UK National Report

Report to the 21st Argo Data Management Team

Authors

UK Argo data team at the British Oceanographic Data Centre, part of the National Oceanography Centre:

- Contributing authors: Matt Donnelly, Clare Bellingham, Kamila Walicka
- Other team members: Katie Gowers, Violetta Paba, Ed Small, Roseanna Wright

With contributions from the wider UK Argo team by:

- Jon Turton and Fiona Carse (Met Office)
- Brian King (National Oceanography Centre)
- Giorgio Dall-Olmo (Plymouth Marine Laboratory)

General Status

Data management team

The British Oceanographic Data Centre (BODC), part of the National Oceanography Centre (NOC), is the data assembly centre for UK Argo funded primarily by the UK Natural Environment Research Council (NERC) and is responsible for data management of UK, Irish and Mauritian floats. In addition, UK Argo is a member of Euro-Argo and is continuing to manage some European Union floats as part of the now ended MOCCA project. Starting as part of the EU H2020 project ENVRI-FAIR, BODC is working towards hosting the Argo reference tables on the NERC Vocabulary Server (NVS). BODC is also the coordinator for the Southern Ocean Argo Regional Centre (SOARC).

BODC Argo Team member	Contributions	Estimated contribution in past year as Full Time Equivalent (FTE)
Matt Donnelly	BODC Argo Lead DAC lead DMQC contributor SOARC coordinating partner	1.00 FTE
Clare Bellingham	DAC operator DMQC operator	0.95 FTE
Kamila Walicka	DMQC lead	0.85 FTE
Ed Small	DMQC software developer	0.75 FTE
Violetta Paba	Argo vocabularies lead BGC QC lead DAC operator	0.35 FTE
Katie Gowers	Senior Argo software developer	0.15 FTE
Roseanna Wright	Metadata investigations	0.05 FTE
Steve/Tom/Matt C./Kay	Server migration support team	0.20 FTE
	TOTAL	4.30 FTE

The composition of the team at BODC has fluctuated during the course of the year, with Violetta Paba and Katie Gowers both currently on parental leave, Ed Small finishing his 1-year position, and Paul McGarrigle stepping back from routine systems support. The core BODC Argo team has been

supported by other members of BODC this year, namely Steve Loch, Tom Garner, Matt Cazaly and Kay Thorne to help with a move of the Argo DAC to a new set of servers. Justin Buck is the ENVRI-FAIR project manager, which includes BODC contributions to the FAIR agenda wider than just Argo. This project has also involved contributions from wider BODC including Gwen Moncoiffe and Alexandra Kokkinaki for NVS content management support, technical advice and training.

General outlook

Core BODC Argo national capability funding from NERC remains static and is decreasing in real-terms. There is also additional funding from NERC associated with particular research projects and the floats they have procured, including from ORCHESTRA, ACSIS, BoBBLE and RoSES projects, but each of these draws to a close in the next year or two. Efforts have been made this year to establish a clear plan for future funding to develop a more sustainable model of UK funding to support the UK contribution to the full-depth multi-disciplinary Argo array, but the funding situation remains challenging.

BODC funding from the EU EASME project MOCCA ended in August 2020 with the end of the project. This project has brought many benefits and efficiency improvements, although the remaining data management commitments for active floats will be covered by UK funding. The Euro-Argo Research Infrastructure Sustainability and Enhancement (Euro-Argo RISE) project provides funding for developing core and deep DMQC (Delayed Mode Quality Control), management of BGC (Bio Geo Chemical) 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.

During 2020 UK Argo contributed to the Argo Data Paper (Wong et al. 2020).

Status

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. BODC currently processing data from floats with Argos communications, Iridium Rudics and Iridium Short Burst Data (SBD) from a diverse fleet of floats manufacturer by TWR, SeaBird and NKE.

Near real-time data delivery

Processing and delivery of incoming data is normally setup within at most one week of float deployment where this capability already exists for a given type of float. At the time of writing, capability to deliver core/deep data exists for almost the entire fleet managed by BODC, with only 3 floats still awaiting a processing stream to be established. This is a dramatic reduction on last year when all of our TWR floats with APF11 controllers and Iridium communications were still outstanding. The remaining floats are a couple of inactive UK Deep Arvor floats and an active Irish Arvor oxygen float. Work is already underway to enable the processing of these floats. Coriolis has been providing the processing for 13 PROVOR BGC floats and delivering the core data to the GTS on BODC's behalf, but it is intended to bring management of these floats back into BODC in the coming year.

During the past year, the BODC Argo System has been developed to process a diverse fleet of APF11i controller board floats, including core, deep and BGC sensor equipped floats. This small fleet ranges from early APF11i floats to the most recent, across several batches and multiple firmware versions, and has posed a challenge in delivering an exception free processing pipeline. The focus for the coming year is to enable the release of all BGC profile data across TWR, SeaBird and NKE float models.

The instance of the Coriolis processing chain deployed at BODC, primarily enabling delivery of data from MOCCA project floats and some Irish floats, was upgraded in January 2020 to fix some processing issues in this software stack.

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.

In May 2020, BODC refactored its workflow for delivering RT NetCDFs to the Met Office, preventing accidental resubmissions to the GTS caused by the previous setup. This should eliminate pollution of GTS metrics by accidental resubmissions that have been noted in the past couple of years.

BODC and the Met Office are planning to review the current approach to GTS distribution in the coming year with the aim of minimising the delays between data recovery and delivery to the GTS, with available options likely to reduce the median delivery time by at least an hour, and possibly more.

Data issued to GDACS after real-time QC

BODC delivers updated meta and tech files for all floats it processes alongside new core profile files to the GDACs as part of every processing run. The transition to generating all technical files in v3.1 was completed in May 2020. Delivery of BGC profile data and many trajectory files are still pending as BODC continues its recovery from the demands of new float types, payloads and the transitional to v3 files. Delivery of BGC profile files is planned as a major focus for 2021. There have also been recent improvements to oxygen and pH sensor metadata in support of APF11i development.

During the past year BODC has returned to routinely addressing Altimetry QC and Objective Analysis reports shortly after receipt.

Data issued for delayed-mode QC and sent to GDACs

The BODC DAC function currently interacts with DMQC operators through two different modes of operation. The first is internal BODC DMQC operators who directly submit DMQC decision to the BODC Argo System, and for which updated D-mode NetCDFs are automatically generated and submitted. For floats managed through the Coriolis processing chain instance deployed at BODC, both internal and external DMQC operators submit updated NetCDFs which are archived within BODC and submitted to the GDACs.

Web pages

NOC 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). NOC also maintains the SOARC website (www.soarc.aq).

Data use and data products

Met Office

At the Met Office Argo data are used operationally:

- they are routinely assimilated into its FOAM (Forecasting Ocean Assimilation Model) suite which is run daily and produces 2 analysis days and a 7-day forecast;
- fields from global FOAM are also used to initialise the ocean component of coupled monthly-to-seasonal forecasts;
- Argo data are also used in the initialisation of ocean conditions in climate models run to make decadal predictions;
- a coupled ocean/atmosphere prediction system has been developed for weather forecasting timescales, and is now being run operationally, delivering ocean forecast information to the Copernicus Marine Environment Monitoring Service (CMEMS);
- near-surface Argo data are used to validate the output from the Met Office's OSTIA (Operational Sea Surface Temperature and Sea Ice Analysis), where the OSTIA fields are used as a lower boundary condition in numerical weather prediction models run by both the Met Office and ECMWF.

Two new Met Office systems which are in the process of being made ready for operational implementation are:

- an improved resolution version of global FOAM with 1/12 degree horizontal resolution, due for operational implementation in 2020. This will continue to make use of Argo data to constrain the T/S fields in the same way as the existing ¼ degree resolution system.
- a coupled weather forecasting system which is initialised using coupled data assimilation, due for operational implementation in 2021. Once this is implemented operationally Argo data will directly contribute to operational weather forecasts as well as ocean forecasts. An assessment of the impact of Argo in a lower atmospheric resolution version of that coupled system was detailed in King et al., 2019.

Met Office research & development applications (non-operational) which have made significant use of Argo data:

- David Ford has done some OSSEs looking at the impact of the planned BGC-Argo array of floats in a global physical-biogeochemical model where he assimilates synthetic versions of the BGC Argo profiles in conjunction with satellite ocean colour data. A paper based on that work is currently in preparation.
- one other project where we made good use of Argo data was in the assimilation of satellite sea surface salinity data from SMOS, Aquarius and SMAP. The near-surface salinity data from Argo was used to bias correct the satellite salinity data and was crucial for the performance of the assimilation of SSS data. That work is written up in Martin et al., 2019.

In the Hadley Centre for Climate Science, Argo data is used to make the following products:

- EN4 contains insitu ocean temperature and salinity profiles and objective analyses. It is updated monthly using real-time Argo profiles and GTSPP data, and annually using delayed-mode Argo profiles (and WOD, GTSPP and ASBO data). EN4 is freely available for scientific research use (see http://www.metoffice.gov.uk/hadobs/en4/). In 2019 a user requirements survey was undertaken about EN4 and an updated version

incorporating more uncertainty information and an updated analysis system is due for release this year (EN5).
- HadIOD (Hadley Centre Integrated Ocean Database) is a database of in situ surface and subsurface ocean temperature and salinity observations supplemented with additional metadata including bias corrections, uncertainties and quality flags. The dataset is global from 1850-present with monthly updates. The current version is HadIOD.1.2.0.0, the chief sources of data are ICOADS.2.5.1, EN4 and CMEMS drifting buoy data. This product has been available to the public since mid-2020 via https://www.metoffice.gov.uk/hadobs/.

Met Office science uses of the EN4 product include OHC analysis, contributions to BAMS, Ocean Obs'19 White Paper and an upcoming Earth Energy Imbalance paper (von Schuckmann et al., 2020).

National Oceanography Centre

Argo data are used widely within NOC, where the science applications include:

- measurement of evolution and drivers of mixed layer processes in the (Indian Ocean);
- inventory and evolution of heat and freshwater establishing controls on budgets (both regional and global);
- deep heat content (N Atlantic).

Plymouth Marine Laboratory

PML have the lead for BGC Argo in the UK, where the data are used for:

- investigating different aspects of the biological carbon pump (e.g., mixed-layer pump, fragmentation);
- investigating export fluxes and efficiency in hypoxic ocean regions;
- 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.

Delayed Mode QC

Core DMQC progress

The strategy adopted to deliver the support to national programmes focused on ensuring a high-quality approach and the progressive enhancement of expertise. This involved upgrading to the most updated DMQC software, adopting the latest reference databases and extensive training for DMQC operators in BODC. DMQC support was offered to any national programme requiring assistance, where BODC has significant physical oceanography expertise. BODC focused on supporting the UK and Irish Argo programmes, data for which are managed by the BODC Argo Data Assembly Centre (DAC) function. Recently BODC has provided support to Polish Argo programme with support in reviewing, providing advice in analysing Argo floats in delayed-mode and DMQC operator training.

The DMQC analysis has been undertaken on floats deployed in five different regions: the North Atlantic, the South Atlantic, the Southern Ocean, the Indian Ocean and a small number of floats in the Pacific Ocean. In some of the regions, such as the Pacific Ocean, this was the first time BODC had undertaken DMQC in the region. The analysis has been undertaken it separate batches for each of these regions, with floats that had already ceased functioning and those with a large number of

profiles being prioritised. This approach provided the opportunity to gain a deeper understanding of the regional oceanography in each region, the variability in float behaviour over time, and the challenges this raises in DMQC.

This work has resulted in a significant improvement in the total amount of delayed-mode profiles delivered by BODC compared with available real-time mode data. From September 2019 until the time of writing, BODC had submitted around 22 527 profiles. The most recent statistics provided by Ifremer, from October 2020, shows that BODC had delivered around 75 % of delayed-mode data from all available data at the DAC. This is a significant improvement compared to the low-point in December 2018 where only around 45 % of profiles had been through DMQC, and this situation is expected to improve further in the coming year.

A combination of the strong focus on comprehensive training provided to BODC Argo team members and the diverse experience of DMQC ensures future sustainability in providing DMQC analysis and regular delivery of Argo data in delayed mode to the Argo Global Data Assembly Centres (GDACs).

DMQC workshop participation and coordination

NOC has undertaken the preparation of hosting the 2nd European Argo/7th International Argo Delayed-mode Workshop for CTD data in Liverpool, UK. This workshop aimed to include the DMQC analysis for both core (2000 m) and deep (4000 m-6000 m) Argo floats, with the latter being a focus of the EuroArgo RISE project WP3 on developing deep DMQC methods. The agenda and registration were advertised with the support of the Euro-Argo ERIC Office via *https://www.euro-argo.eu/News-Meetings/Meetings/Others/2020-DMQC-workshop*. The meeting was planned to happen from 12th May to 15th May 2020. However, due to the ongoing pandemic of COVID-19 virus, this workshop has been postponed. The next DMQC workshop for deep Argo floats is planned to be organised virtually in June/July 2021.

Contributions to DMQC cookbook for Core Argo parameters

BODC (Kamila Walicka) has contributed to the development of an Argo DMQC cookbook for core parameters (*DMQC_cookbook)* led by Ifremer. This contribution covers:

- The guidelines regarding DMQC workflow of Argo core data (pressure, temperature, salinity), providing a list of steps from getting R-files (uncalibrated real-time) from the GDAC to sending the D-files (calibrated delayed-mode) back;
- Description of examples of hydraulic or sensor problems;
- Creating a draft of a template (Clare Bellingham and Kamila Walicka) to produce a common DMQC report template for core Argo parameters of an individual float. The DMQC report template includes the detailed description of visual inspection of the float notes, comparison with satellite altimetry provided by CLS, the OWC configurations for the specific regions, diagnostic plots generated by OWC software, and scientific justification of the decisions made to determine a high-quality at-sea calibration for a given Argo float. The first draft of this template can be found in DMQC report template for core Argo data. The report template and Matlab codes used to generate it are available in a Euro-Argo ERIC GitHub repository dm-report-template.
- The summary of useful information about the water masses, reference data available in the Southern Ocean, aiming to help in the analysis of the floats in delayed mode, for instance, the parameterization of OWC software and the analysis of the results.

- Description of DMQC analysis and decision-making process of an example float deployed in the Southern Ocean.

Conversion of core Argo DMQC software OWC from Matlab to Python

From around September 2019, BODC (Edward Small, Kamila Walicka and Matt Donnelly) has started the conversion of the DMQC software currently available in Matlab to Python, as part of the MOCCA project. This initiative is closely associated with the results from a survey about the existing tool and methods used within DMQC Argo community, that was evaluated as a part of the EuroArgo RISE project WP2 last year. The survey identified barriers and opportunities to improve the efficiency and capacity of the overall community effort. For instance, it was identified that the existing DMQC software being written in Matlab - due to being paid-for licensed software - was a barrier for many institutions. A decision was reached to assess the potential for converting the OWC (Owens, Wong, Cabanes) Matlab code used for DMQC analysis to free software, with the widely used Python being the preferred language.

A year-long development project has consisted of several stages of conversion followed by phases of testing by a small group of DMQC operators and developers on blocks of code. The converted code is now functional, and ready for further User Acceptance Testing by broader Argo community to ensure it is ready for operational use. The development version of the OWC Python 'pyowc' package is currently available from the EuroArgo repository argodmqc_owc. The next steps to finalise the project are to complete the User Acceptance Testing involving the broader Argo community and reach an operational state. After this, development of the code will continue to improve its performance and enhance its functionality.

As part of the development work, there has been close collaboration with Guillaume Maze from Ifremer/LOPS to prepare the software package to be used as part of a Jupyter notebook. Combined with the potential to fully parallelize the analysis code, the conversion of OWC Matlab to OWC Python marks a step-change in capability and sets a new standard in quality control software development for the Argo community.

Deep DMQC

NOC (Kamila Walicka and Brian King) undertook a survey to review the current state of both the real-time and delayed-mode quality control (DMQC) approaches for deep Argo floats used within the global deep Argo community, producing a report based on this survey. This survey was a part of the EuroArgo RISE project WP3, deliverable D3.2_design_comparative_study_DMQC_methods_deep. One of the aims of this survey was to enhance an understanding of the current state of development and practice in deep Argo floats. The common tool used for delayed mode quality control of salinity data is the OWC software, which is currently used for the core Argo DMQC analysis. Some DMQC operators have used the standard core distribution of CTD reference data. Other groups have made use of ship-based CTD profiles collected at float deployments, however, such profiles are not available for all floats. Overall, the standardised methods and tools, mapping scales, reference data, interpreting estimates of salinity drift, thresholds and assignment of QC flags and error fields need to be investigated.

BODC started preparation of the 2nd European Argo-7th International Argo Delayed-mode QC Workshop for CTD data of core and deep Argo data, which was planned on 12-15 May 2020 in Liverpool, UK. The meeting was going to cover the review of Deep-Argo QC procedures used by the international deep Argo community. However, due to the ongoing pandemic of COVID-19 virus, this workshop has been postponed. The next DMQC workshop for deep Argo float data is going to be organised virtually in May 2021.

BGC DMQC

Progress on BGC DMQC has been started in December 2019 with UK Argo holding an oxygen and pH QC workshop between BODC DMQC operators and UK PIs, but progress since has been limited by staff availability and other more urgent demands. During 2021, more attention and progress on BGC DMQC, especially regarding oxygen and pH is expected under the EuroArgo RISE project WP4.

GDAC Functions

NERC Vocabulary Server

Following ADMT-20, BODC led the formation of the Argo Vocabulary Task Team, involving all those involved in managing the content and vocabulary experts at BODC. Since ADMT-20, most of the Argo reference tables have been converted into controlled vocabularies and hosted on the NERC Vocabulary Server (NVS) as part of the EU H2020 ENVRI-FAIR project. The remaining reference tables have been considered, problems identified and potential solutions scoped. Additionally, an assessment of potential opportunities to better constrain Argo metadata has been undertaken and will be investigated further during the course of the coming year. In addition, planning for enhancements of the NVS has been undertaken with development work planned for 2021, and initial training has been provided to Argo vocab editors. A dedicated presentation will be made to ADMT on this topic.

Southern Ocean Argo Regional Centre

Developing SOARC Partnership

In the past year a main focus for BODC has been leading a review of the European perspective on the functioning of Argo Regional Centres. In the context of SOARC this has led to a growth in the partnership to include AWI and additional representation from the US SOCCOM community, whilst Matt Donnelly has contributed to the Southern Ocean Observing system (SOOS) Data Management Sub-Committee (DMSC), and begun engaging with relevant SOOS working groups (WGs). Matt drafted a set of poster slides to promote Argo at the Weddell Sea and Dronning Maud Land WG meeting in October, which were delivered by Birgit Klein from BSH. The August 2020 meeting of the SOOS Observing System Design (OSD) WG was also attended with the aspiration of making better connections in this area of expertise.

Profile characterisation work

Through a collaboration with Kate Hendry, Rhiannon Jones and Luke Roberts at the University of Bristol, BODC has operationalised profile characterisation, with remaining development required to expose these for the benefit of users.

Regional data quality assessments

As part of the Euro Argo RISE project WP5, BODC will work with partners to establish regional data quality assessments in the Southern Ocean. This work has not progresses as far or as quickly as planned due to staff availability during 2020, but will be a major focus for 2021 and 2022.

References

King, R.R., D.J. Lea, M.J. Martin, I. Mirouze and J. Heming. The impact of Argo observations in a global weakly-coupled ocean-atmosphere data assimilation and short-term prediction system. Q J R Meteorol Soc. 2019; doi:10.1002/qj.3682

Martin MJ, King RR, While J, Aguiar AB. Assimilating satellite sea-surface salinity data from SMOS, Aquarius and SMAP into a global ocean forecasting system. Q J R Meteorol Soc 2019;145:705–726. https://doi.org/10.1002/qj.3461

von Schuckmann, K., Cheng, L., Palmer, M. D., Hansen, J., Tassone, C. , Aich, V. , Adusumilli, S. , Beltrami, H. , Boyer, T., Cuesta-Valero, F. J., Desbruyeres, D., Domingues, C., Garcia-Garcia, A., Gentine, P., Gilson, J., Gorfer, M., Haimberger, L., Ishii, M., Johnson, G. C., Killick, R., King, B. A., Kirchengast, G., Kolodziejczyk, N., Lyman, J., Marzeion, B., Mayer, M., Monier, M., Monselesan, D. P., Purkey, S., Roemmich, D., Schweiger, A., Seneviratne, S. I., Shepherd, A., Slater, D. A., Steiner, A. K., Straneo, F., Timmermans, M.-L., Wijffels, S. E.. Heat stored in the Earth system: where does the energy go? Earth Syst. Sci. Data 2020; 12, 3, 2013-2041. https://10.5194/essd-12-2013-2020

Wong Annie P. S., Wijffels Susan E., Riser Stephen C., Pouliquen Sylvie, Hosoda Shigeki, Roemmich Dean, Gilson John, Johnson Gregory C., Martini Kim, Murphy David J., Scanderbeg Megan, Bhaskar T. V. S. Udaya, Buck Justin J. H., Merceur Frederic, Carval Thierry, Maze Guillaume, Cabanes Cécile, André Xavier, Poffa Noé, Yashayaev Igor, Barker Paul M., Guinehut Stéphanie, Belbéoch Mathieu, Ignaszewski Mark, Baringer Molly O'Neil, Schmid Claudia, Lyman John M., McTaggart Kristene E., Purkey Sarah G., Zilberman Nathalie, Alkire Matthew B., Swift Dana, Owens W. Brechner, Jayne Steven R., Hersh Cora, Robbins Pelle, West-Mack Deb, Bahr Frank, Yoshida Sachiko, Sutton Philip J. H., Cancouët Romain, Coatanoan Christine, Dobbler Delphine, Juan Andrea Garcia, Gourrion Jerôme, Kolodziejczyk Nicolas, Bernard Vincent, Bourlès Bernard, Claustre Hervé, D'Ortenzio Fabrizio, Le Reste Serge, Le Traon Pierre-Yve, Rannou Jean-Philippe, Saout-Grit Carole, Speich Sabrina, Thierry Virginie, Verbrugge Nathalie, Angel-Benavides Ingrid M., Klein Birgit, Notarstefano Giulio, Poulain Pierre-Marie, Vélez-Belchí Pedro, Suga Toshio, Ando Kentaro, Iwasaska Naoto, Kobayashi Taiyo, Masuda Shuhei, Oka Eitarou, Sato Kanako, Nakamura Tomoaki, Sato Katsunari, Takatsuki Yasushi, Yoshida Takashi, Cowley Rebecca, Lovell Jenny L., Oke Peter R., van Wijk Esmee M., Carse Fiona, Donnelly Matthew, Gould W. John, Gowers Katie, King Brian A., Loch Stephen G., Mowat Mary, Turton Jon, Rama Rao E. Pattabhi, Ravichandran M., Freeland Howard J., Gaboury Isabelle, Gilbert Denis, Greenan Blair J. W., Ouellet Mathieu, Ross Tetjana, Tran Anh, Dong Mingmei, Liu Zenghong, Xu Jianping, Kang KiRyong, Jo HyeongJun, Kim Sung-Dae, Park Hyuk-Min. Argo Data 1999–2019: Two Million Temperature-Salinity Profiles and Subsurface Velocity Observations From a Global Array of Profiling Floats. Frontiers in Marine Science. 2020; 10.3389/fmars.2020.00700