

**Global Ocean-Atmosphere Prediction and Predictability Network**

**GOAPP Data Management Policy**

Prepared by:

GOAPP Data Management Committee (DMC)  
Jason Chaffey (DFO), Bill Merryfield (GOAPP Investigator), Mathieu Ouellet (DFO),  
Hal Ritchie (GOAPP PI), Keith Thompson (GOAPP PI), Lt (N) Darryl Williams (DND)  
and Susan Woodbury (GOAPP Network Manager)

GOAPP Network Secretariat  
Department of Oceanography  
Room 5660A LSC  
Dalhousie University  
1355 Oxford Street  
Halifax, NS B3H 4J1

Tel: (902) 494-5035

Fax: (902) 494-2885

**Submitted to CFCAS as part of the Annual Progress Report June 26, 2007**  
**Approved by the GOAPP Board of Directors June 21, 2007**

## Table of Contents

|     |   |   |
|-----|---|---|
| 1   | Introduction.....   | 1 |
| 2   | Guiding Principle for Data Management.....                                      | 1 |
| 3   | Issues and Needs.....   | 2 |
| 4   | Data Management Policies .....  | 2 |
| 4.1 | Data Archiving (Issues 1, 2, 3) .....   | 2 |
| 4.2 | Data Access (Issue 4).....  | 3 |
| 4.3 | Data Dissemination to Partners, Public and Private Sector (Issues 5 and 8)..... | 4 |
| 4.4 | Facilitating Collaboration within GOAPP (Issues 6 and 7).....                   | 4 |
| 4.5 | Acknowledgements.....   | 4 |
| 5   | Implementation .....  | 5 |
| 6   | Appendix: GOAPP Definitions.....  | 5 |
| 6.1 | GOAPP Dataset .....   | 5 |
| 6.2 | GOAPP Co-Investigator .....   | 5 |
| 6.3 | GOAPP Collaborator .....  | 6 |
| 6.4 | GOAPP Partner.....  | 6 |
| 6.5 | Non-Participant.....  | 6 |

## **1 Introduction**

The Global Ocean-Atmosphere Prediction and Predictability (GOAPP) Network aims to improve predictions of the ocean and atmosphere on time and space scales of days to decades, and tens of km to global scales. This will be achieved by developing an effective coupled data assimilation and prediction capability.

The proposed research and development is a major challenge that will require advances in scientific understanding of the coupled system and technical developments in modelling and data assimilation. This will be achieved by building on recent Canadian contributions to several major national and international research programs. The research will lead to

- Enhanced modelling and data assimilation capabilities for the ocean and the coupled system;
- A new appreciation of the sources of, and limits to, predictability in the coupled system;
- A better understanding of the physical processes underlying seasonal and longer time-scale predictions made with coupled models;
- Decadal-scale reanalyses of the North Atlantic and North Pacific oceans and a coupled historical forecasting project;
- Analysis techniques and tools for maximizing the utility, and assessing the value, of forecasts.

GOAPP will generate model predictions and also will add value to observation-based products (e.g., time-varying 3D fields of temperature and salinity by combining ARGO profiles with hydrographic climatologies). The model output that requires a data management plan will take one of two forms: routine operational forecasts and reanalyses (e.g., historical reconstructions). The coupled historical forecast project (CHFP) will be a significant computational effort that will produce an archive of coupled model retrospective multi-seasonal forecasts whose skill can be evaluated against past observations and compared with that of an existing non-coupled (two-tier) seasonal forecast system.

## **2 Guiding Principle for Data Management**

The Network Agreement between the Canadian Foundation for Climate and Atmospheric Sciences (CFCAS) and the participating institutions states:

“The guiding principle should be to allow data issuing from the Network research activities to be accessible in a timely manner; and in a form such that the “network data information system” will allow access, which is limited initially to Network Participants and new Collaborators as approved by the Board of Directors”.

### **3 Issues and Needs**

CFCAS, in their official letter of award, asked GOAPP to “note the review committee’s comment on the data management policy.” As part of this review committee’s evaluation, several specific shortcomings in the data management portion of the GOAPP proposal were identified. Some external reviewers also expressed concern that the proposal did not adequately address data management.

Based on the report from the review committee and the comments of the external reviewers we have identified the following list of issues that will be addressed by this data management policy:

1. For how long should the operational forecasts and the reanalyses be archived?
2. Where should the data be archived?
3. What data formats should be used for model archival and dissemination?
4. Who has access to GOAPP products, under what conditions and in what time frame?
5. How, and how quickly, will GOAPP products be made available to partners and the general public?
6. Should GOAPP researchers be required to document their models, products etc. using either FGDC or ISO metadata standards?
7. How do we facilitate interactions between Theme I and Theme II, building on the extensive experience of Theme II researchers in managing large scale data sets?
8. How does GOAPP coordinate its data management policy with that of COMDA (DFO’s new Centre for Ocean Model Development and Applications) and the agencies that will eventually generate operational products for GOAPP’s partners? How does GOAPP interact with international agencies and efforts, e.g., WMO, THORPEX?

### **4 Data Management Policies**

The success of the GOAPP Network will depend heavily upon the timely sharing of data and information. The Network requires an early and continuing commitment to the maintenance, quality assurance, documentation, and distribution of its data sets. The creation of a long-term and high quality data archive will be an important scientific legacy of the Network.

#### **4.1 Data Archiving (Issues 1, 2, 3)**

- (a) Co-investigators will archive "research quality" datasets (defined as “production” model runs, reanalyses, or forecasts serving as a basis for product assessment and scientific investigation) as soon as feasible and no more than six months following production. Questions like “what variable should be stored and at what

frequency?” will depend heavily on the application and will be addressed by the Data Management Committee (DMC) as the plan evolves.

- (b) Archived data will be remotely accessible for download, minimally via a simple means for transferring files such as ftp. (If resources permit, a more sophisticated interface such as OPeNDAP may be employed). Attention will have to be paid to the size of the files to be downloaded. It may be necessary to make arrangements that limit the times that certain users have access to the data. Questions such as these will be addressed by the DMC.
- (c) Archived datasets, both short and long-term, will be distributed among multiple, geographically- distinct centres having the capability to do so. One of our government partners (DFO) has already indicated the possibility of providing an in kind contribution to assist with long term archival plans and has suggested working closely with the DMC to make this a reality. The other partners, (e.g., EC) may also have a role to play but details have to be worked out with the DMC in consultation with the Scientific Steering Committee (SSC) and individual investigators.
- (d) Selected archived data will be regarded as permanent, i.e. will remain accessible for the lifetime of the Network and, subject to storage availability, at least three years following its conclusion, except in cases of (i) operational products having ephemeral value, or for which existence of a complete archive would not be scientifically useful; such data will be archived for a minimum of 60 days; (ii) data which is discovered to be flawed and is withdrawn by the investigators. Archived data will remain accessible for the lifetime of the Network; effort will be made to find a more permanent archive for data deemed to be of particular or long-lasting value, but it will depend on the contributions of the partners (see (c) above) and, in some cases, the efforts of individuals with support of their home institutions.

## **4.2 Data Access (Issue 4)**

- (a) All GOAPP co-investigators and collaborators approved by the SSC will have access to all archived datasets.
- (b) Individual investigators are free to distribute their own data outside of GOAPP to whomever they wish, (e.g., output from the Task Force on Seasonal Prediction Experiment). Of course one investigator cannot provide access to another investigator's data without seeking approval. The SSC should only be involved if there is a problem to resolve.
- (c) Requests for general access to GOAPP’s archived datasets by non-participants will be considered on an individual basis by the responsible GOAPP co-investigators. In cases where the responsible investigator is uncertain as to the

- appropriateness of such a request, the matter may be referred to the SSC for resolution.
- (d) All archived GOAPP datasets will be made freely available (e.g., unrestricted access to non-participants and the general public) no later than 12 months following the conclusion of funded network activity.

#### **4.3 Data Dissemination to Partners, Public and Private Sector (Issues 5 and 8)**

- (a) Selected GOAPP datasets may be exported to GeoConnections / GeoPortal as contour maps, with mandatory metadata attached to it, for users from the public sector and non-participants. This would allow users to get acquainted with GOAPP and its products.
- (b) Selected GOAPP datasets will be formatted according to commonly used GIS data formats and sent regularly via FTP (or posted on OPeNDAP, if resources permit OPeNDAP to be employed) to partners in the public and private sector who require so.
- (c) COMDA is in the process of drafting its data management policy and GOAPP and COMDA will coordinate their efforts where possible. The GOAPP data management plan will also take into account the National Science Data Management policy

#### **4.4 Facilitating Collaboration within GOAPP (Issues 6 and 7)**

- (a) To facilitate collaboration, limited preferred data and metadata formats, such as CF-compliant NetCDF and GRIB format for CMC-based gridded data, will be adopted for archived GOAPP data. Every reasonable effort should be made to comply with this standard, although exceptions may occur in cases where institutional considerations preclude its use.
- (b) GOAPP Centres serving data in other formats will be asked to provide tools enabling users to translate to the GOAPP standard.
- (c) Although data storage will be distributed, the Network should endeavour to provide users with a centralized resource that describes, lists and enables access to all available data.

#### **4.5 Acknowledgements**

GOAPP encourages the use of data sets produced by the project. We would appreciate acknowledgement as follows: “The [\*\*\*] data in this paper were provided by the Global Ocean-Atmosphere Prediction and Predictability Network (GOAPP), funded by the Canadian Foundation for Climate and Atmospheric Sciences.” where [\*\*\*] refers to the actual GOAPP data used.

## 5 Implementation

- The development of an effective data management strategy, and the implementation of the policies described in this document, will be the responsibility of a newly-formed Data Management Committee (DMC).
- The DMC will report to SSC which in turn will report on data management matters to the Board of Directors.
- The DMC will have 3 members representing GOAPP's main government partners (EC, DFO and DND), and two GOAPP investigators representing Theme I and II. If necessary, an additional investigator will be added to represent the SSC.
- The DMC will focus on three specific activities:
  - Facilitating the exchange of data between projects and themes.
  - Facilitating the transfer of data and code between GOAPP and its government partners (taking into account intergovernmental initiatives like CONCEPTS, and also working with parallel international efforts with well developed data management plans, e.g., Task force on Seasonal Predictions and from THORPEX).
  - Overseeing the development of a system for making results available to the general public and interested researchers through the GOAPP web page at [www.goapp.ca](http://www.goapp.ca).
- The DMC will have to consider several general issues in order to realize, and facilitate, data exchange and archival: (i) the choice of common data formats (ii) provision of standardized documentation for the data formats and codes (where applicable). (iii) the costs of data management. The latter can only be tackled with the help of our partners, some of whom have already indicated the possibility of in-kind contributions (see 4.1 (c) above)

## 6 Appendix: GOAPP Definitions

### 6.1 GOAPP Dataset

- Model output
- Forecast products and reanalyzes
- Observation-based products (e.g., time-varying 3D fields of temperature and salinity by combining ARGO profiles with hydrographic climatologies).

### 6.2 GOAPP Co-Investigator

- has been funded through the CFCAS grant to GOAPP, and
- has been recognized as such by the GOAPP Scientific Steering Committee.

### **6.3 GOAPP Collaborator**

- is working collaboratively with a GOAPP Co-Investigator, and
- has been recognized as such by the GOAPP Scientific Steering Committee.

### **6.4 GOAPP Partner**

- contributes scientifically or operationally to GOAPP, but is not a GOAPP Co-Investigator or Collaborator, and
- has been recognized as such by the GOAPP Scientific Steering Committee.

### **6.5 Non-Participant**

Non-participating researchers include any researchers who are not GOAPP Co-Investigators, Collaborators, or GOAPP Partners. Non-Participants include the general public.