



DA-09-02a Task Report

Data Integration and Analysis System

Report by
Toshio Koike, The University of Tokyo, Japan
and
Rick Lawford, University of Manitoba, Canada

ADC Meeting
May 2009
Stresa



DA-09-02a Task Report

(May. 2009)

Sub-task Definition (as given in the 2009-2011 Work Plan):

Coordinate data management approaches that encompass a broad perspective of the observation data life-cycle – from input to processing, archiving, and dissemination, including reprocessing, analysis and visualization of large volumes and diverse types of data.



DA-09-02a Task Report

(May. 2009)

Short Background and objective for DA-09-02a:

GEOSS is being developed to achieve more effectiveness and greater economies in the operation of observational systems through convergence among global, regional and national facilities toward interoperable and integrated observational and information systems. This can be done through increased coordination and communication between the centres, and the development of standards and best practices that can be shared among the centres, In order to achieve its GEOSS objectives GEO must encourage convergence among information systems. This can be achieved most effectively by the development of meaningful linkages and joint initiatives between data centres.



DA-09-02a Task Report

(May. 2009)

Outputs

The following outputs are anticipated as a result of this subtask:

1. Inventories of the significant data centres in all societal benefit areas.
2. Alliances between different centres with similar objectives (e.g., World Data Centres, Research Data systems, etc)



DA-09-02a Task Report

(May. 2009)

Activities

A Phase I survey has been undertaken and the results are being analyzed.

- WDC for Glaciology and Geocryology, China
- Data Integration and Analysis System (DIAS), Japan
- Ground European Network for Earth Science Interoperations – Digital Repositories (GENESI-DR), ESA
- World Data Center Climate (WDCC), Germany
- NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC), USA
- Goddard Interactive Online Visualization ANd aNalysis Infrastructure (Giovanni), USA
- The Global Observing Systems Information Center (GOSIC), USA
- Global Runoff Data Centre (GRDC), Germany
- World Data Center for Glaciology, USA

August 2009: The analysis of the first survey will be published.

August 2009: The questionnaire for the Water Data Centres will be distributed (Phase II – Water Centres)

November 2009: The first data centre alliance workshop will be held

December 2009: The analysis of the second survey will be carried out



Response to the Survey

The survey was sent to 11 centres.

The response was seen as very successful with nine centres completing the survey (82%).

A preliminary analysis of the review follows.



DA-09-02a Task Report

(May. 2009)

THE SURVEY OF THE CENTRES

1. Name of System or Project:
2. Leading Organization(s):
3. Associated Organization(s):
4. Purposes of the System or Projects:
5. Relation to GEOSS SBAs:
6. Involvement in GEO Tasks (Name Tasks):
7. Brief Description of Archived Data:
8. Brief Description of Data Storage Capabilities:
9. Brief Description of Data Interoperability Function:
10. Brief Description of Data Integration and Analysis Capabilities:
11. Level of use and characteristics of users:
12. Homepage Address:
13. Contact:

Centre	Affiliation	Web Site	Contact
World Data Center for Glaciology, Boulder	National Snow and Ice Data Center (NSIDC)	http://nsidc.org and http://nsidc.org/wdc	rbarry@nsidc.org fetterer@nsidc.org
World Data Centre for Glaciology and Geocryology at Lanzhou	Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences	http://wdcdgg.westgis.ac.cn	lixin@lzb.ac.cn
World Data Center Climate (WDCC)	Max Planck Institut for Meteorology (MPI-M) & Deutsches Klimarechenzentrum (DKRZ)	www.wdc-climate.de	data@dkrz.de
Global Runoff Data Centre (GRDC)	Federal Institute of Hydrology (BfG) Am Mainzer Tor 1 56068 Koblenz, Germany	http://grdc.bafg.de	grdc@bafg.de
The Global Observing Systems Information Center (GOSIC)	National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center's (NCDC) US Global Climate Observing System (GCOS) program.	http://gosic.org	christina.lief@noaa.gov

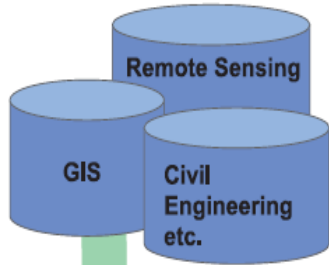
Centre	Affiliation	Web Site	Contact
Goddard Interactive Online Visualization AND aNalysis Infrastructure (GIOVANI)	GES-DISC, a NASA center program operating under the Global Change Data Center (GCDC) at the Goddard Space Flight Center (GSFC)	http://giovanni.gsfc.nasa.gov/ http://disc.sci.gsfc.nasa.gov/	Gregory.leptoukh@nasa.gov
NASA Goddard DAAC	GES-DISC, a NASA center program operating under the Global Change Data Center (GCDC) at the Goddard Space Flight Center (GSFC)	http://disc.gsfc.nasa.gov/	Gregory.leptoukh@nasa.gov
Data Integration and Analysis System (DIAS)	The University of Tokyo	http://www.editoria.u-tokyo.ac.jp/dias/ (Japanese only now)	tkoike@hydra.t.u-tokyo.ac.jp
GENESI-DR: Ground European Network for Earth Science Interoperations – Digital Repositories	European Space Agency, Earth Observation Programmes Directorate, Frascati, Italy	http://www.genesi-dr.eu/	luigi.fusco@esa.int joost.van.bemmelen@esa.int

Brief description of Data Storage Capabilities

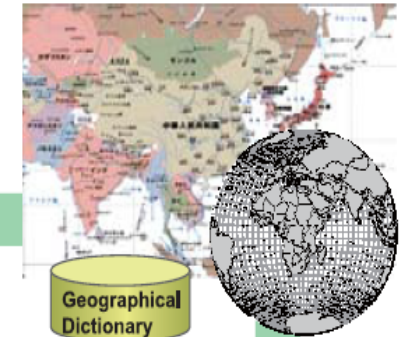
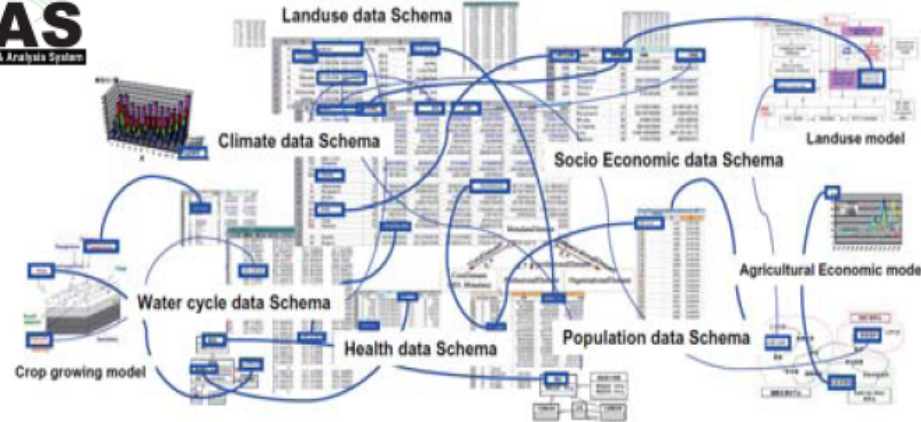


	theme	storage	Data base
WDC Boulder	Glaciology	32TB(RAID)+ 40TB (MAID) 300TB(Tape Storage) +75 TB disk array	<ul style="list-style-type: none"> •Sybase Enterprise Server relational database •MySQL database server
WDC Lanzhou	Glaciology & Geocryology	10TB (RAID)	<ul style="list-style-type: none"> •Catalog metadata database
WDC Hamburg	Climate	Growth of DB system of 100 TB/year will increase to 1 PB/year after 2009	
GRDC: Germany	River Runoff	sufficient storage capabilities	commercial database
GOSIC NOAA/USA	Earth obs. data portal	No storage system but link function	
GIOVANI NASA/USA	Web-based application	20 TB	
GES-DISC NASA/USA	integration and analysis	500 TB Peak daily numbers:3.3 TB for processing 1.7 TB for archiving	
DIAS Japan	integration and analysis	700TB (disk)	
GENESI-DR ESA	e-science infrastructure		OGC Web Services, GridFTP, HTTPS, OPeNDAP, BitTorrent for file-based DRs.

Technical Term Dictionary



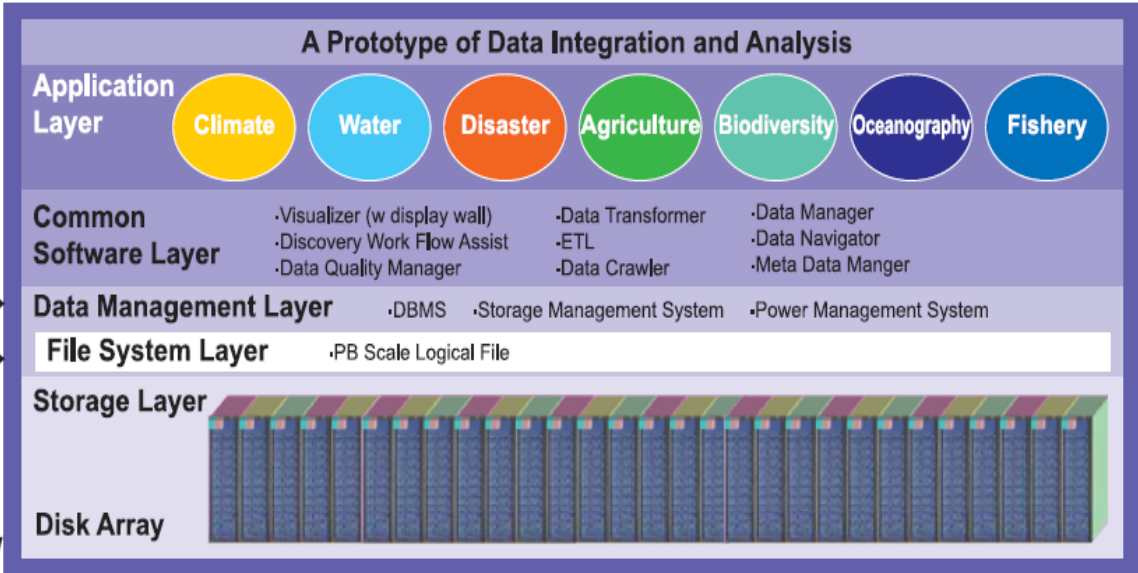
Reverse Dictionary



Geographical Dictionary

Geographical Dictionary

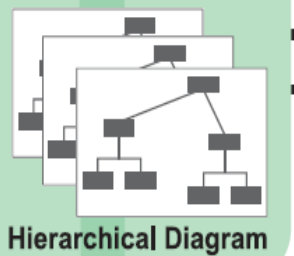
Extra Diversity and Complex Relativity of Data and Information



Data Related information Archive System

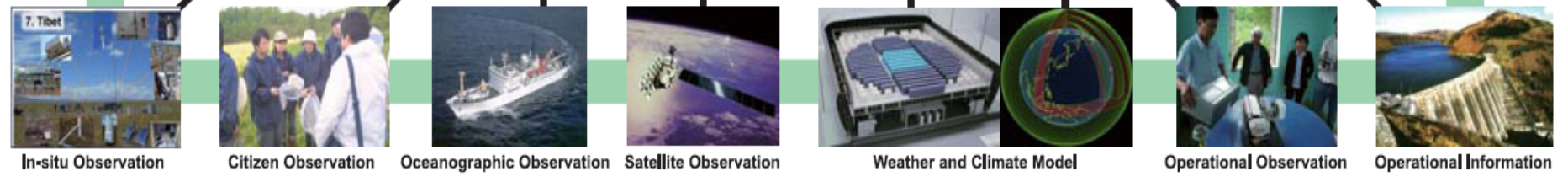


Data model Searching System



Hierarchical Diagram

Extra-Large Volume data from various data and information source



In-situ Observation

Citizen Observation

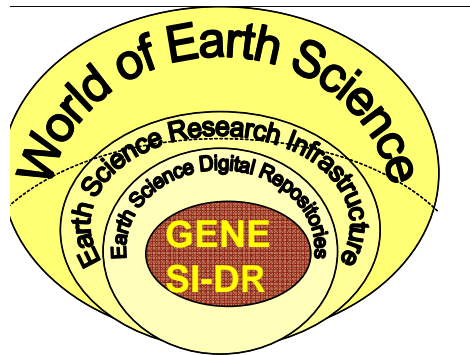
Oceanographic Observation

Satellite Observation

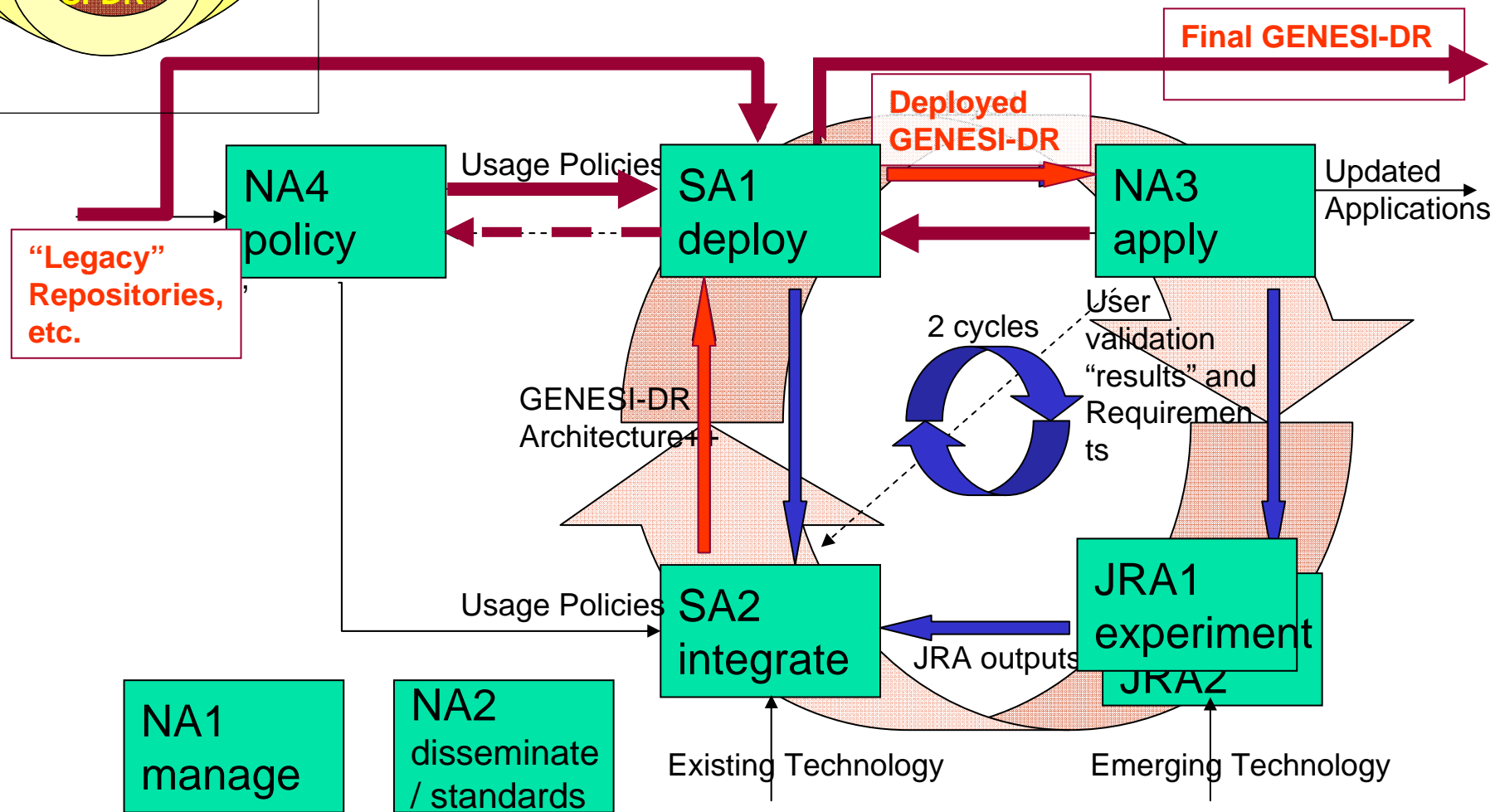
Weather and Climate Model

Operational Observation

Operational Information



The logical sequence for GENESI-DR



From: Luigi Fusco

Preliminary lessons from the analysis

1. The surveys have generated a large amount of relevant information. They have been posted on the CEOP home page.
2. The inputs are not fully consistent from respondent to respondent so it will be necessary to go back to individual contributors to ask for clarification in several cases.
3. There are significant distinctives between those centres which have started recently and make full use of modern “e-technologies” and those which have been operational for many years and adopt new technologies as opportunity arises.



DA-09-02a Task Report

(May. 2009)

- Expected milestones for 2009:
 - A preliminary survey of data centres will be completed.
 - A workshop will be held to explore the feasibility of establishing a small experimental alliance between several centres (likely including Japan (DIAS), Europe (FP-7) and possibly NA (CUASHI)).

Extent of Participation/Recommendations for additional participants:

- The number of large data centres participating in this project must be expanded to ensure it is a success. Other groups that have participated in Phase I and Phase II surveys will be included in the experimental alliance as it grows. In particular, we are planning to involve satellite data archives (e.g. DAACs), ICSU data centres and services, NWP centres and research projects such as CEOP and SERVIR.



DA-09-02a Task Report

(May. 2009)

- Coordination points with other GEO work plan tasks:

The Phase II survey will provide essential background information for Water Tasks and the Water Cycle Community of Practice. This task is being coordinated with SBA tasks that require a high level of data integration such as Task WA-08-01.

- Potential risks that may impede completion of task:

- None identified to date. However, GEO support will be critical to ensure that data centres provide the inputs to the survey and are favourably disposed to participating in the alliance(s).

- Gaps that have been identified that need to be addressed independently:

- Resource requirements are being provided by JAXA, RESTEC and the University of Tokyo. Support from GEO will be needed for the workshops.

- Potential Contributions to the GEO Portal:

- The Data centre registry being developed for the University of Tokyo will be linked to the User Community through the GEO Portal.



DA-09-02a Task Report

(May. 2009)

- Expected products:
 - A comprehensive data base and registry of data centres.
 - A pilot data centre alliance(s) that cover(s) satellite and in-situ data centres, NWP Centres, ICSU data services and research data centres.
 - A framework for future alliances.
 - Reports from several workshops.
- Expected date of completion of task:

September 2010. (Due to a late start on this Task the initial deadline needs to be extended).
- User Engagement and Capacity Building:

Water Cycle Community of Practice including
GEOSS/AWCI, GEOSS/AfWCI, WCRP/GEWEX/CEOP