
An attempt to integrate huge scientific data with interoperability in Japan

Seishi Ninomiya

University of Tokyo

Facing rather poor status of data sharing among scientific community of Japan, Council for Science and Technology Policy, Japan (CSTP, <http://www8.cao.go.jp/cstp/english/index.html>) has been keeping its strong policy to develop databases as a basic platform for researches in last decade. Several funds have been being provided for developing scientific databases by Ministry of Agriculture, Forestry and Fishery (MAFF, <http://www.maff.go.jp/e/index.html>), Ministry of Education, Culture, Sports, Science and Technology (MEXT, <http://www.mext.go.jp/english/>), Japan Society for Promotion of Science (JSPS, <http://www.jsps.go.jp/english/index.html>), Japan Science and Technology Agency (JST, <http://www.jst.go.jp/EN/index.html>), etc.. Actually, several databases have been developed and some of them are targeted for agricultural researches.

Though data sharing and interoperability of databases are always highly requested under such a policy, particularly the interoperability issue was not overcome in most of the cases. Amongst them, Data Integration and Analysis System (DIAS) is an exception. The DIAS project has been being conducted since 2005 under a strong support of MEXT to follow the above policy. It is also conducted as a part of the Japanese contribution to GEOSS (Global Earth Observation System of Systems)

The missions of DIAS are;

- to coordinate the cutting-edge information science and technology and the various research fields addressing the earth environment
- to construct data infrastructure that can integrate earth observation data, numerical model outputs, and socio-economic data effectively
- to create knowledge enabling us to solve the earth environment problems
- to generate socio-economic benefits.

DIAS is tackling a large increase in diversity of the earth observation data by

developing a core system (Fig. 1) for data integration and analysis that includes the supporting functions of life cycle data management, data search, information exploration, scientific analysis, and partial data down-loading. For improving data interoperability, DIAS is developing a system for identifying the relationship between data by using ontology on technical terms and ideas, and geography. DIAS

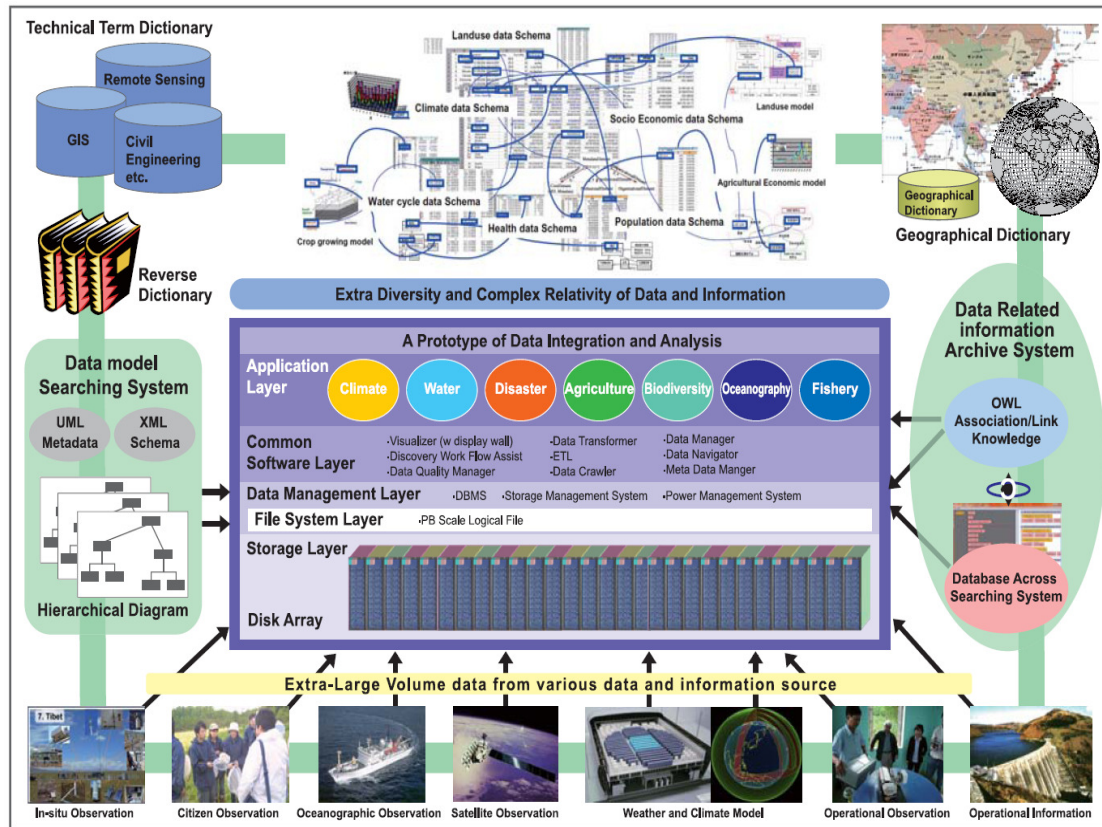


Fig. 1 Core system of DIAS

also is acquiring data base information from various sources by developing a cross-sectoral search engine for various databases.

By utilizing the core system of DIAS, we are to develop several application to contribute to socio-economic benefits;

- Agricultural Production Management

DIAS develops an information system for agricultural production managements by integrating the real-time monitoring data of farmland, the growing condition of each crop cultivar, meteorological data, numerical weather predictions, and climate model predictions. This system will be usable by the farming community, enabling

them to make improved management decisions especially in regions which are susceptible to global warming impacts.

- Ecosystem Conservation and Participatory Monitoring Program

DIAS compiles data bases of a number of important indices of biodiversity, including invasive alien species and endangered species through participatory monitoring programs, integrates to analyze the data with other earth observation data, and disseminates the products in a form

to be easily used for decision making related to biodiversity conservation..

- Integrated Water Resources Management

The Asian countries cooperatively integrate data from earth observation satellites and in-situ networks with other types of data, including numerical weather prediction model outputs, geographical information, and socio-economic data, to generate information for making sound

water resources management decisions.

- Ocean Circulation and Fishery Resources Management

DIAS provides usable information for a sustainable fishery resources management by constructing an oceanography- fishery cooperative platform that enables resource managers to investigate relationships between fluctuations in the fishery resources and the seasonal to decadal ocean variations derived from an ocean re-analysis based on the data assimilation by applying the four dimensional variational assimilation methods.

DIAS contains ca. 1.5 PB data and a part of DIAS database has been open since October, 2010. In addition to the DIAS projects, several database developments are now going on. Table 1 shows some typical databases in Japan useful for agricultural researches. DIAS, MetBroker and YMC are databases being developed under research projects while NIAES and NIES are collections of databases supported by institutions. Such institutional databases were also originally developed under national research projects. AGROPDIA and MAFF STAT are databases managed by MAFF, Japan.

Table 1 Typical databases for agricultural research in Japan

DB names	Information type	Subject scope	amount of digital information/data content	ownership	Access	management standards,	URL
DIAS/Data Integration and Analysis System	Climatic scenario data, satellite image data, GPV data, sensor network data, water resource data, downscaling data	Climate change, integrated water resource management, agricultural production management, environment studies, ocean circulation and	1-2PB	U. Tokyo	half open	ontology based search engine for data interoperability	http://www.editoria.u-tokyo.ac.jp/dias/english/index.html
MAFF/Ministry of Agriculture, Fishery and Forestry Statistics	Agricultural statistics mainly of Japan	Not defined	unknown	MAFF	open, only in Japanese	spread sheet	http://www.maff.go.jp/tokai/index.html
NIAS (National Institute of Agrobiological Sciences) Databases	Mainly genome data	Genome researches	unknown	NIAS	open	international standard for genome	http://www.nias.affrc.go.jp/index_e.html
NIAES (National Institute of Agro-Environmental Sciences) database	land use, environmental data, weather, crop calendar, CO2 flux, soil	Agroenvironmental studies		NIAES	open, only in Japanese		http://www.niaes.affrc.go.jp/techdoc/db_image.html
AGROPEDIA-MAFFIN (MAFF Research Network)	Documents, research reports, satellite, weather	undefined	800TB	MAFF	partially open, mainly in Japanese		http://www.affrc.go.jp/agropedia/
MetBroker	Observed weather data and predicted weather data (GPV)	Any that uses weather data	300GB (30000 data point of 25 weather databases)	NARC/NARO, JP	open	Wrapper based standardization	http://www.agmodel.org/projects/metbroker.html
YMC (Youth Mediated Communication) System	Agricultural Q&A for rice farmers	Agricultural technology transfer to illiterate farmers through their children	hundreds Q&As	YMC Project/Japan	closed	original	http://www.pangaeon.org/blog/japanese/archives/20110520_PressRelease_YMCWorkshop_ver008.p

Contact:

Seishi Ninomiya

Institute of Sustainable Agro-ecosystem Services,

Graduate School of Agriculture and Life Sciences, the University of Tokyo

snino@isas.a.u-tokyo.ac.jp