UK Argo National Data Management Report 2018

Report to the Argo Data Management Team – ADMT-19

Author list

The UK Argo data team (British Oceanographic Data Centre, National Oceanography Centre)
    Contributing authors: Robin McCandliss, Matt Donnelly, Clare Bellingham, Violetta Paba
    Other team members: Katie Gowers, Paul McGarrigle, Elizabeth Bradshaw, Justin Buck, Sarah Chapman, Roseanna Wright

With contributions from the wider UK Argo team by:
    Jon Turton and Fiona Carse (UK Met Office)
    Brian King (National Oceanography Centre)
    Giorgio Dall'Olmo (Plymouth Marine Laboratory)

Status

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).

General Status

BODC Argo Team

The team currently has two members who work full time on Argo (Matt Donnelly and Clare Bellingham). Katie Gowers and Violetta Paba both work part-time on Argo (0.4 FTE and 0.3 FTE, respectively). Liz Bradshaw and Paul McGarrigle provide support to DAC operations by covering daily processing now and again, as required.

Early in 2018, Matt Donnelly took on responsibility for leading the BODC Argo team and used the opportunity to assign responsibility for defined areas across the team. Responsibility for the DAC workstream has been assigned to Clare Bellingham for the core floats and Violetta Paba for the BGC floats. Katie Gowers is leading the software development workstream, with Clare Bellingham as part of her development team. Justin Buck remains involved on a reduced basis, supporting the regeneration of delayed-mode QC and as a source of expertise and advice. Additional short-term support has been provided to the team by Sarah Chapman working on developing a Southern Ocean deployment guide, and Roseanna Wright who has supported data screening.

The Argo Lead role comes with a significant management overhead, requiring time for project management, attending meetings, preparing reports, reviewing finances, responding to new funding initiatives, managing the team and prioritising their work.

It is worth noting that over the past year there was substantial training time required for Clare and Violetta, to develop their capability in Argo processes. Clare and Violetta have been learning how to do DMQC, and Clare has been learning the software development protocols required for development of the system. In addition to Argo specific development work, the BODC Argo team has completed a system-wide Matlab upgrade during early 2018. None of these activities were trivial.

BODC is currently recruiting an addition to the Argo team in order to address DMQC requirements for MOCCA and UK Argo, and free up the existing team members to focus on NRT operations and development of the system to handle new float types.

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 will continue providing additional sources of data management funding, such as from the ORCHESTRA, ACSIS, BoBBLE and RoSES projects. BODC receives funding from the Euro-Argo ERIC MOCCA project for the European Union floats that are managed by BODC, as well as some funding from the EU H2020 AtlantOS project to support delayed-mode QC of Argo extensions.

During 2018, UK Argo was involved in a successful EU bid culminating in Euro-Argo Research Infrastructure Sustainability and Enhancement (Euro-Argo RISE), where BODC's involvement will encompass DMQC coordination, knowledge sharing of DMQC processes, evaluation and development of DMQC tools and to introduce Southern Ocean regional data quality assessments. BODC was also part of the successful ENVRI-FAIR bid, with the intention of using the NVS vocabulary server to support Argo vocabulary management. BODC continues to seek additional sources of funding to support SOARC functions, some of which will be provided by the EU MOCCA and Euro-Argo RISE projects, but a long-term solution is yet to be identified.

DAC Functions

Data acquired from floats

BODC retrieves data for all UK, Irish, Mauritius and EU MOCCA floats from a number of sources and archives these for further processing. Where possible, processing of arriving data is normally setup within one week of float deployment, and during the past year this has typically been achieved much more quickly for those floats where capability exists. Please refer to table 1 for the types of communications used for different floats.

Progress in the past year:
During the past year, BODC has improved management of incoming data for floats that we are not currently distributing data for to enable float position monitoring as both part of the full delivery of data and as an interim monitoring check.

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 kindly providing the processing for a dozen PROVOR BGC floats and delivering the core data to the GTS on BODC's behalf until BODC can take on the management of these floats. The results of improvements to system performance, a proactive approach to loading floats to the BODC Argo System, and increased frequency of data processing and delivery can be seen in figure 1.

Progress in the past year – general processing:
BODC is currently distributing data to the GTS for c. 221 floats at the time of writing, which is an increase from 151 in November 2016 prior to the start of processing floats for the EU MOCCA project. During 2018, BODC has sustained automated data processing four times a day, rather than twice a day as was the case prior to 2017. Generation of BUFR files was transferred from BODC to the Met Office during 2018 following the successful development of the BUFR converter (see below). BODC ceased issuing TESAC messages to the GTS on 1st July 2018, as agreed by ADMT and AST in 2017. Delivery of core data for floats not currently processed has been a major focus during the past year, particularly for all APF9I/N1/N2 floats, but also work on APF11, and remains our highest priority for completion.

Progress in the past year – BUFR converter:

In support of a new Python netCDF-to-BUFR converter, BODC makes all netCDF files available directly to the Met Office via SFTP and this is now occurring on a routine basis four times a day. Testing was completed during early 2018, and the Met Office took over responsibility for the BUFR generation and distribution on the GTS on 5th June 2018. The BUFR messages from the UK Met Office have the header "IOPXII EGRR". During October 2018, the Met Office sent the Python conversion code to Megan Scanderbeg and Rebecca Cowley, to be made freely available on the ADMT software tools web page. The code has been designed to be extensible, where capability for secondary temperature/temperature and salinity, and oxygen profiles can be added plus other BGC variables when required. The Met Office plans to extend the software to include secondary T/S profiles and oxygen profiles during 2019.

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 finish the work ensuring all floats with an APF9I/N1/N2 controller board are effectively managed within the BODC Argo System, and the main attention will then turn to floats with other controller boards such as the APF11. BODC is seeking to collaborate with other DACs in the development of APF11 decoders.

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 not yet completed the conversion to v3.1 for all file types. With the exception of a small number of legacy files, all metadata files are delivered in V3.1. Significant portions of the core profile files are delivered in v3.1, with the remainder currently being addressed. Please refer to table 1 for the types of float and whether they are being fully processed.

Progress in the past year:

BODC has progressed the delivery of core profile files in v3.1 netCDF files, with now ~ 80 % in v3.1 compared to ~ 60 % at the last ADMT as per the last Report on Anomalies. Progress on the conversion has been slowed by other priorities but work to complete the remainder of the profile files is currently underway, with a submission of ~9300 further profiles in late November 2018. Additional progress has been made on delivering tech files in v3.1, now on 30% with additional batches planned for early 2019. A significant backlog on quality control feedback from Objective Analysis and Altimetry QC reports has been mostly cleared at the time of writing and will be completed in the near future.

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.

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

Float type/controller	Comms	Total no. of deployed floats	Total no. of active floats	No. of active floats with ice detection	Mission of active floats					Total no. of active floats being fully processed
					Core only	Core + NST	Core with RBR CTD	Core + oxygen only	Core + other BGC	
Martec Provor	Argos	26	-	-	-	-	-	-	-	-
MetOcean NOVA	Iridium	1	-	-	-	-	-	-	-	-
NKE Arvor	Argos	5	1	-	1	-	-	-	-	1
NKE Provor *	Iridium SBD	13	10	-	-	-	-	-	10	10
NKE Arvor	Iridium SBD	76	74	-	-	74	-	-	-	74
NKE Deep Arvor	Iridium SBD	2	-	-	-	-	-	-	-	-
SBE Navis N1	Iridium Rudics	14	7	7	7	-	-	-	-	6
SBE Navis N1 with BGC	Iridium Rudics	4	1	-	-	-	-	-	1	0
SBE Navis N1 with oxygen	Iridium Rudics	8	8	-	-	-	-	8	-	0
SBE Navis N1 with radiometer	Iridium Rudics	3	3	-	-	-	-	-	3	0
TWR Apex APF7	Argos	8	-	-	-	-	-	-	-	-
TWR Apex APF8	Argos	252	1	1	1	-	-	-	-	1
TWR Apex APF9A (7 types)	Argos	266	139	19	28	111	-	-	-	139
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	-	-	-	0
TWR Apex APF11	Argos	7	7	-	7	-	-	-	-	0
TWR Apex APF11	Iridium Rudics	12	9	n/k	-	-	2	-	7	0
TWR Deep Apex APF11	Iridium Rudics	10	2	n/k	-	-	-	2	-	0
TOTAL		666	260	29	46	188	2	10	21	233

* = processing courtesy of Coriolis

Figure 1: Summary of all BODC internal processing times from time of profile to processing completion, indicative of GTS delays
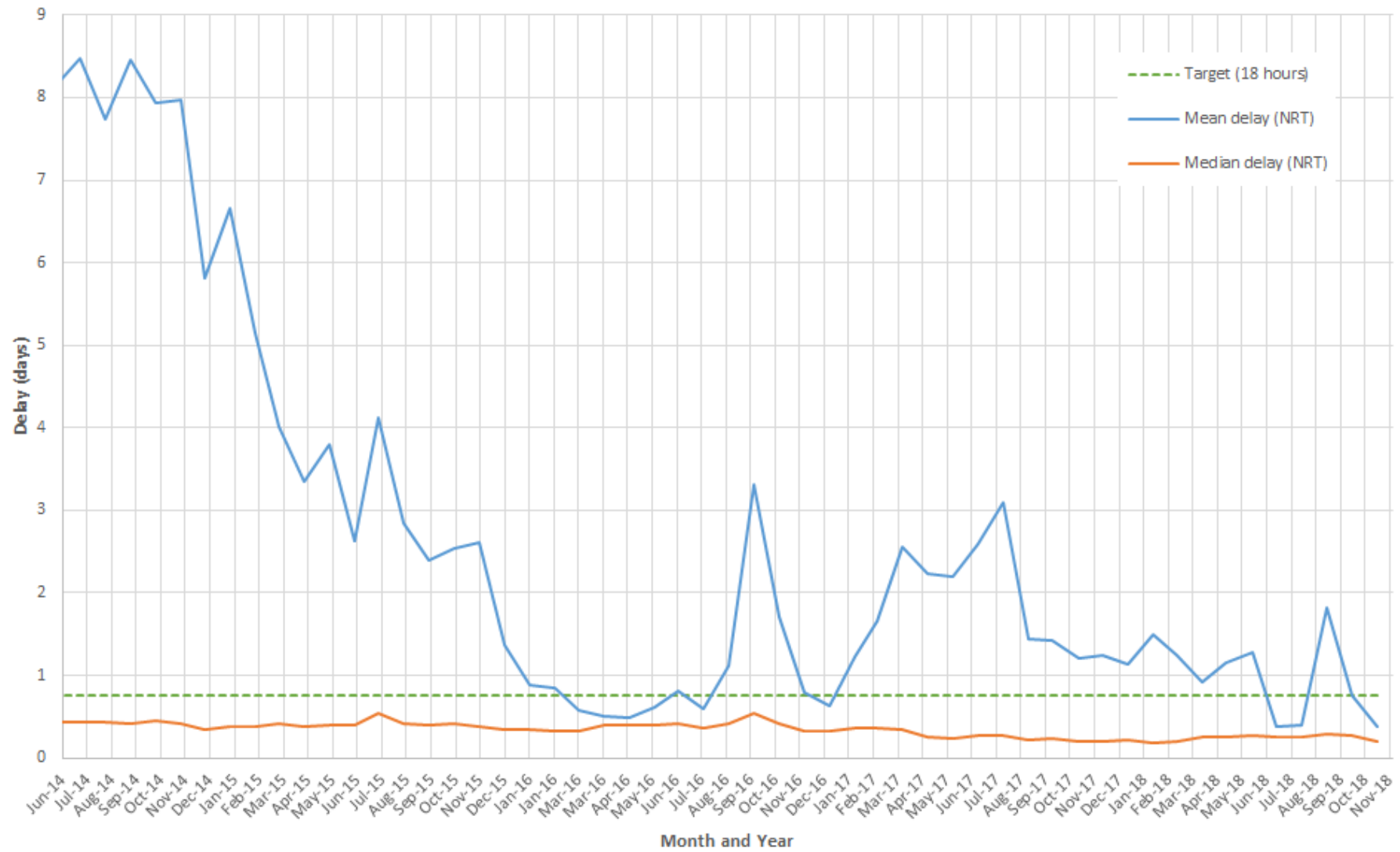
Data issued for delayed-mode QC

All delayed-mode QC on BODC hosted floats is performed within BODC, with the exception of ~35 floats funded under the Euro-Argo MOCCA project.   Currently BODC is only capable of providing data for delayed-mode QC for core data, with work required to finish the delivery of biogeochemical parameters in v3.1. Again, the exception to this are a dozen PROVOR floats that Coriolis is kindly hosting on BODC's behalf.   See section 2 of this report for the status of delayed-mode QC.

For any given float, if the R-mode or A-mode file is available following real-time QC, then any profiles that have been through delayed-mode QC will be available as D-mode files. This applies to float profile files that are in either v2.2 and v3.1 format.

Progress in the past year:
Progress on v3.1 profile files has made more delayed-mode files available in v3.1. BODC has also begun liaising with relevant Euro-Argo partners to manage the delayed-mode files for MOCCA floats as they become eligible for delayed-mode QC.

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. See section 2 of this report for the status of delayed-mode QC.

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)

Elaine McDonagh is also engaged in using Argo data, bidding for float funds, planning strategies, leading analyses and mapped products

The applications of Argo data at NOC 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).

Argo 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;
- and to better understand the nitrogen cycle in oxygen minim zones.

Data Products

National Oceanography Centre

Elaine McDonagh is engaged with 4-D global fields of mapped Argo T and S (Desbruyères et al. 2017), but they are not currently publicly available, however Elaine can be contacted by any interested parties.

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/);
- HadIOD 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, HadIOD obtains sub-surface profile data from EN4. Public release of the data are expected during 2017. HadIOD is expected to supersede the HadGOA data product, which has not been updated for approximately 7 years (http://www.metoffice.gov.uk/hadobs/hadgoa/).

The datasets are used for climate and global change studies, including ocean heat content analysis.

Delayed-Mode QC

Following the regeneration of delayed-mode QC capability though a software and procedural review last year, efforts this year have focused on training additional DMQC operators. This has included BODC (Justin Buck, Matt Donnelly, Clare Bellingham and Violetta Paba) and NOC (Brian King) contributing to the preparation for or attending the 1st European DMQC Workshop, as well as in-house training for Clare Bellingham. BODC use OW software for delayed-mode quality control with the latest reference data available from Coriolis (CTD climatology and Argo profile climatology for guidance).

Following advice from the wider UK Argo team, and particularly from Brian King, we are currently working with the following considerations in-mind:

- To facilitate increasing knowledge/experience of regional oceanography we are addressing DMQC on an ocean basin-by-ocean basin basis. DMQC started with the Indian Ocean last year, with focus moving onto the South Atlantic during 2018, followed by the Southern Ocean and North Atlantic.
- To facilitate increasing knowledge/experience of the use of the OW software and to avoid applying excessive corrections, we are tackling floats with simple pathologies first, then moving onto floats with increasingly complex pathologies as DMQC skills improve within BODC;
- Pursuing delivery of DMQC for the EU MOCCA project floats to meet project deliverables irrespective of basin or complexity.

During Autumn 2018 BODC started the first significant batch of South Atlantic DMQC since 2013 and these are due for submission to the GDACs soon, alongside work on the initial batch of DMQC on EU MOCCA floats. In addition, BODC has started handling DMQC submissions from Coriolis, BSH and OGS as part of the EU MOCCA project.

Violetta Paba at BODC has begun using the SOCCOM/MBARI tool SAGE-O2 to generate the capability at BODC to perform A-mode and eventually D-mode QC on oxygen data.

BODC has recruited a new member of staff who will have a primary focus as a DMQC operator. This person will start in January and as such BODC expects to achieve a significant improvement in the DMQC situation during 2019.

As part of the EU Euro-Argo RISE project, BODC plans to compile a report on globally available Argo DMQC tools to provide an understanding of the current state of development. The aim is to cover matters such as the 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 improving the sustainability of DMQC tool development within the global Argo data system and at the European level. All organisations with DMQC tools are invited to contribute information to this report. Additionally, BODC will be contributing to the further development of oxygen and pH QC procedures.

GDAC Functions

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."

BODC seeks to receive input from the Argo Data Management Team about the perceived issues with existing vocabularies (Argo reference tables) and additional services that could be developed on-top of the NVS to serve Argo and wider applications. Whilst BODC has the infrastructure to provide technical governance, a content governance structure will need to be discussed, agreed and clearly established.

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.

BODC was present at the BGC Float Workshop in Seattle where it highlighted the need to ensure that the legacy of the SOCCOM project is sustained into the future by international coordination in the region through SOARC. Please refer to the workshop report (when available) for further details.

As a result of increased funding for DMQC-related activities and SOARC, BODC aims in the coming 18 months to 4 years to:

- Provide support to other national programmes with DMQC, with a focus on the Southern Ocean (e.g. orphan floats);
- Improve high latitude Southern Ocean deployment opportunities by, for instance, compiling a guide to deployments in the Antarctic Treaty System area, a draft of which is available;
- 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).

As part of achieving the above, in addition to collaboration with SOARC partners, BODC is in the process of establishing collaborations with the University of Liverpool and University of Bristol.

References

Desbruyères, D., E.L. McDonagh, B.A. King, and V. Thierry, 2017: Global and Full-Depth Ocean Temperature Trends during the Early Twenty-First Century from Argo and Repeat Hydrography. *J. Climate,* **30**, 1985–1997, https://doi.org/10.1175/JCLI-D-16-0396.1