



How do I cite Argo data in my publication?

Justin Buck¹ (juck@bodc.ac.uk), Thierry Carval², Kenneth Casey³,
Frederic Merceur², Megan Scanderbeg⁴

1. British Oceanographic Data Centre (BODC) UK, 2. Ifremer France, 3. National Centres for Environmental Information (NCEI) USA, 4. Scripps Institution of Oceanography USA



Figure 1: The Argo array as of June 2015

Introduction

Since 1998 the Argo project has collected >1.3 million ocean profiles. Argo data are used in applications ranging from assimilation in operational ocean models to climate-grade heat content calculations of the upper ocean. Argo data have been used in academic papers 2123 times. Argo data underpinned many findings of the 5th IPCC assessment so are influencing national and international government policy. Consequently it has become essential that Argo data are citable in an unambiguous manner to ensure reproducible and transparent science.

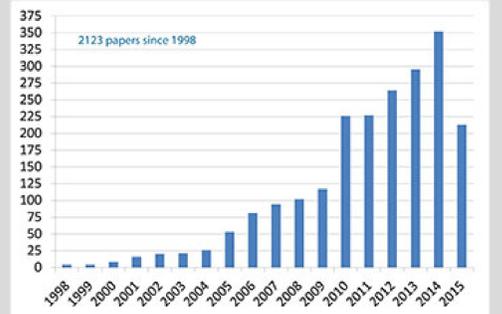


Figure 2: Publications that use Argo data by year

Data growth and updates

The Argo dataset is an aggregation of data supplied by national level Data Assembly Centres (DAC) to two mirrored Global Data Assembly Centres (GDAC). The dataset is “dynamic” in the sense that it grows and updates constantly as new data are collected and quality information/calibrations are updated in delayed mode. The GDACs do not support file versioning so store and supply the most recent version of data.

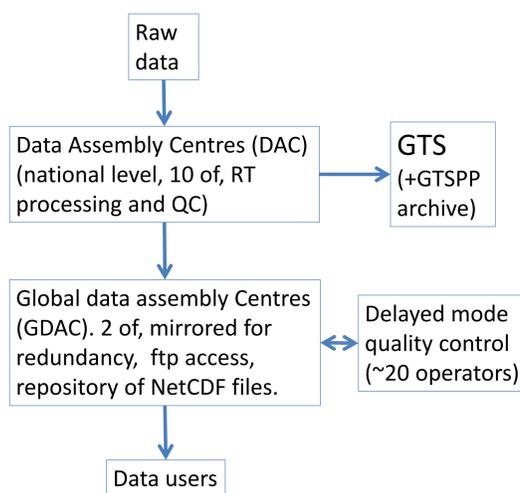


Figure 3: Flow chart showing how Argo data grow and update.

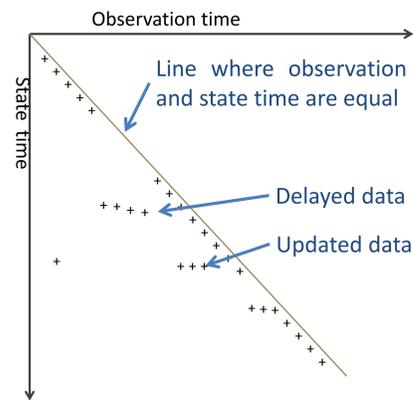


Figure 4: Plot showing bi-temporal growth of dynamic data. Observation time is the time of measurement and state time is the time data are added to the series.

Citing Argo dynamic data

DataCite (Starr *et al.* 2015) provide a number of ways of citing dynamic data.

1. Cite a snapshot which is a full copy of the dataset at a given time.
2. Cite a time slice which is a set of updates/additions to the dataset between two specified times
3. Cite the full dataset with the access time

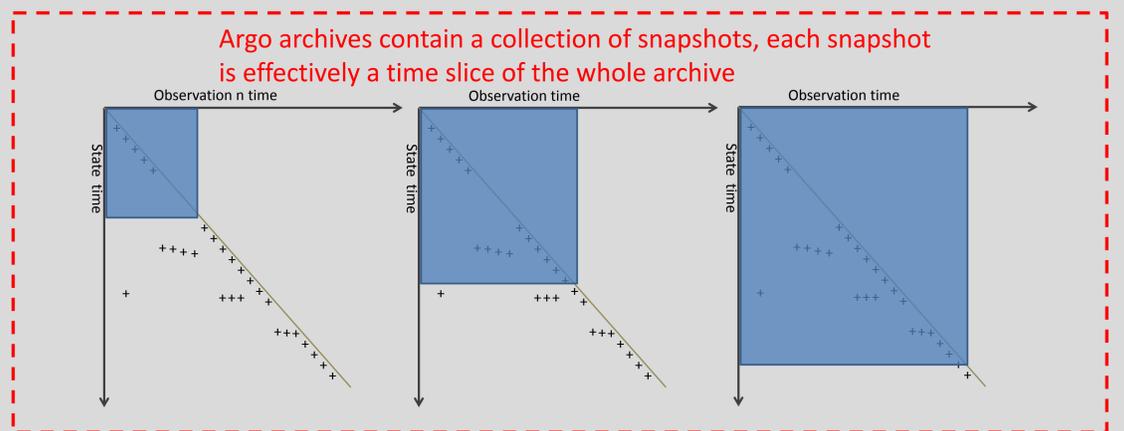
At first sight option 3 is the simplest method but the lack of file versioning at the Argo GDACs means there is no guarantee of data reproducibility.

Snapshots and Argo

Snapshots provide a way to cite Argo data unambiguously. Two archives of Argo data exist:

- Coriolis GDAC offers monthly snapshots of data starting in November 2012.
- NCEI archive number 0042682 creates a snapshot once a week and covers the full history of the Argo project. <http://www.nodc.noaa.gov/archive/arc0022/0042682/>

Figure 5 (right): A visual representation of an archive containing snapshots of a dynamic dataset.



Current citation method

Currently only Ifremer mints Digital Object Identifiers (DOI) for Argo data.

If reproducibility is not required when citing Argo data then there is a DOI for the entire Coriolis GDAC.

<http://dx.doi.org/10.12770/1282383d-9b35-4eaa-a9d6-4b0c24c0cf9>

If reproducibility is required then a snapshot needs to be cited. Each snapshot has its own DOI e.g. the Coriolis GDAC snapshot dated 8th September 2015.

<http://sextant.ifremer.fr/record/ca035889-880d-463e-a523-10aabc3d6be3/>

Moving towards a single DOI

The citation of dynamic data is an active informatics research topic (Watkins *et al.* 2014, Huber *et al.* 2015). Production of an Argo data descriptor paper is on-going. We are working on a prototype approach for reducing the archive of snapshots to a single citable DOI with an additional date in the citation to allow for reproducibility.

The aim is for a citation to have the following format:

[http://dx.doi.org/10.\[ARCHIVE_CENTRE_REF\]/\[Argo_archive_DOI\]/\[time_slice_information\]](http://dx.doi.org/10.[ARCHIVE_CENTRE_REF]/[Argo_archive_DOI]/[time_slice_information])

For more information please see:

<http://www.argodatamgt.org/Access-to-data/Argo-DOI-Digital-Object-Identifier>

And the UCSD page on Argo dataset acknowledgement:

http://www.argo.ucsd.edu/Acknowledging_Argo2.html

References

- Huber, R; Asmi, A.; Buck, J.; de Luca, J.M.; Diepenbroek, D.; Michelini, A. and participants of the joint COOPEUS/ENVR/EUDAT PID workshop (2015). Data citation and digital identifiers for time series data / environmental research infrastructures. Workshop report available at http://figshare.com/articles/Data_citation_and_digital_identifiers_for_time_series_data_environmental_research_infrastructures/1285728
- Starr J. and members of the Metadata Working Group (2015). DataCite Metadata Schema for the Publication and Citation of Research Data. doi:10.5438/0010
- Watkins, J. (2014). Report on RDA Dynamic Data Citation Working Group Workshop. Available from <https://www.rd-alliance.org/system/files/documents/Workshop%20Report%20London%20July%201-2%202014.pdf>

