UK Argo National Data Management Report 2019

Report to the Argo Data Management Team – ADMT-20

Author list

The UK Argo data team (British Oceanographic Data Centre, National Oceanography Centre)

Contributing authors: Matt Donnelly, Clare Bellingham, Violetta Paba, Kamila Walicka

Other team members: Katie Gowers, Ed Small, Paul McGarrigle, Justin Buck, Sarah Chapman, Roseanna Wright

With contributions from the wider UK Argo team by:

Jon Turton (UK Met Office)

Brian King (National Oceanography Centre)

Giorgio Dall'Olmo (Plymouth Marine Laboratory)

General Status

Data management team

The British Oceanographic Data Centre (BODC) is the data assembly centre for UK Argo funded primarily by the UK Natural Environment Research Council (NERC) and responsible for data management of UK, Irish and Mauritian floats. In addition, UK Argo is a member of Euro-Argo and is managing some European Union floats as part of the MOCCA project. BODC is also the lead for the Southern Ocean Argo Regional Centre (SOARC).

The composition of the team at BODC has changed further in the past year, with an overall increase in staffing levels. Liz Bradshaw has stepped away from providing temporary support for DAC operations, whilst Sarah Chapman who undertook some SAORC related work has left BODC. Kamila Walicka and Ed Small have joined the team with a focus on DMQC. Additionally we have temporary support from Matt Cazaly for a short project on BGC derivation equations. The team currently consists of:

BODC Argo Team Member	**Role**	**Estimated contribution this FY as Full time Equivalent (FTE)**
Matt Donnelly	BODC Argo Lead DAC contributor DMQC operator SOARC coordinating partner	1.00 FTE
Clare Bellingham	DAC Lead DMQC operator	1.00 FTE
Ed Small	DMQC tools developer	0.80 FTE
Kamila Walicka	DMQC Lead	0.70 FTE
Violetta Paba	BGC Lead Lead for Argo Vocabs DAC contributor DMQC operator	0.70 FTE
Katie Gowers	Senior Argo Developer DAC contributor	0.60 FTE
Paul McGarrigle	Systems support	0.10 FTE
Justin Buck	ENVRI-FAIR project manager	0.05 FTE
Roseanna Wright	DAC and SOARC support	0.05 FTE
	TOTAL	**5.00 FTE**

Funding outlook

National Capability funding from NERC is currently maintained for BODC at the same rate as previous years. In addition, NERC-funded research projects deploying Argo floats continue provide additional sources of data management funding, such as from the ORCHESTRA, ACSIS, BoBBLE and RoSES projects. Funding for system upgrades has also been secured as part of capital purchases of a type of float in the UK inventory.

BODC funding from the EU H2020 project AtlantOS ceased in early 2019 with the end of that project. BODC continues to receive funding from the Euro-Argo ERIC MOCCA project for the European Union floats that are managed by BODC along with related DMQC activities. The Euro-Argo Research Infrastructure Sustainability and Enhancement (Euro-Argo RISE) project provides funding for developing core and deep DMQC, management of BGC extensions and regional data quality assessments in the Southern Ocean. Additionally, BODC is funded under the EU H2020 project ENVRI-FAIR to introduce the NVS vocabulary server to support Argo vocabulary management.

BODC continues to seek additional sources of funding to support SOARC functions, but a long-term solution for sustained funding is yet to be identified.

DAC Functions

Data acquired from floats

BODC retrieves data for all UK, Irish, Mauritius and assigned EU MOCCA floats from a number of sources and archives these for further processing. Processing of incoming data is normally setup within one week of float deployment. Please refer to table 1 for the types of communications used for different floats.

Data issued to GTS

BODC delivers core data in netCDF format to the UK Met Office four times a day, where it is subsequently issued to the GTS in BUFR format. Over 95% of the netCDF files are delivered within 24 hours of the data being available to BODC. Coriolis is providing the processing for 12 PROVOR BGC floats and delivering the core data to the GTS on BODC's behalf. Delivery times to the GTS can be seen in figure 1.

Progress in the past year – general processing:

BODC is currently distributing data to the GTS for c. 256 floats at the time of writing, which is an increase from 221 in November 2018. During 2019, BODC has sustained automated data processing four times a day. Delivery of core data for floats previously unprocessed has been a major focus during the past year, for all APF9I/N1/N2 and APF11 Argos floats. Work on APF11 iridium floats remains a priority for completion.

Current activity and future plans:

Distribution of all core data to the GTS from all BODC managed floats is a priority, including core data from floats with any type of Argo extension (deep, BGC or auxiliary data). BODC's current focus is to ensure all floats with APF11 iridium controller boards are effectively managed within the BODC Argo System. Delivering on this objective is the highest priority for UK Argo. BODC is seeking to collaborate with other DACs in the development of BGC parameters especially regarding QC techniques.

Data issued to GDACs after real-time QC

All core data received for currently processed floats are distributed to the GDACs within one hour of the data arriving at BODC, with the real-time quality control tests applied. Any file that fails to be transferred is queued for the next transfer attempt. BODC has completed the conversion to v3.1 for meta and core data with the exception of a small number of legacy fil. Please refer to table 1 for the types of float and whether they are being fully processed.

Progress in the past year:

BODC Argo has developed capability for the delivery of data from floats that house the APF9 Iridium and N1/N2 controller boards, enabling data delivery of core data for 23 additional floats. We have also developed our system to deliver APF11 Argos floats, enabling data delivery of core data for another 6 floats. Work has also progressed on managing BGC profile data providing the infrastructure for its delivery after all core data is available. Work is underway to deliver data from all of our APF11 Iridium floats.

Current activity and future plans:

There remains substantial further work to complete the delivery of the remaining core profile, tech and trajectory files, in that order of priority. We are not currently issuing any BGC-Argo files for UK floats due to the current focus on core profile data. The exception to this is the dozen PROVOR floats kindly hosted for BODC by Coriolis until such time as BODC can take over the real-time processing.

Data issued for delayed-mode QC

Delayed-mode QC on BODC hosted floats is performed within BODC and as well as through external operators on some floats in the MOCCA project. Currently BODC is only capable of providing data for delayed-mode QC for core data. Investigative work has progressed to enable the delivery of biogeochemical parameters in v3.1. The exception to this is 13 PROVOR floats that Coriolis is hosting on BODC's behalf. See section 2 of this report for the status of delayed-mode QC.

Progress in the past year:

Progress on v3.1 profile files has made more delayed-mode files available in v3.1. BODC Argo has made significant progress this year on the number of profiles with delayed mode QC of core data on the GDAC. BODC Argo is now also able to visually QC core data from floats which use the Coriolis processing stream i.e. MOCCA floats as they become eligible for delayed-mode QC using the SCOOP software (but some issues remain with this software).

Current activity and future plans:

Progress on DMQC will continue in the coming year, and BODC expects to reach > 90% of core profiles in D-mode by ADMT-21

Delayed-mode data sent to GDACs

All delayed-mode QC on BODC hosted floats is submitted to the GDACs the same day that delayed mode QC is complete for a profile when completed by BODC, or as soon as the data has been accepted following submission by external DMQC partners. Submissions from external partners are issued with accession numbers for tracking purposes within BODC archives. See section 2 of this report for the status of delayed-mode QC.

Table 1: Summary of all BODC managed Argo floats, with a focus on those that are currently active

					Mission of active floats					
Float type/controller	**Comms**	**Total no. of deployed floats**	**Total no. of active floats**	**No. of active floats with ice detection**	**Core only**	**Core + NST**	**Core with RBR CTD**	**Core + oxygen only**	**Core + other BGC**	**Total no. of active floats being fully processed**
Martec Provor	Argos	26	-	-	-	-	-	-	-	-
MetOcean NOVA	Iridium	1	-	-	-	-	-	-	-	-
NKE Arvor	Argos	5	-	-	-	-	-	-	-	-
NKE Provor *	Iridium SBD	13	8	-	-	-	-	-	8	8
NKE Arvor	Iridium SBD	77	72	-	-	72	-	-	-	72
NKE Deep Arvor	Iridium SBD	2	-	-	-	-	-	-	-	-
SBE Navis N1	Iridium Rudics	12	5	5	5	-	-	-	-	5
SBE Navis N1 with BGC	Iridium Rudics	4	-	-	-	-	-	-	-	-
SBE Navis N1 with oxygen	Iridium Rudics	8	8	-	-	-	-	8	-	8
SBE Navis N1 with radiometer	Iridium Rudics	3	3	-	-	-	-	-	3	3
TWR Apex APF7	Argos	8	-	-	-	-	-	-	-	-
TWR Apex APF8	Argos	252	-	-	-	-	-	-	-	-
TWR Apex APF9A (7 types)	Argos	269	124	6	24	100	-	-	-	124
TWR Apex APF9I	Iridium Rudics	20	2	2	2	-	-	-	-	2
TWR Apex APF9I with BGC	Iridium Rudics	4	-	-	-	-	-	-	-	-
TWR Apex APF9I with STS	Iridium Rudics	4	3	-	-	3	-	-	-	3
TWR Apex APF11	Argos	6	6	-	6	-	-	-	-	6
TWR Apex APF11	Iridium Rudics	14	10	n/k	-	-	2	2	6	0
TWR Deep Apex APF11	Iridium Rudics	14	3	n/k	2	-	-	1	-	0
TOTAL		742	244	13	39	175	2	11	17	231

* = processing courtesy of Coriolis

Web pages

BODC continues to maintain the UK Argo website (www.ukargo.net) along with a Facebook page (www.facebook.com/UKArgofloats/) and a Twitter account (twitter.com/ukargo). BODC/NOC also maintains the SOARC website (www.soarc.aq).

Data use and data products

Statistics of Argo Data Usage

National Oceanography Centre

Argo data are used widely within NOC science with the following regional leads for float deployment and science:

- Alex Sanchez Franks (Indian Ocean)
- Yvonne Firing (Southern Ocean)
- Penny Holiday (Sub-polar N Atlantic)
- Brian King (everywhere else)

The applications of Argo data at NOC include:

- Measurement of evolution and drivers of mixed layer processes in the (Indian Ocean);
- Inventory, transports and evolution of heat and freshwater establishing controls on budgets (both regional and global);
- Deep heat content (N Atlantic).

UK Met Office

All Argo data together with other ocean data, received over the WMO GTS, are routinely assimilated into the Met Office's FOAM (Forecasting Ocean Assimilation Model) which is run daily. The FOAM suite runs daily in an early morning slot and produces 2 analysis days and a 7-day forecast. The 3-D temperature, salinity and current fields from the global model run are used as boundary conditions for the regional models. There are 4 different configurations: ¼ degree global, 1/12 degree North Atlantic, 1/12 degree Mediterranean, 1/12 degree Indian Ocean and ~6km European North West Shelf. More details are at: http://www.ocean-sci.net/12/217/2016/os-12-217-2016.pdf and http://www.geosci-model-dev.net/7/2613/2014/gmd-7-2613-2014.html . The global FOAM system is used to initialise the ocean component of coupled monthly-to-seasonal forecasts, and so the requirements for Argo for that application are the same as for FOAM.

A coupled ocean/atmosphere prediction system has been developed for weather forecasting timescales, including assimilating Argo data in a coupled data assimilation framework (Lea et al., 2015), and is now being run operationally, delivering ocean forecast information to the Copernicus Marine Environment Monitoring Service (CMEMS). The timeliness constraints on Argo for this application are more stringent (data need to be available within 24 hours of measurement, and preferably within 6 hours). The impact of Argo on this system was assessed as part of the E-AIMS EU project (King et al., 2015). It is likely that future versions of coupled data assimilation schemes will require Argo data with timeliness of 3 hours (Chris Harris, Met Office Coupled Data Assimilation Manager, pers. comm., October 2018).

Near-surface Argo data are used to validate the output from the Met Office's OSTIA (Operational Sea Surface Temperature and Sea Ice Analysis) – the OSTIA fields are in turn used as a lower boundary condition in numerical weather prediction models run by both the Met Office and ECMWF.
Argo data are also used in the initialization of ocean conditions in models run to make decadal predictions, see: http://www.metoffice.gov.uk/research/modelling-systems/unified-model/climate-models .

Plymouth Marine Laboratory

Giorgio Dall'Olmo is the lead PI for BGC data in the UK. Bio-Argo data from 13 Provor floats are now available from the GDACs, thanks to processing courtesy of Coriolis.
Core-Argo data are used at PML for:

- providing a description of the physical environment in the framework of biological (e.g. mapping eel migration routes) and biogeochemical studies;
- developing techniques to generate 3D fields of biogeochemical variables by merging ocean-colour and in-situ data;
- investigating mesoscale structures by combining altimetry and in-situ profiles with a special focus on Agulhas rings.

BGC-Argo data focuses on investigating new methods to:

- efficiently monitor the ocean biological carbon pump;
- quantify particle flux attenuation;
- vertically-resolve seasonal remineralisation rates;
- to better understand the nitrogen cycle in oxygen minimum zones.

Data Products

UK Met Office

The Hadley Centre maintains two data products that incorporate Argo observations:

- EN4 contain in-situ ocean temperature and salinity profiles and objective analyses. It is updated monthly using real-time Argo profiles, and annually using delayed-mode Argo profiles. EN4 is freely available for scientific research use (see ***http://www.metoffice.gov.uk/hadobs/en4/***);
- HadGOA is an integrated database of surface and sub-surface temperature and salinity observations for the period 1850 to present. It includes quality flags, bias corrections and uncertainty information (Atkinson et al., 2014). At present, HadGOA obtains sub-surface profile data from EN4. The data is publicly available at: https://www.metoffice.gov.uk/hadobs/hadgoa/data/download.html.
- The datasets are used for climate and global change studies, including ocean heat content analysis.

Delayed-Mode QC

BODC has improved delayed mode QC capability through a number of activities:

2.1 Workflow improvement

Shortly after the previous ADMT-19 meeting, BODC installed and implemented the new DMQC software OWC 2.0.1, released in December 2018 by Coriolis. The update of software was accompanied by an installation of the newly updated CTD and Argo reference dataset (CTD_for_DMQC_2018v1 and Argo_for_DMQC_2018v2). This activity required an update of in-house DMQC procedural documents.

In the early January BODC employed a new staff member, Kamila Walicka who received in-house training, mostly from Matt Donnelly with the support of Justin Bulk. Kamila has obtained extensive training about the structure and procedures of OWC software and has learnt about the DMQC methodology developed since the beginning of the Argo observing network. From February, Kamila started to submit a vast major of BODC's floats to GDAC. In March, Kamila visited our external European partners (Ifremer and BSH) in order to obtain more experience in using OWC software and benefit from wider knowledge and experience.

Another workflow improvement was the creation of a new DMQC report generator by Clare and Kamila, using the open-source LaTex (https://www.latex-project.org/) typesetting system. The newest improved DMQC report generator includes the detailed description of the OWC configurations for the specific regions, sequence of the checking procedures of data quality and scientific justification of the decisions made to provide the a high quality report for long-term reference.

2.2 Current activities

2.2.1 DMQC core progress

From December 2018 until the time of writing, BODC has submitted 101 floats, including 15 213 profiles:

- 31 EU MOCCA project floats, including 2028 profiles
- 13 UK MetBio Provor floats, including 3868 profiles
- 57 UK Argo floats, including 9317 profiles

By the end of September 2019, 55 % of BODC hosted floats profiles eligible for delayed mode QC have been processed and submitted to the GDACs. BODC expects to continue improving the number of submitted floats in the coming year, involving Kamila, Matt and Clare.

2.2.2 Argo DMQC Community Tools Survey

As a part of the Euro-Argo RISE project, Task 2.4: "Development & Implementation of DMQC methods" BODC is committed to undertaking a review of DMQC tools used within the Argo community. The aim of this report is to cover matters such as the report assessments of tool capabilities, interoperability between organisations, state of development, minimum system requirements, availability of access to the tool and number of current users. The final output from the report will be one or more recommended pathways to improve the sustainability of DMQC tools development within the global Argo data system. The aspiration is to improve collaboration to allow the DMQC process to be more sustainable, transparent, and easier to implement for groups managing a modest fleet of Argo floats or simply freshly involved in the complex task of DMQC.

This work was originally started under regeneration of our DMQC procedures for MOCCA in late 2017 but has been put on hold due to other demands. This task was undertaken by Kamila Walicka under Euro-Argo RISE WP2that will be completed before the Euro-Argo RISE WP2 meeting in November 2019. The main effort is to

produce a report on DMQC tools detailing their capability and operating requirements. Initial findings of the survey will be shared at ADMT20 for consideration by the community.

2.2.3 Argo OWC Conversion from Matlab to Python

The currently available OWC software has been developed in Matlab over the course of the past 10+ years. The computational efficiency of this process is not clear, but it is a time-consuming process. However, we know that various Argo partners do not have Matlab licenses, or have old versions of Matlab, or do not have access to Matlab toolboxes, such as the Optimisation Toolbox used by OWC. The licensing issue restricts the pool of potential DMQC operators, exacerbating a human resource shortage. Transferring the Maltab version of OWC to open-source software has the potential to enable more groups to use the most up-to-date version of OWC, and to enable DMQC operators from newly contributing institutions.

In order to address the issues mentioned above, BODC aims to deliver a translatedversion of OWC in the Python language.,To address this task, in September, BODC employed software engineer Edward Small, who is responsible for transferring the OWC software from Matlab to Python. This project is undertaken within the MOCCA project that runs up until June 2020. However, the aim is for this work to be completed well in advance of this close-down date. The main objectives of this project are to retain all existing functionality, achieve the same outputs as the Matlab version, enhance the computational efficiency, and ensure the Python version of the code is suitable for integration into a GUI interface.

2.2.4 Quality Control of BGC data

BODC has downloaded and implemented the newest version of software SAGE-O2 that is used for verification of the O2 and pH data quality and applying proposed corrections. Currently, these works are focused on understanding the SAGE software and provide the training for the UK Argo team, with an aspiration to begin applying corrections to the netCDFs in the coming months. This task is led by Violetta Paba. Additionally, BODC will be contributing to the further development of oxygen and pH QC procedures.

2.3 Future Plans

BODC is considering the implementation of Jupyter Notebooks to manage workflow and efficiency improvements under MOCCA. The Jupyter Notebooks (https://jupyter.org/) is a tool that could potentially be used to streamline the DMQC process, integrating Python OWC version with LaTeX report generation to move to an increasingly seamless approach.

GDAC Functions

NERC Vocaulary Server

As part of a wider environmental sciences infrastructure application, BODC has secured funding from the EU's H2020 funding programme to undertake significant work on adding the Argo vocabulary to the NERC Vocabulary Server (NVS). The outline for this package of work submitted for the proposal was:

> "The provenance of data in the Argo Data System is underpinned by rich metadata which is standardised across the data system using vocabularies currently held in manuals and associated spreadsheets. The accuracy, controlled evolution and semantic value of this metadata can be further enhanced by migrating these existing vocabularies to a controlled vocabulary management environment and server such as the NVS vocabulary server. The NVS manages controlled vocabularies according to internationally agreed W3C-compliant standards. Its existing infrastructure and associated tools underpin various environmental data systems in Europe, Australia and the USA. As part of the European SeaDataCloud project the NVS is being further enhanced to improve the transparency of the governance model and provide editorial access to external users. High quality management of Argo's vocabularies (including list of codes, terms and their definitions) will involve reviewing and potentially enhancing/refining existing definitions to create a set of well managed catalogues, introducing new catalogues where required, and performing detailed concept mapping within and between catalogues. Such mappings will facilitate and enhance the accessibility of the Argo netCDF repositories and interoperability with other research infrastructures through inter and intra domain mappings, as well as facilitate future efficiencies at Data Assembly Centres (DACs) by introducing new catalogues of manufacturer metadata concepts mapped to Argo data system terms. This work will prioritise vocabularies and mappings that would have the highest impact. This activity will be undertaken through close cooperation with the global Argo Data Management Team to ensure that appropriate governance is maintained for migrated vocabularies."

Since last year, in consultation with other members of ADMT, BODC has developed plans for this project and begun implementing the use of the NVS for Argo Vocabularies. This includes developing an understanding of how this work will improve the 'FAIRness' of the Argo system:

- Findable
  - Fully described on a vocabulary server
  - Mappings within and beyond the Argo data system
  - the NVS will make Argo metadata readable by machines
- Accessible
  - All metadata definitions in one place
  - Potential for e.g. GDAC checks automation
  - the NVS can be accessed through several APIs, including ReSTful, SOAP and SPARQL.
- Interoperable
  - NVS collections easily exported to machine-readable formats, e.g. JSON, CSV etc.
  - Holding the Argo vocabularies in the NVS will make Argo NetCDFs fully self-descriptive (to humans as well as machines).
- Re-usable
  - Facilitate aggregation with other data

In addition to the development of 'FAIRness', there are additional benefits for the Argo Data System. Firstly it should decrease the overall workload and decrease the risk of error by having a single version-controlled

definitive source of information which is fully described, machine and human readable. Secondly, it will facilitate connections with other data infrastructure through clear an unambiguous mappings, and in doing so will set the standard for other ocean observing networks as they develop their data systems. The NVS provides a single source of information for a wide variety of vocabularies, whilst supporting different governing authorities. Governance policies have been put in place for the different vocab editors within Argo, overseen by BODC technical governance, see:

- C30 Active vocabulary content governance authorities
- C88 BODC asset access right roles

As part of the upload of Argo reference tales to the NVS their content will be reviewed to ensure each term has an appropriate ID, name and definition. Argo reference tables on the NVS will carry the 'R' prefix, such as R03, R25, R26, R27 – named to match existing reference table names were possible. These four vocabularies have been partially uploaded as demonstrators.

BODC aims to provide training to vocab editors and other contributors regarding the NVS to support them in adopting the service. BODC will consult with the Argo community regarding further development of the services provided by the NVS.

Regional Centre Functions

BODC continues to provide the coordinating role between the SOARC partners and hosts the SOARC website (www.soarc.aq). Feedback on the website is welcome and can be submitted either via the website contact form or direct to argo@bodc.ac.uk. Matt Donnelly is the SOARC lead at BODC.

Following discussions at the BGC Workshop in Seattle, Tanya Maurer and Josh Plant now represent the SOCCOM project in SOARC. As of September 2019, Matt Donnelly represents SOARC on the Southern Ocean Observing System (SOOS) Data Management Sub-Committee (DMSC).

BODC activities progressed in the past year include:

- Completion of the 'Argo and the Antarctic Treaty' guidance document, now available on the SOARC website;
- Beginning a project on Argo profile characterisation in the Southern Ocean in collaboration with the University of Bristol;
- Restored full DMQC capability and able to support other national programmes in the Southern Ocean;
- Developed a better understanding of other work on under-ice float position research ready to pursue operationalising that knowledge;
- Various website updates and creation of a SOARC GitHub organisational repository;
- Research on potential approaches to Southern Ocean regional data quality assessments. The initial focus is on using the work of Reeve et al. (2016a), Reeve et al. (2016b) and Reeve et al. (2019) to establish routine Weddell Gyre data quality assessments.

Future work for BODC on SOARC includes:

- Developing more deployment opportunities in collaboration with Jcommops and SOOS, e.g. through other organisations such as IAATO and CCAMLR;
- Improving under-ice positioning methods;
- Improve the availability of Southern Ocean DMQC resources;
- Develop regional data quality assessments for the Antarctic Circumpolar Current (ACC) and Weddell Gyre, in collaboration with NOC (UK) and BSH (Germany), with the latter intended to be based upon Reeve et al (2016a), Reeve et al. (2016b) and Reeve et al. (2019).

A more comprehensive summary about BODC's activities, and those of other partners, will be available in the SOARC presentation and written report.

References

Reeve, Krissy A; Boebel, Olaf; Kanzow, Torsten; Strass, Volker H; Rohardt, Gerd; Fahrbach, Eberhard (2016a): Objective Mapping of Argo data in the Weddell Gyre: a gridded dataset of upper ocean water properties, link to data files in NetCDF format. *PANGAEA*, https://doi.org/10.1594/PANGAEA.842876

Reeve, KA et al. (2016b): A gridded data set of upper-ocean hydrographic properties in the Weddell Gyre obtained by objective mapping of Argo float measurements. *Earth System Science Data*, 8(1), 15-40, https://doi.org/10.5194/essd-8-15-2016

Krissy Anne Reeve, Olaf Boebel, Volker Strass, Torsten Kanzow, Rüdiger Gerdes (2019) Horizontal circulation and volume transports in the Weddell Gyre derived from Argo float data, Progress in Oceanography, Volume 175, Pages 263-283, https://doi.org/10.1016/j.pocean.2019.04.006