Global Argo Data Repository Status Report of US NODC for 2008

Prepared by Charles Sun
with contribution from Tim Boyer

October 2008

1. Summary

The US National Oceanographic Data Center (NODC) intended to use this report as input for the ninth Argo Data Management Team annual meeting at the East-West Center in Honolulu, USA from 28 to 31 October 2008. The report summarized the Argo user statistics and the highlights of the Global Argo Data Repository (GADR) activities since the eighth Argo Data Management Meeting at the Marine and Atmospheric Research of the Commonwealth Scientific and Industrial Research Organisation (CSIRO) of Australia in Hobart, Australia from 14 to 16 November 2007.

2. GADR Functions and Operations

The NODC operates the Global Argo Data Repository[1] (GADR), known as the Argo long-term archive, for managing and archiving the Argo data and information. The GADR performs six functions as defined at the 4th ADMT meeting in Monterey, CA:

- Archive profiles, metadata, trajectory and technical information received from the GDAC on a monthly basis.
- Provide tools to allow transformation of Argo netCDF data into other forms.
- Provide usage statistics, data system monitoring information and problem reporting facility.
- Provide data integration tools to allow client to get Argo float data combined with data collected with other instruments.
- Provide hardcopy data sets for distribution to users.
- Provide offsite storage of data.

3. Usage Statistics

This analysis was produced by analog 5.24[2]. We use the following basic definitions:

a) The number of distinct hosts is the number of different computers requests has come from. The host is the computer (often called the "client"), which has asked for a file.
b) The file might be a page (i.e., an HTML document) or it might be something else, such as an image. By default filenames ending in (case insensitive) .html, .htm, or / count as pages.
c) The number of requests is the total number of files downloaded, including graphics. The total requests counts all the files which have been requested, including pages, graphics,

[1] http://www.nodc.noaa.gov/argo/
[2] http://www.analog.cx

etc. (Some people call this the number of hits). The requests for pages only count pages. One user can generate many requests by requesting lots of different files, or the same file many times.

Figure 1 shows the number of monthly distinct hosts served by the GADR from 1 October 2007 to 30 September 2008. The monthly average of distinct hosts served by the GADR is 2,151 during this time period.

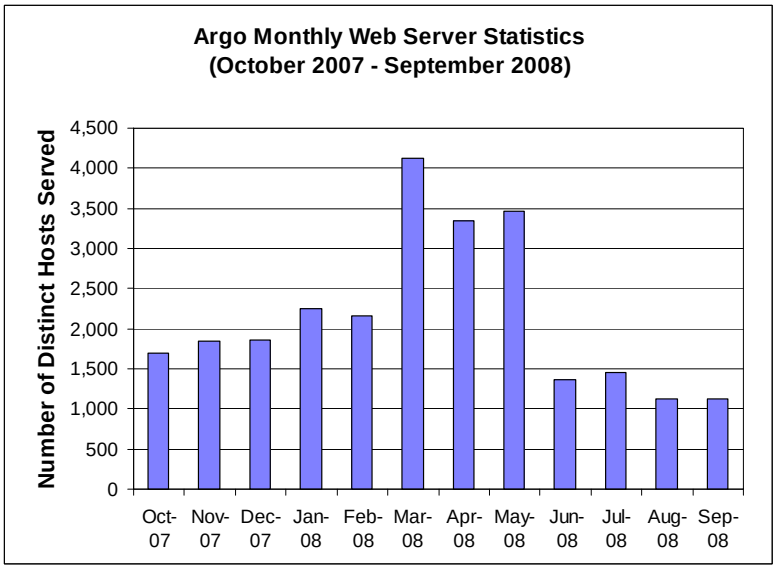

Figure 1 Monthly distinct hosts served by the Global Argo Data Repository.

Figure 2 illustrates the monthly comparison of the numbers of the Argo data files downloaded from the GADR Web site over the past 24 months ending September 2008. The GADR receives an average of 1,010,865 requests per month in the period from October 2007 to September 2008, increased from 455,909 requests per month last year over the same period of time, while the monthly-averaged Argo data downloaded increased from 17.85GB in 2007 to 39.17GB, about 119% increase, this year.

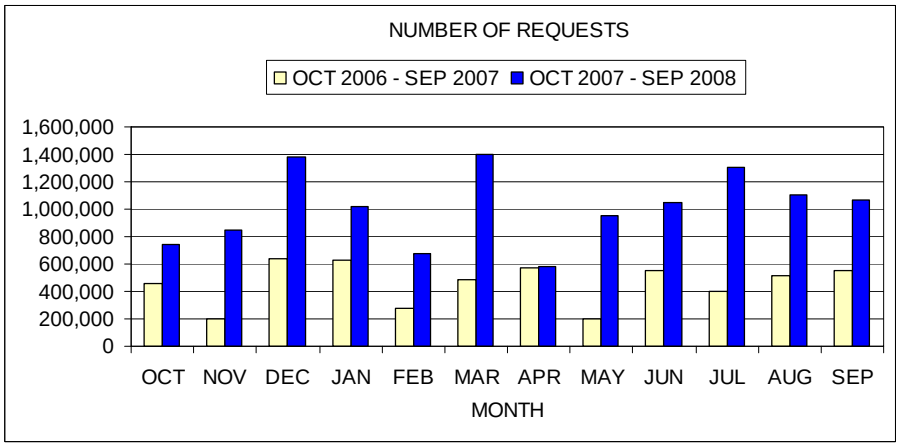

Figure 2 Monthly data transfer statistics of the Argo Data

4. Activities in support of Argo

The main support to Argo from the NODC World Ocean Database (WOD) is in relation to the Argo CTD Reference Database used for delayed-mode quality control of Argo salinity data. The quality control of Argo salinity data requires high quality CTD measurements delivered in a timely manner. The WOD provides fully quality controlled data sets approximately every 4 years. To increase the timely dissemination of more recent data to the Argo community, the WOD[3] is now updated on a quarterly schedule. The quality control is not as complete as for the full updates every 4 years and is considered preliminary. The WOD has now been updated 3 times in 2008, most recently in late September. To date, 284,244 CTD or bottle casts containing salinity data have been added in the 3 updates since WOD05. Of these, 26,767 came from the Global Temperature-Salinity Profile Program[4] (GTSPP). 33,204 of these additional casts are from cruises completed between January, 2006 and August, 2008, with 19,806 coming from GTSPP. Figure 3 shows cruise tracks of CTD/bottle casts with salinity data in the World Ocean Database taken in years 2006-2008.

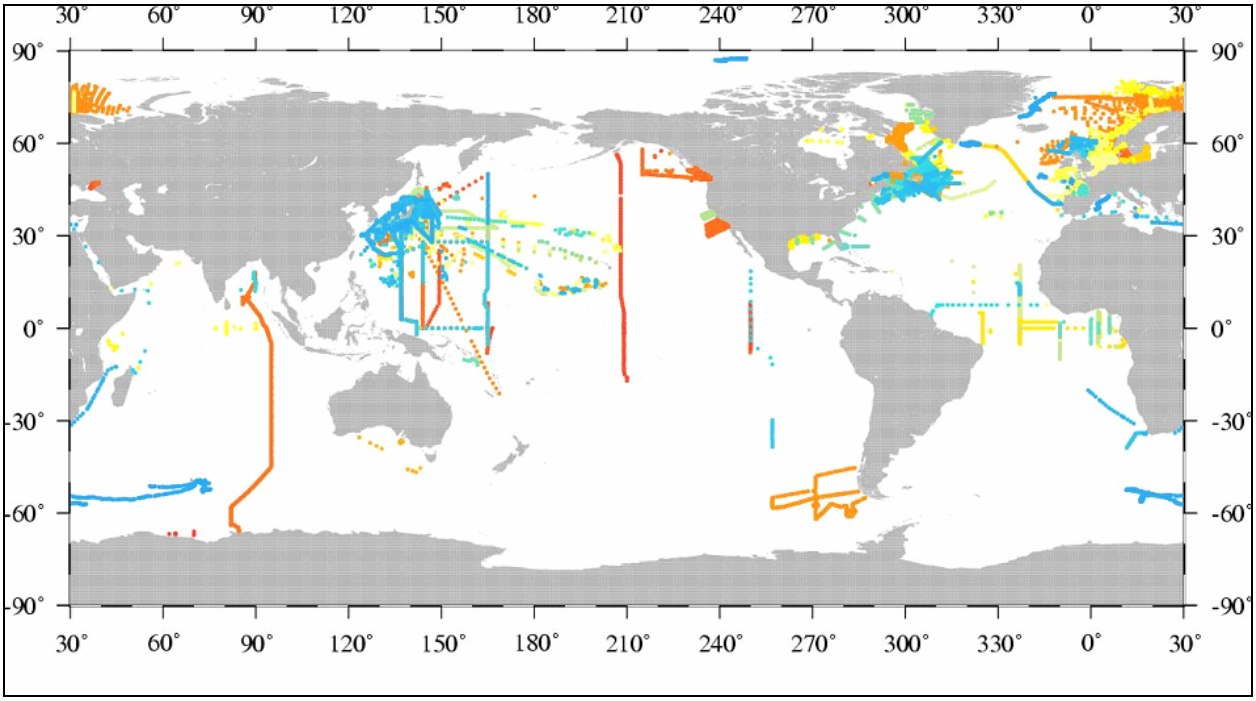

Figure 3 Cruise tracks of CTD/bottle casts with salinity data in the World Ocean Database taken in years 2006-2008

Figure 4 shows the cruise tracks of real-time CTD casts with salinity data in World Ocean Database from GTSPP taken in years 2006-2008.

[3] http://www.nodc.noaa.gov/OC5/WOD05/updates05.html
[4] http://www.nodc.noaa.gov/GTSPP/

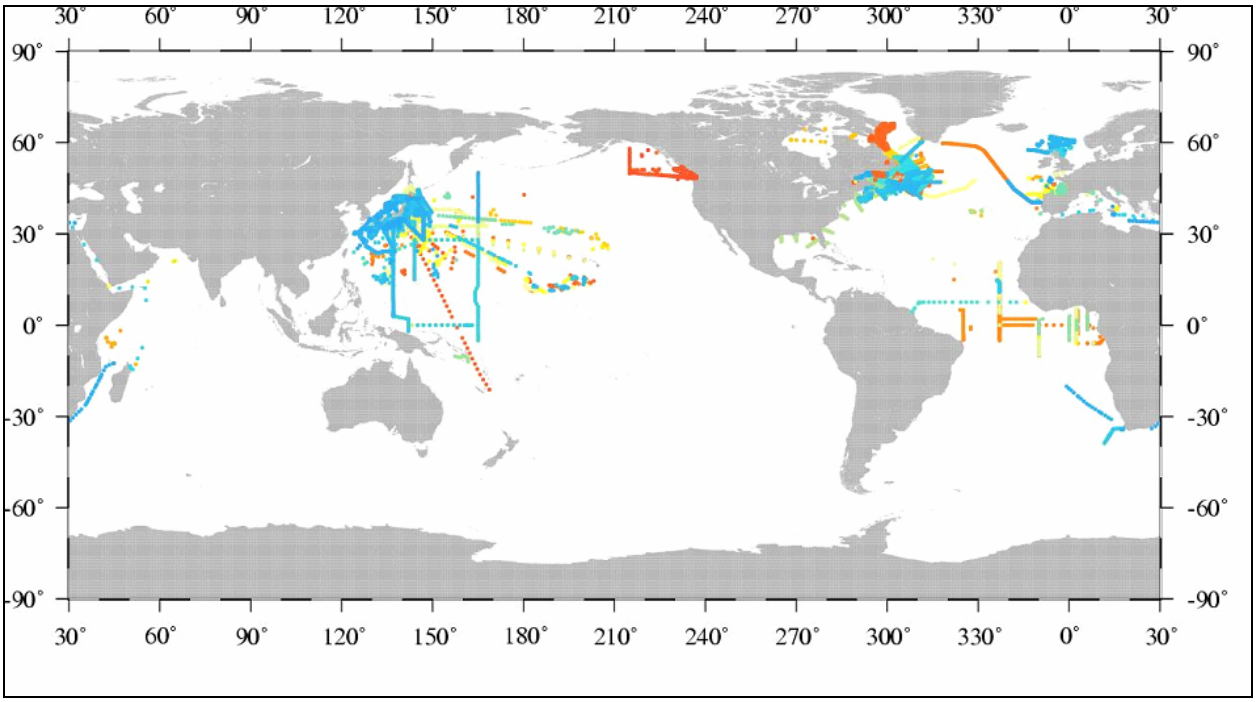

Figure 4 Cruise tracks of real-time CTD casts with salinity data in World Ocean Database from GTSPP taken in years 2006-2008.

NODC has made significant efforts to acquire recent and historical data from members of the Argo community and the larger oceanographic community. Within the Argo community, NOAA Pacific Marine Environmental Laboratory (PMEL), the Commonwealth Scientific and Industrial Research Organisation (CSIRO), and the Indian Argo Regional Center have made contributions of recent data to WOD directly. Many others have contributed directly and indirectly through National and International Data Centers, and projects such as IOC Global Oceanographic Data Archeology and Rescue Project (GODAR). There has been some disappointment voiced over a lack of recent salinity data for delayed-mode quality control. Part of this is due to the time between completion of a cruise and final calibration of the salinity data. The Argo community should decide whether it is more important to get salinity data as soon as possible after a cruise, or rather to wait for final calibrations. GTSPP real-time CTD data is a significant source for recent data. The standards of calibration and operation of CTDs of contributors to this data pool are not usually known in real-time. Also, the data are not usually of high vertical resolution. The Argo community and GTSPP managers should decide whether it is worth the effort to attempt to obtain calibration and operational information from GTSPP contributors to answer questions about data quality.

The WOD is also updated quarterly from the Argo database (US GODAE server version). Real-time data are replaced with delayed-mode data when available. Argo has high quality control standards, but some problems do escape their quality control procedures. NODC prepared a list of problems found during our quarterly updates and sent it to the Argo GDAC at Coriolis, Brest, France[5] for corrections.

[5] http://www.coriolis.eu.org//cdc/argo_problem_reporting.htm

Two functional problems which still slow down our processing of Argo data:

I. When floats are renumbered by Argo processing, or when floats exceed 255 cycles, some float data can disappear at the US GODAE server. For instance, for some Indian floats, physical cycle 256 is given cycle number 1, and the data for the original cycle number 1 is overwritten.
II. When going from real-time to delayed mode, WMO instrument type codes for some floats were altered, and they no longer match the text instrument type information. For instance float 6900271 has the following: INST_REFERENCE = "NEMO Profiling Float" ; and WMO_INST_TYPE = "846 " ; Code 846 is “Web Research/Seabird Conductivity Sensor”

For automatic processing, NODC suggests that it is easier to use the WMO code, so it is important to make sure the code matches the instrument.

5. Other Activities

5.1. Continue to preserve the Argo data transferred from the Argo US GDAC – The NODC continues to use the improved "mirror" facility of the UNIX "lftp" command. The GDAC’s files are copied from "http://www.usgodae.org/ftp/outgoing/argo/", the "geo" subdirectory is skipped, and files which are no longer present on that site are removed from the local mirror. This command runs automatically daily at 12am UTC. The mirroring process completes by about 8:30 am UTC.
5.2. Monitor Argo floats reporting pressure as depth on the GTS once a month and report problem floats found to AIC, if needed.
5.3. Implement an automated procedure for acquiring the CLIVAR & Carbon Hydrographic Data Office (CCHDO) data from the Web for archive accession.
5.4. Produced a compressed archive of the Argo data archived at the NODC as of 1 October 2008 at http://argo.nodc.noaa.gov/ .

6. Future Plans
6.1. Continue to operate the Global Argo Data Repository.
6.2. Continue to acquire the CCHDO data via the Internet on a quarterly basis.
6.3. Continue to update the WOD in support of Argo on a quarterly basis.