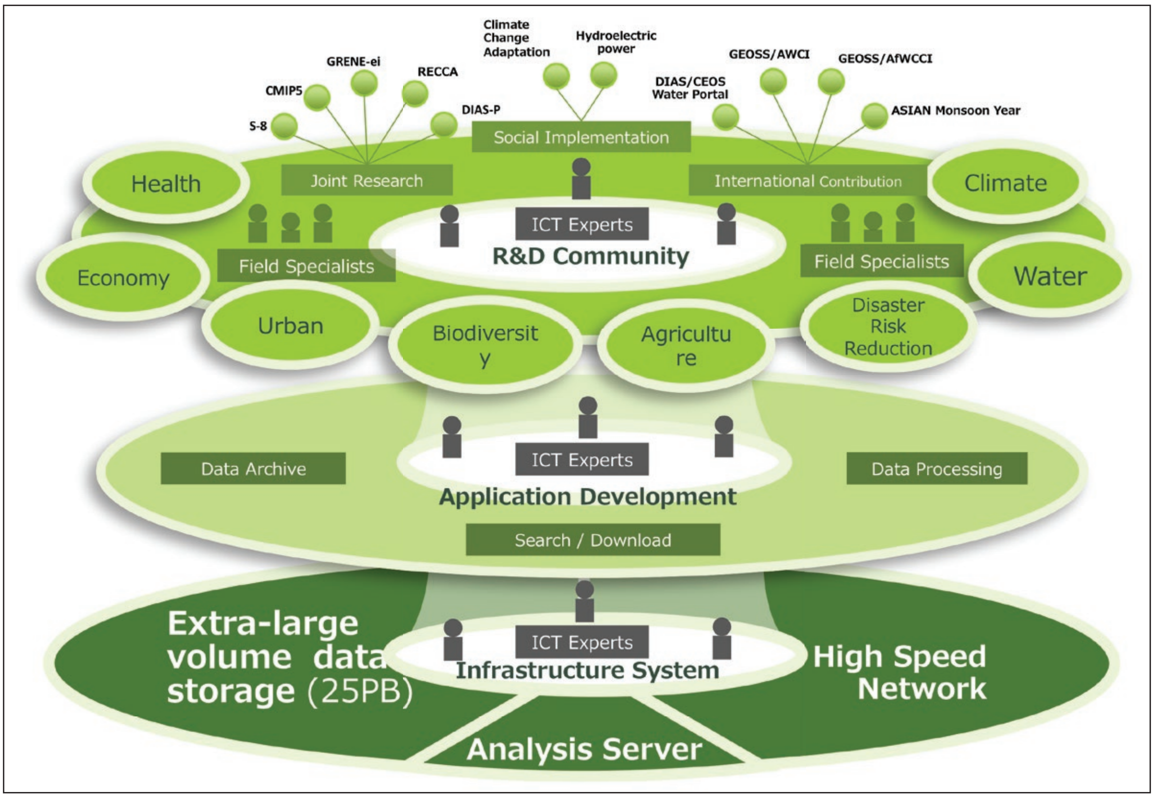

Figure 2: Three components of DIAS.

temporal or spatial range of interest. As of October 2016, the DIAS application website shows 31 tools available for public use (**Figure 3**).

There are five categories of DIAS applications and tools: “Data search and discovery,” “Real-time data process and archive,” “Data and model integration,” “Visualization and analysis,” and “Crowd-sourced science.” Examples of each of these five categories are given below.

- ‘Data search and discovery’. CEOS (The Committee on Earth Observation Satellites) Water Portal was developed by the Japan Aerospace Exploration Agency (JAXA). Scientists studying matters related to water use this portal in the development of services associated with data integration and distribution. General users (non-researchers) also use the portal.
- ‘Real-time data process and archive’. AMeNOW! (Precipitation Real-time Visualization Service) enables the display of XRAIN (X-band polarimetric (multi parameter) Radar Information Network) rainfall observation data on PCs, smart phones, tablets, and other devices. In addition to displaying still images and animations showing precipitation patterns (previous 15 minutes-24 hours), the service allows users to check precipitation information and forecasts for their current position. This is done through the use of GPS and other location information.
- ‘Data and model integration’. DIAS has an development platform named Water Cycle Integrator for coupling various dataset and models for a specific problem solving in the field of water. An example is near real-time data integration ensemble flood prediction system (Shibuo et al., 2014). In order to realize near real-time ensemble flood prediction, meteorological ground observation data, radar data, river discharge data, and dam inflow and release data were integrated and processed for hydrological model on real-time based with ensemble rainfall prediction. This system provides water cycle parameter monitoring, flood predictions, and dam management support information for the Tone River in Japan.
- ‘Visualization and analysis’. The development of a climate change analysis support tool is explained in detail in Section 3. To date, climate change support tools have been applied mainly to the fields of climate and hydrology. In future, these tools will incorporate data from biological and ecological systems, and from multidisciplinary fields such as health, agriculture, and urban planning.

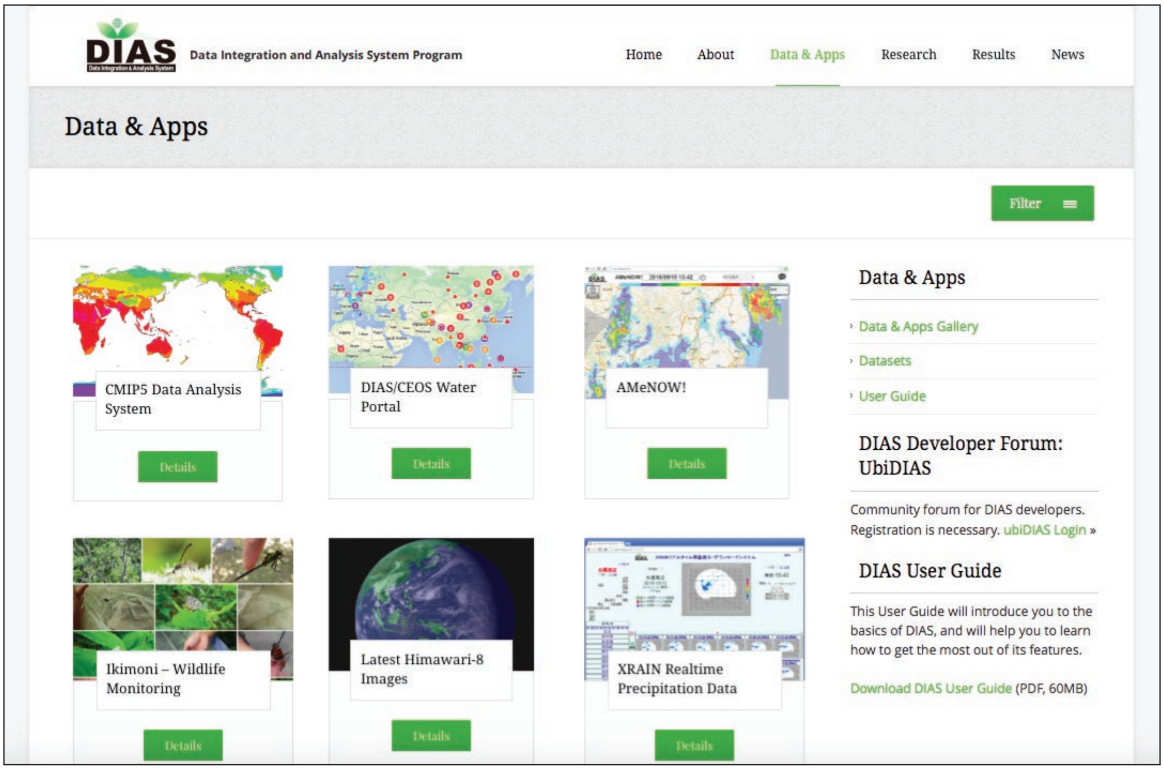

Figure 3: DIAS applications and tools available to the public.

- 'Crowd-sourced science'. The Citizen-participatory Wildlife Monitoring Survey (Ikimoni) is an application that aims to collect and release data on butterflies in Tokyo. This web-based system builds understanding of biodiversity through citizen involvement in the accumulation of data. Citizens use a data input page to upload survey entries and pictures of butterflies. Butterfly specialists then cleanse the data using a data quality management tool. Once the quality check is complete, the data is released and made available for anyone to use.

Adding new and enhancing existing tools will lead to integrated problem-solving and stimulate the creation of new research approaches.

The DIAS R&D community

One of the features of DIAS is its diversity. Approximately 90 researchers from 14 universities and institutions have participated in its research and development. Among the members of this community, a DIAS R&D meeting function has existed for 16 years, the time through which domain scientists working in environmental fields and IT experts have held regular discussions. The input of scientists is needed if the applications are to be used in addressing social issues, and the contribution of IT experts is vital to incorporate cutting-edge information and communication technology into the DIAS infrastructure system. This DIAS R&D meeting has played an important role in determining topics to be developed as a DIAS R&D project via intense discussions between domain scientists and IT experts. This means that only valuable research topics with a high scientific challenge and involving big data analysis in an interdisciplinary and/or transdisciplinary manner can be approved as DIAS R&D projects. This process has established a kind of standard and uniqueness with respect to DIAS R&D. After critical discussion and screening at the DIAS R&D meeting, the domain scientists and IT experts work closely together to develop tools and applications of high value for users. Our approach at DIAS may be unique in terms of this development process: IT experts work to develop a tool or application once it has been identified as a top scientific contribution in a domain, rather than the data-handling capability being addressed by domain scientists.

DIAS application for climate change analysis

The Coupled Model Intercomparison Project (CMIP; Meehl et al., 2000; Covey et al., 2003) is a standard experimental protocol for studying the output of coupled atmosphere–ocean general circulation models (coupled GCMs). CMIP was established in 1995 under the World Climate Research Programme's (WCRP) Working Group on Coupled Modelling (WGCM). GCMs have been employed for decades in theoretical investigations of the mechanisms of climatic change, including the detection of anthropogenic effects in past climate records and the projection of future climate change due to the human production of greenhouse gases and aerosols. The third phase of CMIP, known as CMIP3 (Meehl et al., 2007), delivered an unprecedented, comprehensive, and coordinated set of model outputs that served the Intergovernmental Panel on Climate Change (IPCC) Working Group I. These outputs were used to prepare the IPCC Fourth Assessment Report in 2007 (IPCC, 2007). Future climate projections delivered by CMIP3 were based on a set of greenhouse gas emission scenarios for the 21st century, as formulated in the Special Report on Emission Scenarios (SRES; Nakicenovic et al., 2000). The thoughtful design of the experiments and open access to the coordinated dataset enabled a large group of climate researchers to analyze these data and carry out an extensive range of studies based on them. Nonetheless, due to the complex format of CMIP3 data in the Program for Climate Model Diagnosis and Intercomparison (PCMDI) home archive and its relatively large volume (33 TB), the data were not easily accessible, especially to researchers from other disciplines and users outside the scientific community.

To make the CMIP3 data more easily accessible to a wide range of users, a dedicated CMIP3 analysis tool was developed on DIAS (accessible at http://www.dias.nii.ac.jp/modelvis/cmip3/) and launched in 2010. This tool enables users to quickly access, browse, visualize, analyze, subset, and download CMIP3 data, which have also been replicated to the DIAS archive from the original site at PCMDI. These functions can be performed for a specified time or space through simple screen operations and without the need for users to have the data on their own computers. The DIAS CMIP3 tool has been used extensively, especially by hydrologists in Asia studying the impact of future climate change on hydrological regimes and water resources at the river-basin scale. With water resources being a key prerequisite for sustainable development, such assessment studies are in demand by those managing and developing policies for these resources. For relevant examples, see Koike et al., 2015.

Following the IPCC AR4 publication and reflecting on gaps in information that CMIP3 could fill, new coordinated climate model experiments were designed under CMIP. These experiments comprise the fifth phase of CMIP (CMIP5; Taylor et al., 2012) and are much broader in scope than those undertaken for CMIP3. The CMIP5 dataset, which includes outputs from more than 20 modeling groups worldwide, was used to prepare the IPCC's Fifth Assessment Report (IPCC, 2013). The future climate projections provided through CMIP5 are based on new IPCC greenhouse gas emission scenarios—the so-called representative concentration pathways (RCPs; Moss et al., 2008; van Vuuren et al., 2011). Since the CMIP5 dataset is massive and multidimensional, only users in command of advanced data processing techniques can handle its original format. Building on its experience with the CMIP3 tool, the DIAS community has developed a similar analysis tool with extended capabilities, which is dedicated to CMIP5 data. The CMIP5 dataset was a new challenge for the DIAS team, mainly due to its volume (1.7PB) and complexity in terms of the number of experiments, GCMs, their versions, and ensemble runs.

Development of the CMIP5 tool was possible due to the unique DIAS environment, which was created by the integration of three components: a top-technology infrastructure with large storage capability and a high-performance analysis system, an application development platform, and an R&D community. While the advanced IT infrastructure of DIAS is vital, the development of an application such as the CMIP5 tool would not be possible without the close R&D environment DIAS affords. Intensive and iterative collaboration between the DIAS IT team and a group of CMIP5 data users from the hydrological and environmental science community led to the creation of a tool based on user needs. Its interface is simple and intuitive, and use of the tool can be explained through short training sessions or self-learned via the Help document attached to the tool.

The CMIP5 tool offers the following functions:

1. Selection, visualization, and download of particular pieces of data, whether for a given time period, area, or climate variable.
2. Comparison of model output with reference data (observation-based or reanalysis). This function includes 2-D statistical analysis, a seasonal time-series visual inspection function (1-D plots), and a 2-D wind vector diagram inspection function with difference images (**Figure 4**).
3. Visualization of inter-annual change, including 1-D time series and 2-D areal plots.
4. Intercomparison of global warming projections among models, including 2-D plots and statistical analysis.
5. Tools for CMIP5 analysis, including daily and monthly data download functions, a statistical precipitation bias correction function (Nyunt et al., 2013, 2016) (**Figure 5**), and a Mann–Kendall trend analysis function (Mann, 1945; Kendall, 1975; Gilbert, 1987).

The DIAS CMIP5 tool has made future climate change projection data available to Japanese research institutions and public and private bodies. At the same time, international contributions to the promotion and social implementation of research in the field of climate change have expanded through the use of the CMIP5 tool as part of Official Development Assistance (ODA) in the Philippines, Indonesia, Tunisia, and other countries. The CMIP5 tool has greatly simplified the selection and processing of GCM output for hydrological and water resource applications at the basin scale.

Koike et al. (2015) described the methodology for using CMIP3 or CMIP5 data in assessing the impact of climate change on hydrological regimes and water budgets in river basins. GCM outputs suitable for regions of interest are used as forcing data for hydrological models. The CMIP5 tool function, which allows for 2-D comparison of model output with reference data, can be used to select suitable GCM outputs. Since there is substantial bias in the precipitation output of GCMs across the entire intensity spectrum (insufficient extreme events, biased mean intensity, extensive light intensity drizzle, too few dry days), this precipitation output cannot be used to force hydrological or other impact models without some form of prior bias correction. A three-step statistical bias correction method for precipitation, developed by Nyunt et al. (2013, 2016), has been incorporated into the CMIP5 tool.

To yield solid results, this method requires reliable precipitation observation data of high spatial resolution. All users of the CMIP5 tool have the option of using precipitation data from the Asian Precipitation Highly-resolved Observational Data Integration Towards Evaluation of Water Resources (APRHODITE) project (Yatagai et al., 2012). APRHODITE develops state-of-the-art daily precipitation datasets with high-resolution grids for Asia (0.25° and 0.5° resolutions). The datasets are created primarily with data obtained from a

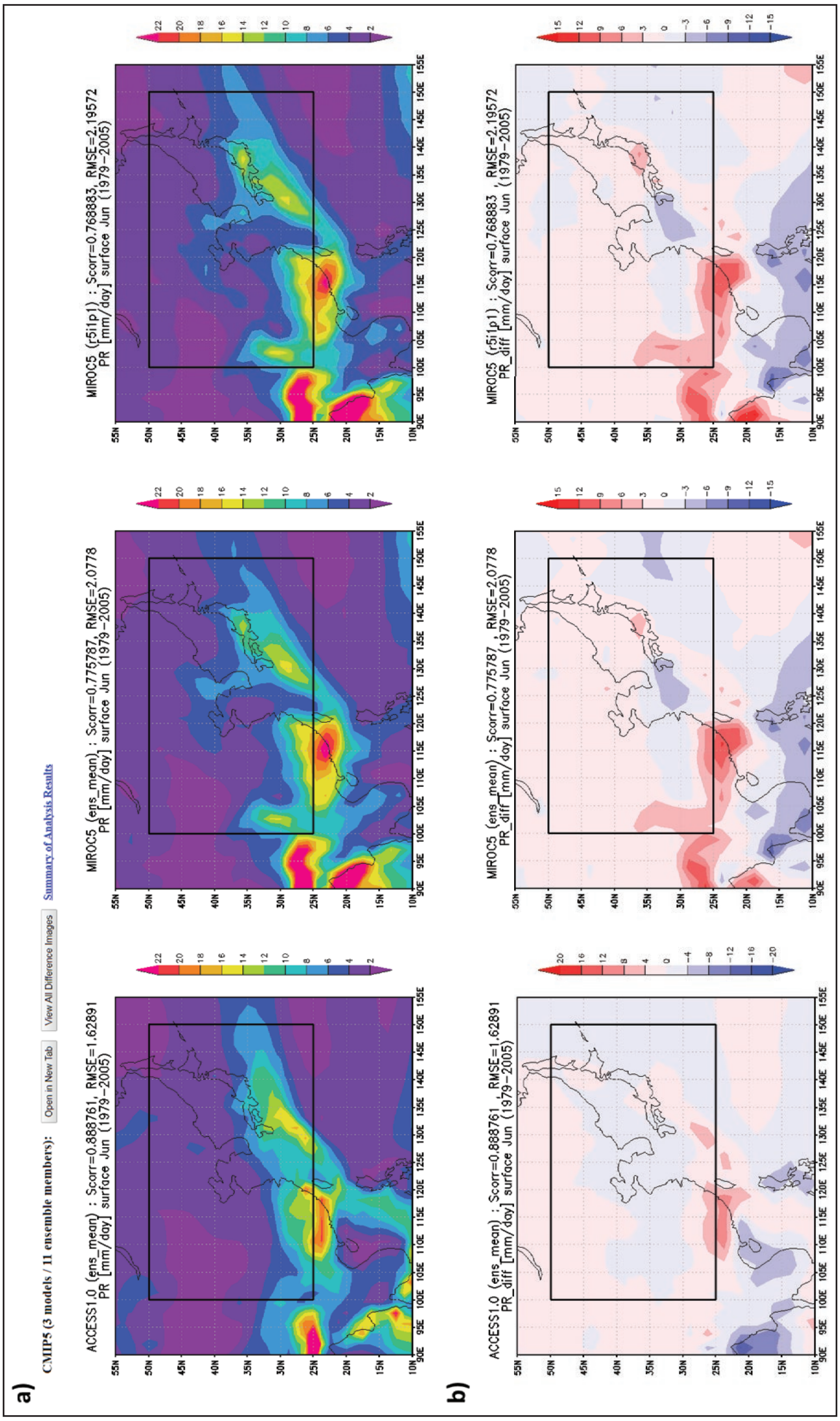

Figure 4: Example of results provided by the 2-D intercomparison function of the DIAS CMIP5 tool. **(a)** 2-D plots of precipitation output of selected ensemble runs of selected GCM(s) for the selected analysis area (black box). The variable is the average daily precipitation in June for the period 1979–2005 (historical simulation experiment of CMIP5). The display area (whole plot) can be defined by the user. The tool also provides analysis results by comparisons of model output against reference data over the selected analysis area: the root mean square error (RMSE) and spatial correlation coefficient (Scorr) values. In the case of precipitation, reference data come from the observation-based Global Precipitation Climatology Project (GPCP). Climatological parameters are compared with reference data to evaluate model performance. **(b)** 2-D images showing differences between selected model output and reference data for the same precipitation variable as a).

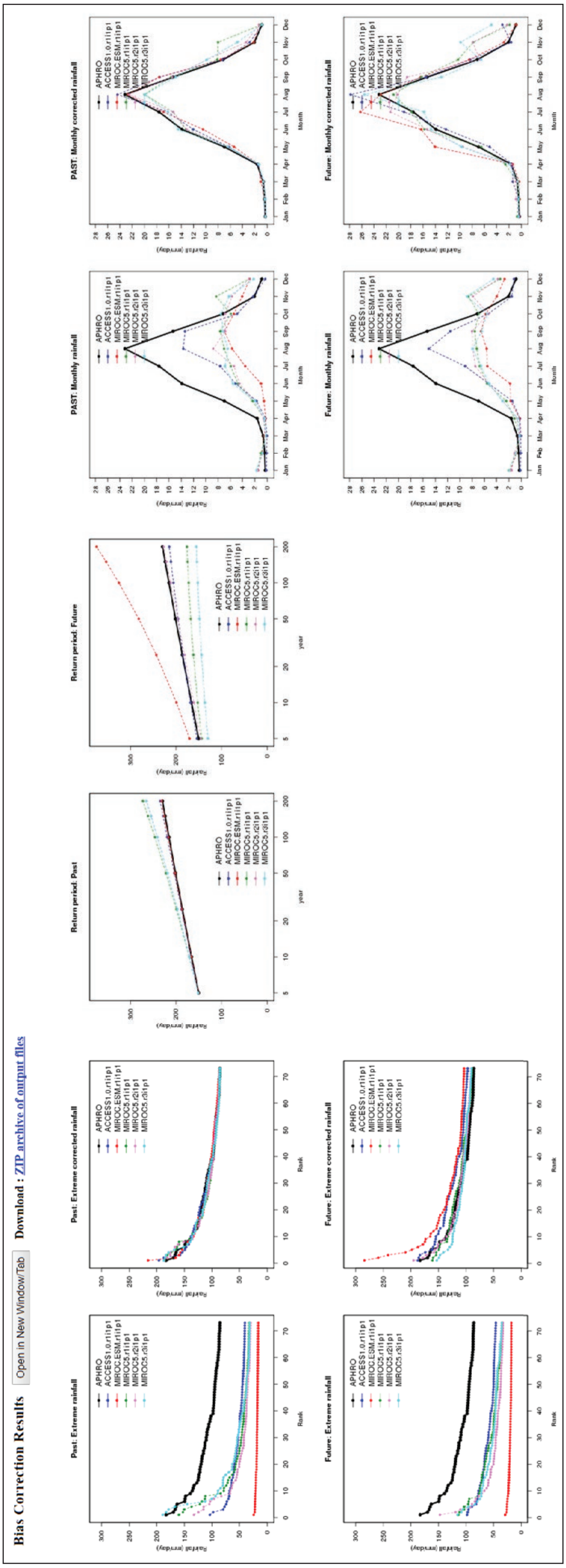

Figure 5: Example of results provided by the precipitation bias correction function of the DIAS CMIP5 tool. The bias-corrected data are downloadable as a ZIP archive. The comparison plots provide an instantaneous look at raw and corrected data (historical simulation and future projection) against observation data used as reference for the bias correction (bold black line on the plots). This example shows results for three selected models, including the available ensembles for each model.

rain-gauge observation network and the resulting gridded product is freely available to all users. The accuracy of APRHODITE data depends on the density of the rain-gauge network, which is used to generate the gridded product. In addition, the spatial resolution of the freely available product may not be sufficient to capture the precipitation distribution at the basin scale, nor do APRHODITE datasets cover the whole globe. The CMIP5 tool, however, offers an option to use in situ precipitation observations to correct for bias in GCM precipitation outputs. This option is only available to users who provide their own in situ observation data, and depends on the data provider and the DIAS team having reached agreement regarding data policy. Such functions, dedicated and tailored to individual users, are the result of the strong R&D component of DIAS and its versatile application development platform. The Asian Water Cycle Initiative (AWCI) has used in situ observation data for bias correction in climate change impact assessment studies conducted in AWCI member countries. Section 4 includes more information regarding AWCI.

In 2013, discussions on and preparations for the sixth phase of CMIP were initiated (CMIP6; Meehl et al., 2014). The great interest of the wider scientific community in CMIP6 planning resulted in a wider range of scientific questions and a substantially larger number of proposed and endorsed experiments (Eyring et al., 2016). The CMIP6 simulations commenced in 2016 and will be completed in several years. The total CMIP6 output volume has been estimated to reach approximately 20 PB (Balaji, 2016), which poses a great challenge for data replication and storage. Encouraged by the success of the CMIP3 and CMIP5 tools, the DIAS team is considering the development of a similar tool for the CMIP6 output. However, the actual scope of the experiments to be embraced and the set of provided functions is subject to discussion. The functions available on the CMIP3 and CMIP5 tools, as described above, will be maintained but can also be expanded based on the further requirements of users. While the DIAS CMIP3 tool might be the first instrument to enable easy access to the CMIP data, which were tailored to the needs of the climate impact research community, other similar endeavors have emerged during the CMIP5 phase. One is the Infrastructure for the European Network for Earth System Modeling (IS-ENES) project under the European Union's Seventh Framework Programme for research (FP7), which developed the Climate4impact tool (Plieger et al., 2015)—a platform enabling impact researchers to explore climate data and perform analyses. There are certain similarities as well as differences between the DIAS CMIP5 and Climate4impact tools. DIAS is a centralized system with its own large data archive, and the DIAS CMIP5 tool works with the data stored there. The Climate4impact tool is a data analysis platform built on the distributed Earth System Grid Federation (ESGF) network and accesses data and some services on remote servers of the ESGF data nodes. Regarding the provided services, the data search, selection, and visualization functions are similar in scope, but there are differences in the data processing and analysis functions. Both tools offer subsetting with respect to time and area, 2-D visualization of selected indices over a specific area, and model intercomparison of 2-D plots or time series. In addition, the DIAS CMIP5 tool provides analysis functions such as the comparison of model output with reference data, inter-annual variation (time series and 2-D plots), and statistical trend analysis and rainfall bias correction, as noted above. On the other hand, the Climate4impact tool provides the calculation of climate indices supported by a dedicated wizard and offers statistical downscaling. Also, the Climate4impact tool provides more explanatory text including generic use cases and other valuable advice regarding climate impact studies. This may be one of the next steps for the DIAS team to make the DIAS CMIP5 (and later CMIP6) tool more self-explanatory.

Social problem-solving based on data sharing and networking

One of the main aims of DIAS is to facilitate and support problem-solving by integrating the assessment of social problems with relevant data. In this section, we focus on the water cycle and water resources, since the provision of ample and clean water is essential for reducing poverty, promoting health, ensuring food and energy security, and maintaining healthy ecosystems and biodiversity.

Accurate, timely, and long-term water cycle information is essential for integrated water resources management, particularly given the threats posed by climate change. With this in mind, GEOSS AWCI was established in 2007, bringing together 18 Asian countries. A similar collaborative framework was established in Africa in 2009. Several countries and transboundary river basin authorities and organizations are involved in the GEOSS Africa Water Cycle Coordination Initiative (AfWCCI). Recognizing the advanced capabilities of DIAS—data sharing, integration, the translation of data into information relevant to water resource practitioners and policy decision makers—both AWCI and AfWCCI collaborate closely with DIAS. Practically, this means that AWCI and AfWCCI member countries and authorities provide DIAS with meteorological and hydrological in situ observations from selected basins. In turn, DIAS shares satellite observations and GCM output data, as well as analysis and research. Such collaboration would not be possible without a strong commitment from policymakers in the AWCI and AfWCCI countries. To ensure a commitment to ongoing collaboration, the AWCI and AfWCCI research communities have demonstrated the capabilities of DIAS and the benefits of its use to members of government, agencies, and academia. A number of meetings were held in which representatives of government organizations and academia were invited to present outcomes of successful implementations of DIAS as a data integration platform, e.g., the Water–Climate–Agriculture workbench in Cambodia (Monichoth et al., 2014) and the real-time flood prediction and dam operation optimization system in Japan (Shibuo et al., 2016). In addition, selected functions of DIAS were demonstrated at these events, e.g., the support system for in situ data providers to submit data, perform data quality control, and generate metadata and metadata documents compliant with ISO standards. Moreover, training sessions were held regarding the use and applicability of CMIP tools, which were open to government organizations and their nominees. These face-to-face meeting opportunities were crucial to the continued commitments and collaboration. Recognition of the AWCI and AfWCCI regional collaborative frameworks by the international GEO community has also been very important for ensuring these commitments from member-country policymakers. Under the AWCI and AfWCCI frameworks and in collaboration with DIAS, programs for system development and capacity-building have been launched in Asia and Africa. In these programs, global earth observations are effectively used to develop flood-control/drought-prevention measures, water environment improvements, and climate change adaptations.

One AWCI activity contributing to social issue solutions has involved assessing the impact of climate change on water resources, i.e. how precipitation patterns will change in future and how the frequency of floods and droughts will be affected as a result. This information and analysis is essential for drafting river basin management plans and adaptation measures. Implementation of this activity included:

1. Selection of a target demonstration basin in each country.
2. Collection of long-term meteorological and hydrological in-situ data for the target river basin.
3. Development of hydrological simulation technologies suitable for the given area using the Water and Energy Budget based Distributed Hydrological Model (WEB-DHM; Wang et al., 2009; Koike et al., 2015).
4. Selection of an ensemble of GCM outputs suitable for the given region, execution of the GCM precipitation bias correction (using the statistical method mentioned in Section 3), and downscaling using the in-situ precipitation data from the given basin.
5. Running the WEB-DHM system using the GCM meteorological variables as forcing data for past (1981–2000) and future (2046–2065) periods.
6. Analysis of climate change trends for precipitation and discharge, including for extreme events (floods and droughts).
7. Holding a workshop in a specified country with representatives of government and other relevant institutions and agencies.
8. Holding an AWCI workshop at which member states discuss and share experiences of climate change impacts in individual countries and across regions.

The DIAS CMIP3 and CMIP5 tools have enabled such studies. Aside from pilot research and demonstration studies, comprehensive assessments of the impact of climate change on water resources have been requested in the Philippines (Pampanga River basin; Jaranilla-Sanchez et al., 2013; JICA Final Report, 2013), Japan (Yoshino River basin), and Indonesia (Musi River basin). These assessments have formed part of broader national studies on water resources and planning for the future in this regard.

Similarly, under the auspices of AfWCCI, government agencies and research institutions in African nations are promoting climate change analysis and adaptation. The GEOSS Joint Asia-Africa Water Cycle Symposium, held in Tokyo in November 2013, brought together approximately 200 people involved in AWCI and AfWCCI. Stakeholders had the opportunity to consider water-related issues in Asia and Africa, share knowledge and experience, and contribute to meaningful debate on integrated water resources management and watershed management.

A case study: climate change analysis and reducing flood risk in Myanmar

Rapid urbanization and the effects of climate change mean there is urgent need for Myanmar to reduce the risk of floods in the country (Htut et al. 2014; Win et al. 2015; Kawasaki et al. 2017). These twin challenges will lead to a greater number of devastating floods if left unchecked. Scientific analysis and planning that considers the effects of current flood characteristics and climate change, and the ways in which the flood risk will change, is required. With such analysis, decision makers and government agencies can make informed and careful plans for development, while working to reduce the flood risk and adapt to future scenarios. In addition, local, regional and national disaster management systems in Myanmar should be formulated on a scientific basis, especially given the potential effects of climate change.

Climate change analysis requires the use, storage and processing of big data. **Figure 6** shows the research framework used in an on-going climate change study (Myanmar). There are three main steps in climate change analysis:

1) production of high-resolution data (concerning precipitation, mainly)
2) bias correction (precipitation and other variables)
3) local impact assessment

Central to climate change analysis for flood risk reduction is the provision of high-resolution precipitation data. This data is necessary for understanding flood hazards in specific regions and river basins. High-resolution precipitation is produced using dynamic downscaling through the usage of a regional climate model (RCM), in this case the Weather Research and Forecasting model. Using GCM model outputs from the CMIP5 project in DIAS as forcings to RCMs, high-resolution past/present and future climate projection data for a 30 year period each. However, there are high computational needs and huge data storage needs (several GB of space to ~100 GB per 1 year simulation of the ~25 km grid resolution) in using RCMs to downscale precipitation. Also, the production of climate projection data face two main issues related to RCMs for downscaling and GCM outputs as inputs to RCMs.

Domain settings and sub-grid physics selection in RCMs, which previous studies have shown to be highly sensitive in simulating past events, are especially important in setting-up the model correctly. So, multi-observation comparison (**Figure 6**) with simulation results is needed to validate the high-resolution (downscaled) products. This ensures the model is capable of reproducing past simulation events, and its correct set-up. When not set properly, these two components can add to the uncertainty involved in climate projection outputs.

Before GCM outputs are used as forcing for producing high-resolution data, GCMs are selected using the DIAS CMIP5 tool, which analyzes the capability of the GCM to reproduce the historical run. This ensures that the selected GCMs have the ability to adequately reproduce the past/present climate over Myanmar region. GCMs are selected are based on a scoring procedure (**Figure 7**) for capturing the relative distribution and absolute value of the precipitation (including other variables defined important for Myanmar climate). The selection of GCMs limits the uncertainty in the climate projections data. Due to GCMs coarse resolution and usage of RCMs, the downscaled historical (past/present) runs can still have biases as compared to the observations. The biases of the high-resolution products are minimized through bias-correction methods. After setting up WRF model and selecting GCMs, high-resolution data (historical and future projections) are produced based on the Pseudo-Global-Warming (PGW) method of F. Kimura and A. Kitoh (2007).

On production of the 30-year historic and future climate projection runs, bias correction of products such as precipitation and temperature is needed before these can be used as inputs for impact assessment. This requires computational needs and storage, but not to the level of the previous step. Bias correction is needed due to the coarse resolution of the GCMs and inherent errors therein. Thereafter, hydrologic (Water and Energy Budget Distributed Hydrological Model, WEB-DHM) and inundation models can be used for local scale assessment: this is necessary in order to understand how floods in the past and floods in the future may differ.

The DIAS environment provides the model running and model evaluation procedure of the downscaling precipitation component of the framework. The DIAS environment analysis cluster is used to run multiple cases considered in the WRF model set-up and GCM selection procedure. Bias correction analysis, hydrologic modeling and inundation modeling are done on the analysis cluster and archives allow for validation. The DIAS environment links data storage needs, access to archived datasets, and computational demands, and as such provides the analytical backbone of the downscaling precipitation and basin scale impact assessment.

Essentially, the key to climate change analysis is to have a platform that can provide multi-observation and reanalysis data, as well as GCM outputs (CMIP5 climate projections), and handle storage needs and high computational demands. The DIAS environment meets all these requirements. Without such a platform, it

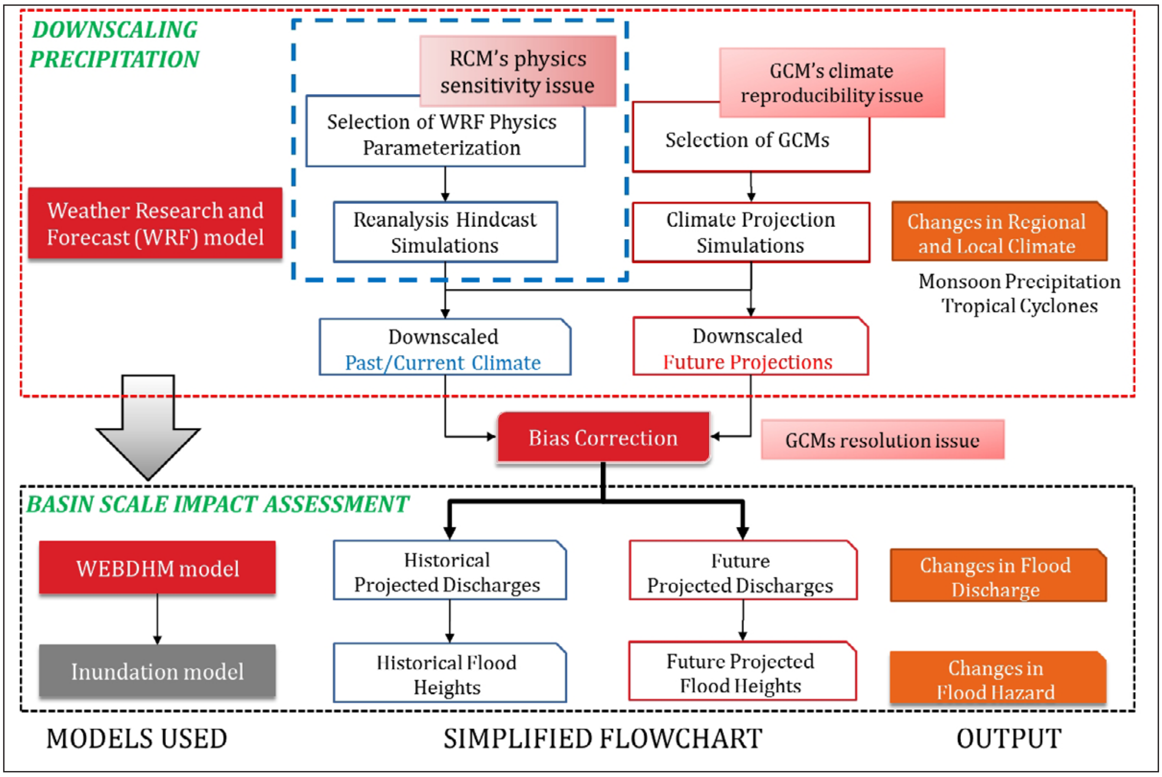

Figure 6: Research framework for climate change analysis relevant to water-related disasters.

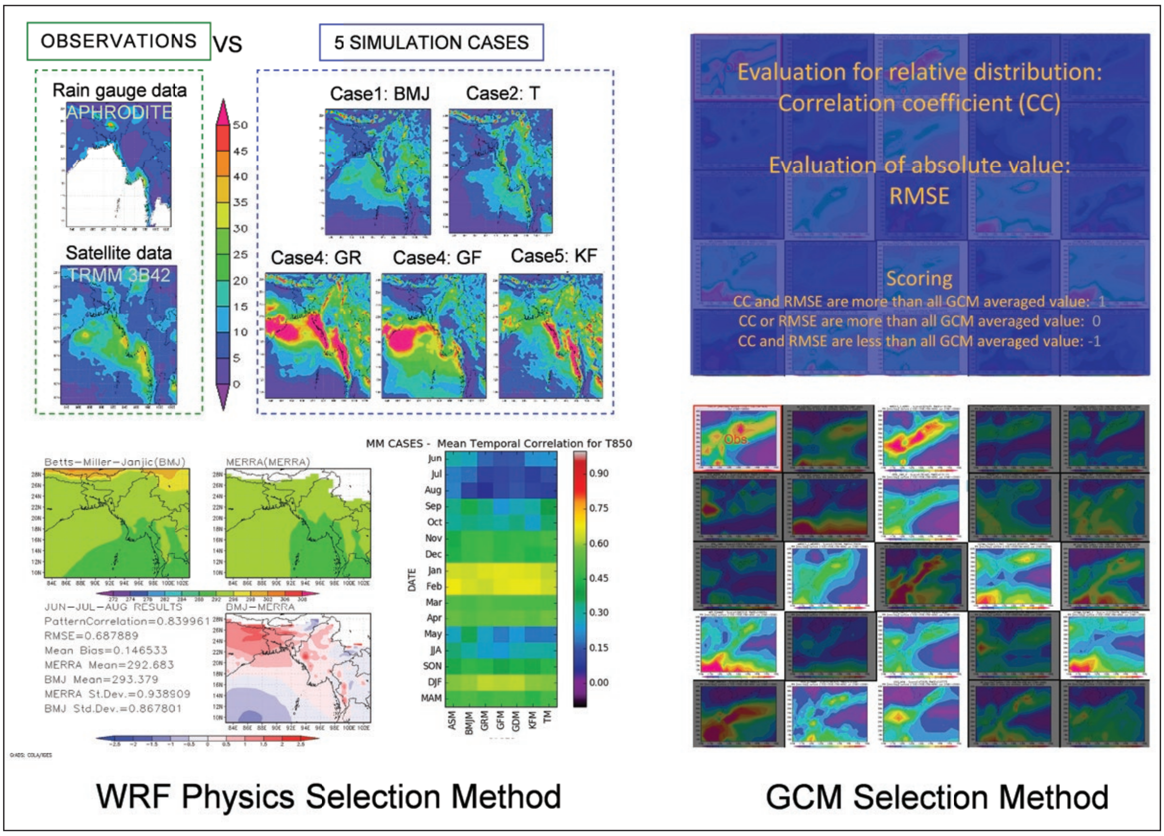

Figure 7: Components of climate change analysis that use the DIAS environment (cluster analysis, data storage, some archived datasets).

would take more time and effort to assess the impact of climate change. The study in Myanmar is testament to this, and the effectiveness of the DIAS platform.

Conclusion

With regard to research on the global environment, DIAS is involved in all aspects of the data lifecycle—measurement, retrieval, analysis, archiving, and sharing of information. Although there are other specialized data measurement/management infrastructure systems, DIAS is distinctive in the way it integrates the infrastructure system, application development, and R&D community to facilitate interdisciplinary and transdisciplinary research and support solution development with a broad cross-sectorial approach. Over the 10 years of DIAS' implementation, the DIAS team has learned a number of lessons that may be useful to those undertaking other similar initiatives:

- An advanced and sustainable infrastructure is a prerequisite for a data platform, but one of the keys to DIAS' success is its strong R&D community originating from DIAS partner organizations and the established links between DIAS and a wide range of users through a number of projects.
- Successful coordination of involved partners requires continued discussion among the DIAS R&D members, i.e., the IT experts and domain scientists of the partner organizations. For this purpose, the DIAS R&D meeting function was established, which includes regular teleconferences, videoconferences, and face-to-face meetings.
- To maintain the interest of the involved partners and funding institutions, it was important to deliver actual services and benefits shortly after launching the first phase of DIAS. This was ensured by initially focusing on a smaller set of tools and functions (e.g., the interoperable data archive; the advanced data visualization system with intercomparison functions; the in situ data upload, quality check, and metadata registration tools for data providers; and the CMIP3 tool), which were in high demanded by the research community and central to the further development of DIAS. These tools were used in pilot studies and activities and the resulting benefits have been demonstrated at project meetings and scientific conferences.
- The DIAS approach regarding the scope and development of new systems and tools has also been vital to its success. It is important to consider the requirements of the user community when planning new developments, but it is also essential to focus on tools with a high added value, i.e., tools providing capabilities that cannot be implemented without the DIAS-integrated data platform. Therefore, proposals for new developments are assessed at the DIAS R&D meeting against the DIAS criteria, including a high scientific challenge, involvement of big data analysis, and interdisciplinarity and/or transdisciplinarity. The actual development of an approved system/tool then becomes an interactive process between the domain scientists and the IT experts of the DIAS R&D team.
- Ensuring easy data search and data interoperability is a basic requirement for data integration systems. Dedicated efforts focusing on interoperability management have been crucial to the success of DIAS. DIAS developed an interoperability portal that provides metadata searches with various vocabularies and the metadata itself is developed based on ISO 19115 and ISO 19139 standards. Further details can be found in, e.g., Nagai et al. (2014) and Takahashi et al. (2010).
- The benefits of the system services to the scientific community and society at large must be communicated regularly to relevant stakeholders. To this end, it is important to invite stakeholders to appropriate project meetings and to organize sessions devoted to the demonstration of the system capabilities and the explanation of its benefits. In particular, presentations of concrete outcomes of successful cases in which DIAS has been used have had great impact. Such demonstrations are also helpful in securing further funding.

Political decision-making based on data has a decisive role to play in future climate change analysis and disaster risk reduction. In this regard, funding for platforms such as DIAS must be ongoing. A commitment to DIAS and other integrated platforms will also lead to greater community understanding of climate change and its effects. DIAS combines analysis methods, data assimilation, and prediction models from across fields and disciplines. In this way, DIAS aims to initiate cooperation between a variety of stakeholders, and contribute to the creation of scientific knowledge and the goal of sustainable development.

Acknowledgements
The authors are grateful to the EDITORIA science team for their support of the DIAS project, and thank those who provided data for climate change analysis. The DIAS project was funded by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT). The analysis in Section 5 was supported, in part, by the Japan Science and Technology Agency (JST)/Japan International Cooperation Agency (JICA), and the Science and Technology Research Partnership for Sustainable Development Program (SATREPS).

Competing Interests
The authors have no competing interests to declare.

References
Aitsi-Selmi, A, Murray, V, Wannous, C, Dickinson, C, Johnston, D, Kawasaki, A, Stevance, A-S, Yeung, T, et al. 2016 Reflections on a science and technology agenda for 21st century disaster risk reduction. Based on the scientific content of the 2016 UNISDR Science and Technology Conference on the Implementation of the Sendai Framework for Disaster Risk Reduction 2015–2030. *International Journal of Disaster Risk Science*, 7(1): 1–29. DOI: https://doi.org/10.1007/s13753-016-0081-x

Balaji, V 2016 Status of Infrastructure for CMIP6. Oral presentation at the 20th Session of WGCM, Princeton University, New Jersey, USA, 1–2 November 2016. Retrieved from: www.wcrp-climate.org/images/modelling/WGCM/WGCM20/documents/04_Balaji_wgcm20-vb.pdf (accessed on 27 April 2017).

Belmont Forum e-Infrastructures and Data Management Collaborative Research Action Steering Committee 2015 A Place to Stand: e-Infrastructures and Data Management for Global Change Research: Belmont Forum e-Infrastructures & Data Management Community Strategy and Implementation Plan.

Future Earth 2013 Future Earth Initial Design: Report of the Transition Team . Paris: International Council for Science (ICSU).

Gilbert, R O 1987 Statistical Methods for Environmental Pollution Monitoring, Wiley, NY.

Htut, A Y, Shrestha, S, Nitivattananon, V and **Kawasaki, A** 2014 Forecasting climate change scenarios in the Bago River Basin, Myanmar. *J. Earth Sci. & Clim. Change*, 5(9). DOI: https://doi.org/10.4172/2157-7617.1000228

IPCC 2007 Climate Change 2007: The Physical Science Basis. In: Solomon, S, Qin, D, Manning, M, Chen, Z, Marquis, M, Averyt, K B, Tignor, M and Miller, H L (Eds.) *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA.

IPCC 2013 Climate Change 2013: The Physical Science Basis. In: Stocker, T F, Qin, D, Plattner, G-K, Tignor, M, Allen, S K, Boschung, J, Nauels, A, Xia, Y, Bex, V and Midgley, P M (Eds.) *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 1535 pp. DOI: https://doi.org/10.1017/CBO9781107415324

Japan International Cooperation Agency (JICA) 2013 Final Report: The Republic of the Philippines, The Study of Water Security Master Plan for Metro Manila and Its Adjoining Areas. Climate Change Impact Assessment and Hydrological Simulation.

Jaranilla-Sanchez, P A, Koike, T, Nyunt, C T, et al. 2013 Hydrological impacts of a changing climate on floods and droughts in Philippine river basins. *Ann J Hydraulic Hydraul Eng-JSCE*, 57: 13–18. DOI: https://doi.org/10.2208/jscejhe.69.I_13

Kawasaki, A, Ichihara, N, Ochii, Y, Acierto, R A, Kodaka, A and **Win, W Z** 2017 Disaster response and river infrastructure management during the 2015 Myanmar floods: a case in the Bago River Basin. *International Journal of Disaster Risk Reduction*, 24, 151–159. DOI: https://doi.org/10.1016/j.ijdrr.2017.06.004

Kendall, M G 1975 Rank Correlation Methods, 4th edition, Charles Griffin, London.

Kimura, F and **Kitoh, A** 2007 Downscaling by pseudo global warming method. In: *The Final Report of the ICCAP*. Research Institute for Humanity and Nature (RIHN): Kyoto, Japan.

Koike, T, Koudelova, P, Jaranilla-Sanchez, P A, et al. 2015 River management system development in Asia based on data integration and analysis system (DIAS) under GEOSS. *Science China: Earth Sciences*, 58: 76–95. DOI: https://doi.org/10.1007/s11430-014-5004-3

Mann, H B 1945 Non-parametric tests against trend. *Econometrica*, 13: 163–171. DOI: https://doi.org/10.2307/1907187

Meehl, G A, Boer, G J, Covey, C, Latif, M and **Stouffer, R J** 2000 The Coupled Model Intercomparison Project (CMIP). *Bull. Amer. Meteor. Soc.*, 81: 313–318. DOI: https://doi.org/10.1175/1520-0477(2000)081<0313:TCMIPC>2.3.CO;2

Meehl, G A, Covey, C, Delworth, T, Latif, M, McAvaney, B, Mitchell, J F B, Stouffer, R J and **Taylor, K E** 2007 The WCRP CMIP3 multi-model dataset: A new era in climate change research. *Bull Am Meteorol Soc*, 88: 1383–1394. DOI: https://doi.org/10.1175/BAMS-88-9-1383

Meehl, G A, Moss, R, Taylor, K E, Eyring, V, Stouffer, R J, Bony, S and **Stevens, B** 2014 Climate Model Intercomparisons: Preparing for the Next Phase. *Eos Trans. AGU*, 59: 77. DOI: https://doi.org/10.1002/2014EO09001

Monichoth, S I, Tsujimoto, K, Aida, K, et al. 2014 Water and Food Security under Climate Change in Cambodia. *Trans. JSASS Aerospace Tech. Japan*, 12(29).

Moss, R, Babiker, M, Brinkman, S, Calvo, E, Carter, T, Edmonds, J, Elgizouli, I, Emori, S, Erda, L, Hibbard, K, Jones, R, Kainuma, M, Kelleher, J, Lamarque, J F, Manning, M, Matthews, B, Meehl, J, Meyer, L, Mitchell, J, Nakicenovic, N, O'Neill, B, Pichs, R, Riahi, K, Rose, S, Runci, P, Stouffer, R, van Vuuren, D, Weyant, J, Wilbanks, T, Pascal van Ypersele, J and **Zurek, M** 2008 *Towards New Scenarios for Analysis of Emissions, Climate Change, Impacts, and Response Strategies*. Intergovernmental Panel on Climate Change, Geneva, 132 pp.

Nagai, M, Rajbhandari, S, Ono, M and **Shibasaki, R** 2014 Earth Observation Data Interoperability Arrangement with Vocabulary Registry. In: Kawtrakul, A, et al. (Eds.) *ISIP2013*. CCIS 421, pp. 128–136. DOI: https://doi.org/10.1007/978-3-319-08732-0_10

Nakicenovic, N and **Coauthors** 2000 *IPCC Special Report on Emissions Scenarios*. Cambridge University Press, 599 pp.

Nyunt, C T, Koike, T, Jaranilla-Sanchez, P A, et al. 2013 Bias correction method for climate change impact assessments in the Philippines. *Ann J Hydraulic Hydraul Eng-JSCE*, 57.

Nyunt, C T, Koike, T and **Yamamoto, A** 2016 Statistical bias correction for climate change impact on the basin scale precipitation in Sri Lanka, Philippines, Japan and Tunisia. *Hydrol. Earth Syst. Sci. Discuss.* DOI: https://doi.org/10.5194/hess-2016-14

Plieger, M, de Cerff, W S, de Vreede, E, Page, C, Tatarinova, N, Cofino, A, Vega Saldarriaga, M, Hutjes, R, de Jong, F, Barring, L and **Sjokvist, E** 2015 Bridging CMIP5 and CORDEX data infrastructure to impact users. European Geosciences Union General Assembly, 12–17 April 2015, Session ESSI2.1.

Rasmy, M, Koike, T, Lawford, P, Hara, M, Fujita, M and **Kimura, F** 2015 Assessment of future water resources in the Tone river basin using a combined dynamical-statistical downscaling approach. *Ann J Hydraulic Hydraul. Eng.-JSCE*, 59: I_73–I_78.

Sawada, Y, Koike, T and **Walker, J P** 2015 A land data assimilation system for simultaneous simulation of soil moisture and vegetation dynamics, *J. of Geophys. Res. – Atmospheres*, 120: 5910–5930.

Seto, R, Koike, T and **Rasmy, M** 2013 Analysis of the vertical structure of the atmospheric heating process and its seasonal variation over the Tibetan Plateau using a land data assimilation system, *J. Geophys. Res. Atmos.*, 118. DOI: https://doi.org/10.1002/2013JD020072

Shibuo, Y, Ikoma, E, Saavedra, O C, Wang, L, Koudelova, P, Kitsuregawa, M and **Koike, T** 2014 Development of operational realtime ensemble flood forecast system. Journal of Japan Society of Civil Engineers. *Ser. B1 (Hydraulic Engineering)*, 70(4): 397–402. DOI: https://doi.org/10.2208/jscejhe.70.I_397

Shibuo, Y, Ikoma, E, Saavedra Valeriano, O, Wang, L, Lawford, P, Kitsuregawa, M and **Koike, T** 2016 Implementation of real-time flood prediction and its application to dam operations by data integration analysis system. *J. of Disaster Res.*, 11(6): 1052–1061. DOI: https://doi.org/10.20965/jdr.2016.p1052

Takahashi, A, Tatedoko, M, Shimizu, T, Kinutani, H and **Yoshikawa, M** 2010 Metadata Management for Integration and Analysis of Earth Observation Data. *J. of Software*, 5(2): 168–178.

Taylor, K E, Stouffer R J and **Meehl, G A** 2012 A summary of the CMIP5 experiment design. *Bull. Am. Meteorol. Soc.*, 93: 485–498. DOI: https://doi.org/10.1175/BAMS-D-11-00094.1

van Vuuren, D P, Edmonds, J, Thomson, A, Riahi, K, Kainuma, M, Matsui, T, Hurtt, G C, Lamarque, J-F, Meinshausen, M, Smith, S, Granier, C, Rose, S K and **Hibbard, K A** 2011 The Representative Concentration Pathways: an overview, *Climatic Change*, 109: 5–31. DOI: https://doi.org/10.1007/s10584-011-0148-z

Wang, L and **Koike, T** 2009 Comparison of a distributed biosphere hy-drological model with GBHM. *Ann J Hydraulic Hydraul Eng-JSCE*: 103–108

Win, W Z, Kawasaki, A and **Win, S** 2015 River Flood Inundation Mapping in the Bago River Basin, Myanmar. *Hydrological Research Letter*, 9(4): 97–102

Yasukawa, M, Kitsuregawa, M, Taniguchi, K and **Koike, T** 2008 PVES: Powered Visualizer for Earth Environmental Science. *IEEE Systems Journal*, 2(3). DOI: https://doi.org/10.1109/JSYST.2008.925980

Yatagai, A, Kamiguchi, K, Arakawa, O, Hamada, A, Yasutomi, N and **Kitoh, A** 2012 APHRODITE: Constructing a Long-Term Daily Gridded Precipitation Dataset for Asia Based on a Dense Network of Rain Gauges. *Bull. Amer. Meteor. Soc.*, 93, 1401–1415. DOI: https://doi.org/10.1175/BAMS-D-11-00122

How to cite this article: Kawasaki, A, Yamamoto, A, Koudelova, P, Acierto, R, Nemoto, T, Kitsuregawa, M and Koike, T 2017 Data Integration and Analysis System (DIAS) Contributing to Climate Change Analysis and Disaster Risk Reduction. *Data Science Journal,* 16: 41, pp. 1–17, DOI: https://doi.org/10.5334/dsj-2017-041

Submitted: 31 October 2016 **Accepted:** 26 May 2017 **Published:** 04 September 2017

]u[*Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS