UK Argo National Data Management Report for the 24th Argo Data Management Team meeting

Authors

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

● Contributing authors: Kamila Walicka, Clare Bellingham, Violetta Paba, Emma Gardner, Helen Snaith, Justin Buck

● Other team members: Roseanna Wright, Charlotte Dempster, Katy Baldwin, Danielle Wright, Jordan Atherton

With contributions from the wider UK Argo team by:

● Jon Turton and Fiona Carse (Met Office)

● Brian King, Nathan Briggs, Darren Rayner (National Oceanography Centre)

Data Management Team

The British Oceanographic Data Centre (BODC), part of the National Oceanography Centre (NOC), is the data assembly centre for UK Argo. It is funded primarily by the UK Natural Environment Research Council (NERC) of UK Research and Innovation (UKRI) and is responsible for data management of UK and Irish floats. In addition, UK Argo is a member of Euro-Argo. BODC hosts the NERC Vocabulary Server (NVS) which now hosts the Argo reference tables. BODC is also a member of the Southern Ocean Argo Regional Centre (SOARC).

Emma Gardner has been recently appointed to the role of Autonomous Platform Lead within BODC, having worked at the BODC for the past 10 years. She is providing a leadership and co-ordination role for Argo, Gliders and AUV (Autosub) platforms, with the aim of bringing together the existing teams in support of a more sustainable and unified approach. Currently, she oversees the BODC Argo team in this role.Clare Bellingham manages the BODC Argo DAC, loading and enabling the processing of new UK Argo floats and managing the day to day running of the DAC. This includes applying real time quality control corrections and working with BODC software developers on system requirements. Two new BODC data managers (Katy Baldwin and Charlotte Dempster) are currently being trained in the real-time data management of Argo data. Kamila Walika is the BODC Argo Delayed mode operator responsible for the DMQC analysis and application of required corrections for UK Argo floats. Kamila maintains the fast salty drifter list and works with the delayed mode community to ensure UK Argo data meets the Argo community standards. Kamila is the BODC representative for SOARC.

Roseanna Wright works on the Objective analysis reports and altimetry QC feedback from Ifremer

Violetta Paba focuses on supporting the Argo Vocabulary Task Team, with the newest contribution of two experienced members of the BODC Vocabulary Management Group, Danielle Wright and Jordan Atherton. Our role is that of managing NVS tasks pertaining Argo metadata lists and vocabularies and provide guidance on wider efforts to enhance the interoperability of the Argo (meta)data system.

Funding Outlook

Core BODC Argo national capability funding from NERC remained static for 2022-23 and is therefore still decreasing in real terms. There is additional funding from NERC associated with research projects and the floats they have procured, such as PICCOLO. NOC/BODC secured funding to develop data infrastructure for NKE BGC floats (the ASBAN UK project) purchased by NOC through NERC Capital funding. Efforts have continued to establish a clear plan for a more sustainable model of UK funding for Argo data management to support the UK contribution to the full-depth multi-disciplinary Argo array, but the funding situation remains challenging. BODC has been funded under the EU H2020 project ENVRI-FAIR (ended in mid-2023) to introduce the NVS vocabulary server to support Argo vocabulary management. BODC has been unable to source sustainable funding to support SOARC functions, so the ARC remains unfunded in the UK to date.

- BODC Argo team has been undertaking activities to provide a more secure long-term funding position and are keen to be part of future funding opportunities especially around strengthening the DAC in light of the expansion of OneArgo.
  - The team prepared a UK Argo business case document, focusing on current and imminent gaps in funding as a basis to identify new and sustainable funding streams to sustain the UK Argo program.
  - UK Argo team has drafted a OneArgo proposal for a 5-year plan of UK goals and aspirations. This document will be presented to the G7 meeting by the NOC lead Scientist for IL Global Observing, Katy Hill.
- Some of the key sources of funding for enhancements of the BODC Argo program are now coming to an end. BODC contributes resource required to manage the project and maintain the systems required for Argo to operate from its operational budget.
- UK based floats draw down data management funding available from NERC for projects utilising Argo float data, when the PI has completed a data management plan (DMP).

OneArgo Vision

A major new commitment is now required to enable the UK's contribution to OneArgo. This would significantly strengthen the UK's contribution to the Global Ocean Observing System (GOOS) and the Global Climate Observing System (GCOS). It would secure the UK a leading role in shaping and optimising OneArgo alongside other key G7 members USA, Canada, France, Germany and Japan.The UK Argo team are working on a UK briefing document to present to Defra (Department for Environment Food & Rural Affairs) on the UK funding landscape for Argo and the priority topic recommendations to implement OneArgo. These briefing notes will be used to inform key persons who will be present at the G7 Future of the Seas and Oceans Initiative Working Group Meeting in Tokyo, Japan in November.

Real Time Status

Data acquired from floats

BODC retrieves data for all UK, some Irish and assigned EU MOCCA floats from a number of sources and archives these for further processing. BODC currently processes data from floats with Argos communications, Iridium Rudics and Iridium Short Burst Data (SBD) from a diverse fleet of floats manufactured by TWR, SeaBird, NKE and SOLO.

To date this year UK Argo have deployed 43 floats with a further 19 expected for 2023. These are a mix of core and BGC where the core floats are Apex, Apex with RBR and NKE Arvor, and the BGC floats are NKE Provor III and 1 Navis N2.
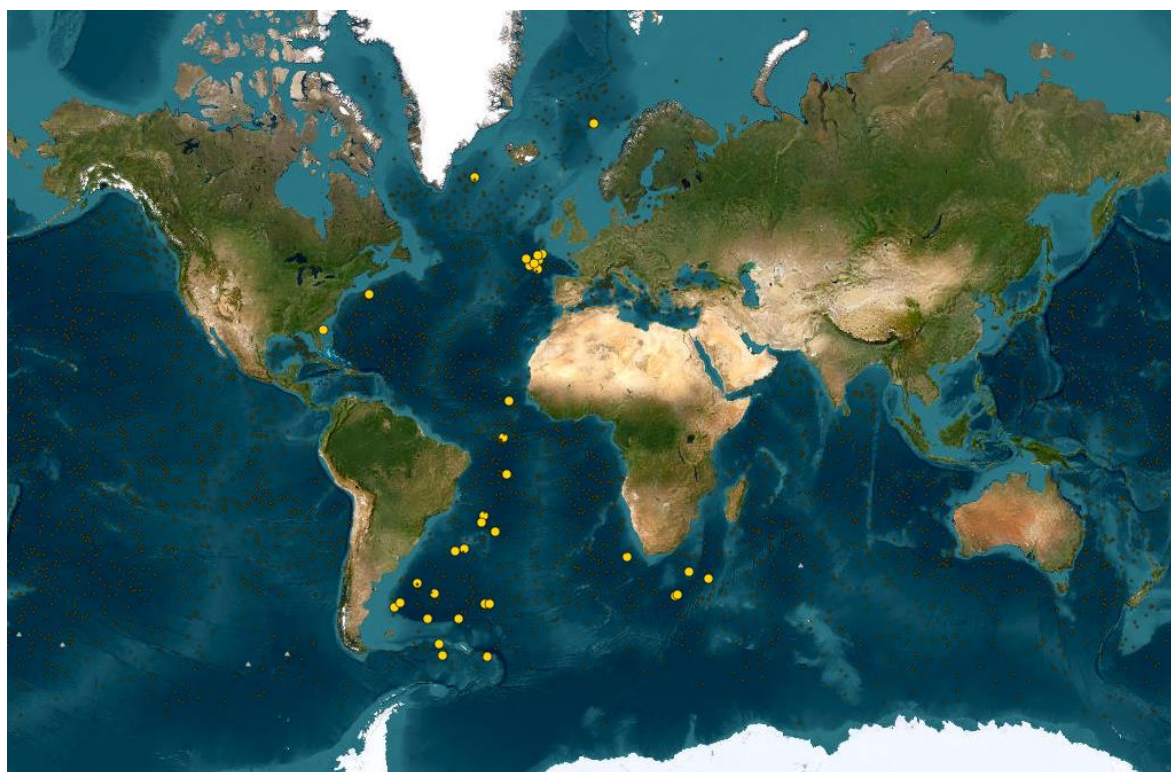
Map showing current locations (16/10/2023) of UK Argo deployments Jan-Oct 2023 (https://fleetmonitoring.euro-argo.eu/dashboard)

BODC is currently processing the temperature and salinity coming from the RBR sensor in real time. BODC has 21 RBR Apex floats in their fleet, where 10 of them are currently active. UK Argo is planning to deploy another 8 RBR Apex floats by spring 2024.

BODC is currently processing the temperature and salinity coming from the UK Deep Argo floats using the SBE61 sensor in real-time. BODC has 22 deep Argo floats in their fleet, where all of them are currently inactive. NOC deployed a SOLO deep float in 2022 which is the first float of this type in the UK Argo fleet. BODC has been working with software developers with help from AOML to enable capability to process and deliver the data from this float to the GDAC. This work is nearing completion and files can be seen on the GDAC for float 2903791

Near real-time data delivery

Processing and delivery of incoming data is normally set up within one week of deployment where the capability already exists for a given float type. We have found that development work for new float types can be significant and manufacturer's data decoders and discussions with other DACs have been important.

BODC continue to operate two parallel processing chains. The BODC-developed chain processes and delivers all UK Navis, Apex and SOLO floats. A version of the Coriolis processing chain has been implemented in the BODC software stack and processes and delivers UK NKE, the EU MOCCA NKE that BODC have responsibility for and Irish NKE floats.

During 2023 BODC DAC completed the development work required to process NKE BGC floats purchased through the Natural Environment Research Council (NERC) and NOC investment project UK Atlantic Sector Biogeochemical Argo Network (ASBAN-UK) meaning that these floats now deliver real time core and BGC data to the GDAC. Following the BGC workshop earlier in the year DAC operators were able to calculate the DOXY gain for these floats meaning that DOXY adjusted is also delivered in real time to the GDAC. Delivery of real time BGC parameters has enabled NOC scientists to strategically deploy Argo floats alongside gliders and use the Argo BGC data as intelligence to pilot the gliders. This use case scenario was recently utilised in the Gulf Stream as part of the ongoing C-Streams project

Data issued to GTS

BODC delivers core data, and oxygen from NKE floats, 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.

One issue we have found impacting our GTS delivery is a recent issue with Apex floats not making a fix with GPS on the first attempt. Our experience has seen that the fix is made shortly before the float dives again and so cannot be retrieved until 10 days later, so while we have the science data at the DAC we cannot process and deliver the profile due to the lack of GPS fix. This is something that UK Argo have started to discuss with Teledyne but a resolution has not yet been found and 20 of our floats have found to be affected by this during some cycles.

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. Delivery of BGC profile data for most floats and many trajectory files are still pending.

Core

BODC endeavours to address any RTQC from Objective Analysis reports and Altimetry QC issued by Ifremer and OceanOps on a regular basis and made updates to the meta files following reports from the GDAC file checker. Additionally, BODC is undertaking visual inspection of the core Argo parameters from core, deep and BGC float types of all profiles which then undergo further DMQC analysis.

The workflow to implement real-time adjusted corrections to RBR Argo floats is still under development. This activity will be strongly dependent on very limited funding and resources in BODC.

BGC

BODC developed their workflow of applying the real-time adjusted corrections to the dissolved oxygen data using the Sage02 software. Currently, the corrections have been applied to 10 recently deployed BGC NOCBio Provor floats.

The workflow to implement the real-time adjusted corrections to Arvor and Apex BGC Argo floats is still under development.

Deep

The workflow to implement the real-time adjusted corrections to Apex and Solo Argo floats is still under development.

Uk Argo metadata update

BODC Argo floats have had a metadata review from the wider community

Submitting of real time adjusted and d-mode data from external partners

The BODC Argo DAC function currently interacts with DMQC operators through two different modes of operation. The first is internal BODC DMQC operators who directly submit DMQC decisions to the BODC Argo System, and for which updated D-mode NetCDFs are automatically generated and submitted. The second mode is used for MOCCA floats managed through the Coriolis processing chain at BODC, both internal and DMQC operators from Ifremer, OGS and BSH submit updated NetCDFs which are archived within BODC and submitted to the GDACs.

Additionally, BODC has received support from LOV in processing Irradiance (d-mode), Chlorophyll and Backscatter data of the real-time adjusted mode.

All float data submitted to the CODC Argo DAC are then 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.

Delayed Mode QC status

- ### CORE

From January 2023 BDOC Argo submitted to GDAC 15 core Argo floats with ~1600 new core profiles in D-mode. The DMQC analysis has been focused to process the suspected salty drifting floats listed in the Argo greylist.

Additionally, in 2023 some of the BODC Argo floats coming from MOCCA project has been continuously DMQC-ed by external European partners. We have received and submitted to GDAC 8 analyzed core Argo floats (with ~380 new D-mode profiles) from BSH.

The key focus for BODC is to implement the workflow of DMQC Deep and RBR Argo in BODC including analysis of these floats and continue DMQC analysis of core Argo floats. These activities will be strongly dependent on very limited funding and resources in BODC.

- ### DEEP

BODC have greatly improved their understanding of the current procedures and guidelines for RT and DMQC of deep Argo data. BODC performed initial DMQC analysis for two Deep Argo floats following the most recent recommendations from the Argo manual. However, these data have not been submitted to the GDACs because the BODC processing chain and BODC database is not yet adapted to new deep Argo procedures.

- ### BGC

The BODC Argo team has greatly expanded their knowledge of the DMQC analysis of BGC Argo floats. BODC has started implementing the workflow of the DMQC BGC procedures of the DOXY parameter for the BGC Argo floats.

The UK Argo fleet currently has 64 BGC Argo BODC floats requiring DMQC analysis (>1 year old). However, BODC is currently able to process in DMQC only parameters from 17 BGC floats. The remaining 47 legacy BGC Argo floats are still not available in R-time in BODC. These require additional development works in the BODC Argo processing chain.

In early 2023, BODC Argo performed DMQC analysis of Dissolved Oxygen parameters from 13 MetBio NKE float and submitted to GDAC about 4500 profiles in D-mode. Our further ambition is to start the DMQC analysis of other BGC parameters.

Software in DMQC
BODC has regularly adopted the latest reference databases for DMQC analysis of core and deep Argo floats (CTD_for_DMQC_2021V02 and ARGO_for_DMQC_2022V03).

BODC has adopted the procedures and SAGE_O2 Argo [software](#) for estimates of the Gain of the DOXY parameters of Argo floats. Additionally, we also implemented the procedures from https://github.com/catsch/DM_FILLER allowing applying corrections in D-mode BGC floats, generation and population of the D-mode NetCDF files.

Improvements to the quality of the UK Argo fleet data
The UK core Argo fleet data went through the international DMQC audit run by external partners from the DMQC core Argo group. The audit was motivated by the fact that a higher percentage of SBE CTDs are now experiencing sensor drifts, which may not be easily identifiable by only examining individual time series. All identified BODC profiles with some issues were reviewed. Any additional revisions or corrections have been completed and re-submitted to the GDACs. BODC was not able to resubmit the few remaining floats from the beginning of the Argo project in early 2000s due to technical issues with the float data related to a depracated NetCDFversion (Format 2.2) of Argo profiles which are not acceptable by the GDAC filechecker.

Support to national programs

The strategy adopted to deliver the support to national programs focused on ensuring a high-quality approach and the progressive enhancement of expertise. This supports OGS in implementing the DMQC-PCM software, contribution to the working groups. Moreover, BODC also provides support in analysis of the dissolved oxygen parameters of Argo floats to IOPAN.

Working groups

BODC actively contributed to activities related to the Abrupt Salty Drift (ASD) group, focusing on estimating the best practices, guidance and examples on how to treat salinity data that are affected by sensor drift to produce optimal adjustment in d-mode. This involved actively contributing to updating the shared list of floats affected by the salty drift and reviewing best practices and procedures for DMQC operators of core Argo floats.

Furthermore, BODC and NOC contributed to the research article with the collaboration of the international Deep Argo community on "*The implementation plan of the global Deep Argo array to measure the full ocean volume*". The article is currently under review in Frontiers journal.

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) workflow which is run daily and produces 2 analysis days and a 7-day forecast.
- A coupled ocean/atmosphere/sea-ice/land global prediction system is now operational for producing the main Met Office weather forecasts. This coupled NWP system assimilates data in all components of the coupled model, including Argo data in the ocean component. These data therefore affect both weather forecasts and short-range 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.
- Initial conditions for coupled monthly-to-seasonal forecasts are taken from the global coupled NWP system so the Argo data are used to initialize these forecasts and are used in ocean reanalyses.
- Argo data are also used in the initialization of ocean conditions in climate models run to make decadal predictions;
- Near-surface Argo data are used to validate the output from the Met Office's OSTIA (Operational Sea Surface Temperature and Sea Ice Analysis).

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

- A new paper describing work to develop an ensemble ocean system, including improvements to the way data (including Argo) are assimilated, has been published (Lea et al., 2022);
- A paper was published on OSSEs to investigate the potential impact of expanding the Argo array (Mao et al., 2020);
- 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 (Ford, 2021);
- A PhD project is currently looking at the impact of real BGC Argo data in a global physical-biogeochemical model. The BGC Argo data are assimilated into the model and the impact on air-sea CO_2 fluxes is being investigated.
- A paper was published jointly with the University of Reading on the application of a simple smoother algorithm to make better use of Argo data in ocean reanalysis (Dong et al., 2021).
- A 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. Another paper was published investigating impact in FOAM and the Mercator system of satellite SSS assimilation which used Argo for assessment (Martin et al., 2020).

In the Hadley Centre for Climate Science and Services, Argo data is used in the following products:

- EN4 contains in-situ 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/). The latest version is EN.4.2.2, which includes a fresh download of all the source data and a substantial update to the XBT/MBT correction schemes. EN.4.2.2 contains four ensemble members where previously there was only two. There is also a new product user guide (based on both the Argo Users' Manual and the HadIOD user guide), including FAQs and example code. EN4 is also forming part of a GEWEX EEI project - comparing Ocean Heat Content calculated from reanalyses, in situ data and satellite products (the project website is https://sites.google.com/magellium.fr/eeiassessment/dissemination/documents?authuser=0).
- 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, seasonal forecasting, decadal forecasting, climate model initialization and evaluation.

National Oceanography Centre (Brian, Nathan)

At NOC we produce a 4-D global map of Argo T and S data at 2 degree lat and long resolution from 60S to 60N. The data are gridded in 10-day windows using objective mapping on sigma-1 or neutral density levels and then interpolated back to 20 dbar vertical resolution. This is generally updated towards the end of each calendar year. A time series of global heat content is calculated and reduced to annual averages and then incorporated into the synthesis of global heat content calculations led by K von Schuckmann. The full 4-D gridded fields can be made available by contacting Brian King at NOC.

A two-year NOC-led project called GLOBESINK started in August 2022 to generate a global dataset of particle size and downward particulate organic carbon flux from BGC Argo measurements of optical backscattering. This dataset contributes to the wider NERC BIO-CARBON programme, which aims to improve our ability to predict changes in biological carbon update by the oceans. One output of the project will be a publicly available particle dataset using BGC Argo data through 2022 (to be delivered in 2024). NOC aims to maintain this product in the future through single center NERC bid AtlantiS. New BIO-CARBON projects PARTITRICS and IDAPro, led by NOC and University of Southampton, will deploy two UK BGC Argo floats and fund their data delivery and QC, and also deploy three French BGC Argo floats. The float data will be used for estimates of primary production, net community production, and downward POC flux as part of 2024 BIO-CARBON fieldwork.

Currently, three NOC-led PhD projects have a large component utilizing BGC Argo data. One focuses on net community production in the Weddell Gyre, another is exploring methods to optimally interpolate subsurface chlorophyll data, and a third is looking into the drivers of variability in the remineralization depth of sinking organic carbon in the ocean. A fourth NOC-based project led by the University of Southampton will develop methods to QC and correct pH data from BGC Argo.

GDAC Functions: Argo NERC Vocabulary Server (NVS) activities

As the ENVRI-FAIR project came to an end in June 2023, BODC completed the transfer of all the original Argo metadata lists onto NVS collections, and full governance was handed over to the designated NVS Editors from the wider Argo Data Management Team/Argo Vocabulary Task Team.

BODC continues its role of NVS support and guidance to the ADMT community, routinely monitoring AVTT GitHub tickets, answering queries and making suggested edits, and participating in technical meetings with the team and its stakeholders.

To unleash the full potential of a metadata system underpinned by the NVS, tools and templates are being developed by various groups in order to incorporate the NVS machine-readable technology into current processes and enhance automation; specific efforts include the upgrade of the GDAC file checker, and the standardisation of sensor metadata harvesting and parsing from manufacturers to DACs.

As BODC and Argo hold an ambitious vision in pioneering cross-domain interoperability and FAIR development of this remarkable data system underpinned by the NVS, additional funding is being sought to pursue these objectives and further raise Argo's status in the Oceanographic Observational Networks of the modern world.

Value Added items

Webpages

- NOC continues to maintain the UK Argo website (www.ukargo.net)
- BODC Argo website (https://www.bodc.ac.uk/data/hosted_data_systems/argo_floats/)
- NVS VocPrez website (http://vocab.nerc.ac.uk/)
- Argo Vocabulary Task Team (AVTT) GitHub space: https://github.com/orgs/nvs-vocabs/teams/avtt
- Facebook page (www.facebook.com/UKArgofloats/)
- Twitter account (twitter.com/ukargo)
- NOC maintains the SOARC website (www.soarc.aq)
- NVS Argo collections (new):
- R14 - Argo technical parameter names
- R18 - Argo configuration parameter names
- R28 - Argo controller board types and generations
- R40 - Argo Principal Investigator (PI) names

Software tools

- A Python implementation of the "OWC" salinity calibration method traditionally available for Matlab used in Argo floats Delayed Mode Quality Control https://github.com/euroargodev/argodmqc_owc
- A software for an infrastructure agnostic set of common BGC parameter derivation equation functions https://github.com/euroargodev/bgc_derivation
- Real time QC automated tests for Argo data. https://github.com/euroargodev/argortqcpy
- The quality assessment method in the Southern Ocean (SO) uses the pre-classified core Argo float and climatological data belonging to similar water mass regimes using the Profile

Characterization Model (PCM). https://github.com/euroargodev/DMQC-PCM/tree/SO_assesment

- This repository includes the report template and Matlab codes used to generate plots required in the DMQC report for core Argo parameters. https://github.com/euroargodev/dm-report-template
- BODC has provided the material to update the 'Argo vocabulary server' web page on the Argo data management website: http://www.argodatamgt.org/Documentation/Argo-vocabulary-server

Manufacturer engagement

NOC (Brian King) Since the 2022 ADMT NOC has been engaged with SBS and RBR to develop machine-readable files for exchange of metadata between platform and sensor suppliers and DACs or other users. The present status of this work will be presented for discussion at ADMT. The development so far is shared on GitHub: https://github.com/euroargodev/sensor_metadata_json and a document describing the project and progress has been distributed for comment on the argo-dm mailing list.

Workshops and training

BODC Argo team attended the 1st DMQC workshop of the BioGeoChemical parameters. The meeting was focused on reviewing the available procedures, tools and methods of processing the BGC Argo parameters in delayed mode and also upskill the DMQC operators.

The key outcome of this meeting for BODC and NOC were:
- improve the international collaboration with the BGC Argo community and
- BODC Argo team for the first time have started DMQC analysis of Oxygen data in D-mode. As for the time of writing BODC Argo team submitted 695 **BGC Argo profiles** to GDAC. The profiles come from the MetOffice BGC Argo floats.

DMQC Deep Argo workshop. This workshop was intended for delayed-mode operators of Deep-Argo floats to share their experiences regarding two important aspects of DMQC for Deep-Argo data: the application of a CPcor adjustment for salinity from SBE CTDs; and the evaluation of sensor drift to the expected Deep-Argo salinity accuracy of 0.004.

Kamila and Emma participated at the 10th International EuroGoos conference in Galway. Kamila Walicka presented a poster on Implementing a machine learning method based on a profile classification approach in the delayed mode quality control of Argo floats.

Kamila and Brian as the experts in the DMQC analysis of core Argo parameters, participated at the series of the International DMQC discussion meeting led by CSIRO. These virtual discussions helped to promote collaboration between DMQC operators and interested members of the Argo community. This forum gives an opportunity for newer operators to improve their skills and get advice on concerning and difficult floats, promote a sense of community, and contribute to the adoption of more consistent DMQC practices.

References

Lea, D.J., While, J., Martin, M.J., Weaver, A., Storto, A. & Chrust, M.(2022) A new global ocean ensemble system at the Met Office: Assessing the impact of hybrid data assimilation and inflation settings. Quarterly Journal of the Royal Meteorological Society, 148(745), 1996-2030. https://doi.org/10.1002/qj.4292