GEO Water Cycle Data Centre Alliances

Rick Lawford
Toshio Koike

IGWCO Planning Meeting
New York, NY

March 2011

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)

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)

Phase 1 Survey 2009

Activities

A Phase I survey has been undertaken and the results have been 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

Brief description of Data Storage Capabilities

	theme	storage	Data base
WDC Boulder	Glaciology	32TB(RAID)+ 40TB (MAID) 300TB(Tape Storage) +75 TB disk array	•Sybase Enterprise Server relational database •MySQL database server •Catalog metadata database
WDC Lanzhou	Glaciology & Geocryology	10TB（RAID)	
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.

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.

Steps to developing an alliance:

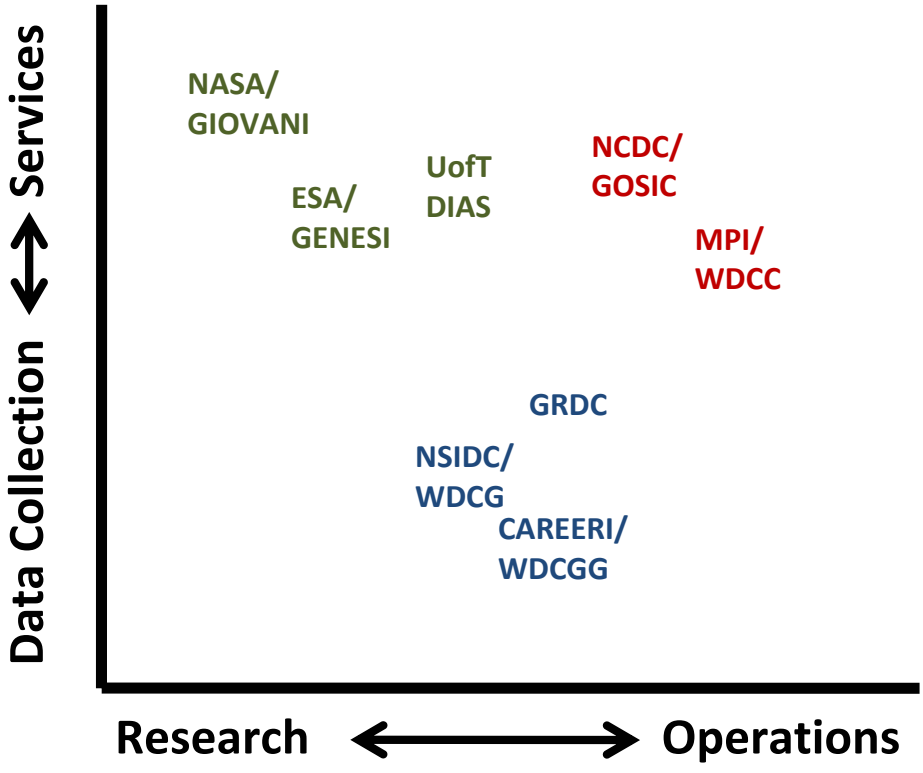
1. The candidate centres will be invited to be part of an alliance. Each alliance will form an Alliance team and name an Alliance coordinator.

2. The Alliance Team will define a short-term (12-month) projects and, as appropriate, several longer term projects that will contribute to GEOSS.

3. In collaboration with a nominee from the Harmonization SubTask, each Alliance team will map the structure and activities of each centre against the Harmonization model.

4. The Alliance will explore the differences between the centres and propose interfaces that could be developed in order to facilitate better collaboration and interoperability.

Preliminary Classification of Data Centres



Possible Issues to be addressed in the Alliances:

"Data Integration System Development Alliance":
1.Provenance
2.Visualizations
3.Ontologies

"Information Services Alliance"
1.User requirements

"Data Product Development and Archival Alliance"
1. Data transfer from collection point to Global Data Centre

Phase 2 Survey (2010)

Activities

A Phase 2 survey has been undertaken and the results have been analyzed.
Data centers surveyed:

1) Aquastat,FAO's global information on water and agriculture
2) Coordinated Asia-European long-term Observing system of Qinghai–Tibet Plateau (France, China)
3) Hydrologic Information Systems, CUASHI
4) FLUXNET project at NASA's Oak Ridge National Laboratory DAAC
5) Monitoring database water quality and water quantity (Belgium)
6) Global Irrigated Areas Mapping (GIAM) Project (USGS)
7) Global Soil Moisture Data Bank (Rutgers)
8) HYDROWEB (CNES)
9) International Groundwater Resources Assessment Centre (IGRAC)
10) Indian Space Research Organisation
11) National Geophysical Data Center (NOAA)
12) National Water Management System (South Africa)
13) United Nations Environment Programme
14) Global Environment Monitoring System (GEMS) Water Programme
15) National Water Information System (USGS)
16) United Kingdom National River Flow Archive

Table 6: Value-added Data Products and Their Dissemination (#10, 11, 13)

	Global	Regional	National	In-house Analytical Tools	Models	Product Formats	Open Access to Products	Approval before Access	Multiple Acess Points
Aquastat	X	X	X						
CEOP – AEGIS		X		X	X				
CUAHSI		X							
VMM		X	X						
FLUXNET		X			X				
GIAM	X								
GSMD	X	X							
HYDROWEB	X						X		
IGRAC	X								
ISRO		X	X	X	X		X		
NGDC	X	X			X				
Oak Ridge	X				X				
Water Management System		X	X	X	X		X		
UNEP/ GEMS	X	X	X	X	X		X		
NWIS		X	X	X		Fixed; comma delimited; RDB	X	X	X
NRFA			X	X	X	Summaries, reviews, reports	X		

Next Steps:

1)Complete and publish the phase 2 analysis
2)Develop a list of Centres from each SBA to receive the Phase 3 Survey
3) Send out the surveys and complete the analysis.