



A single DOI for Argo:

A generic approach to making datasets that grow and evolve with time citable on legacy data architecture

Justin Buck, National Oceanography Centre (UK), juck@bodc.ac.uk

Thierry Carval, Ifremer (France), thierry.carval@ifremer.fr

Thomas Loubrieu, Ifremer (France), thomas.loubrieu@ifremer.fr

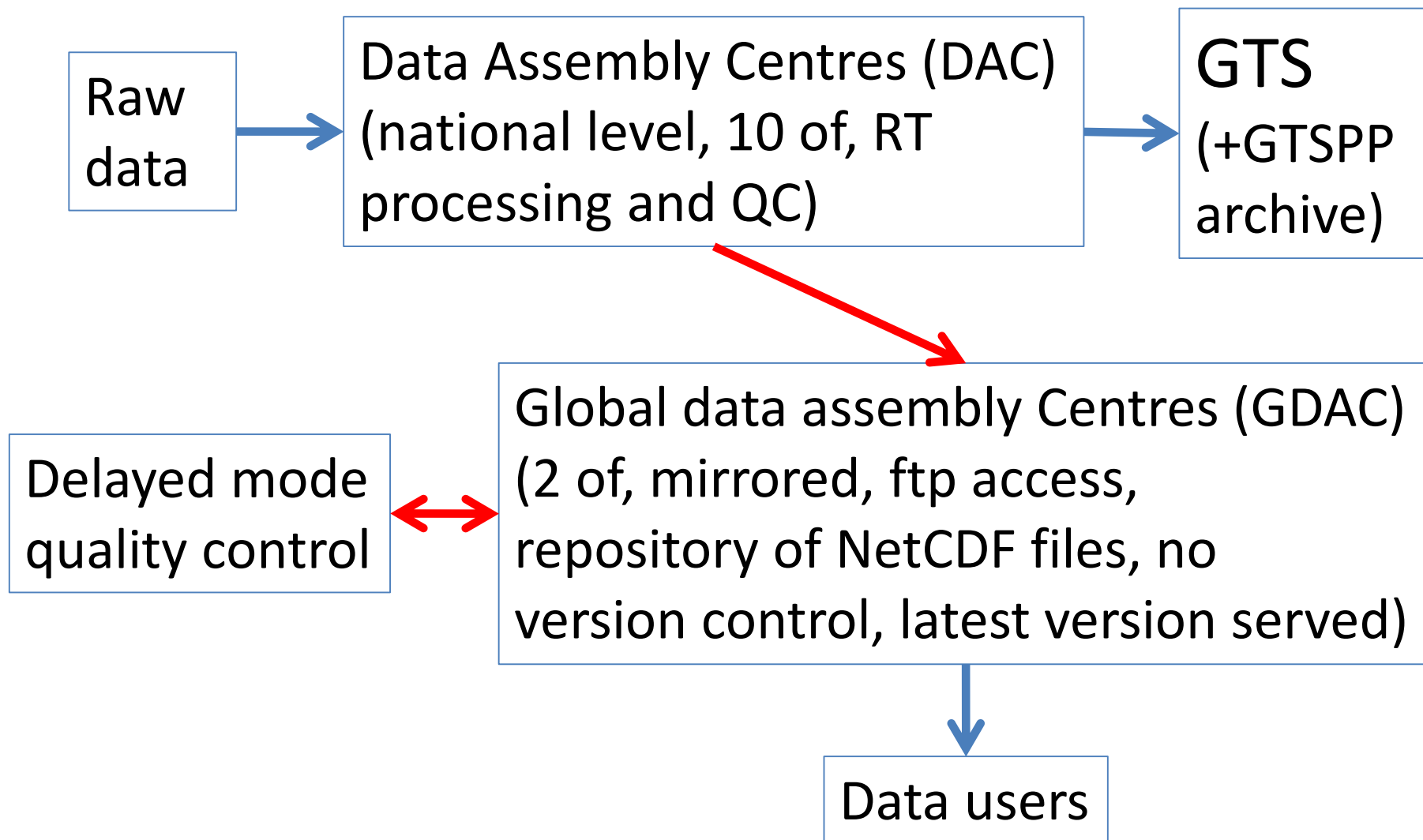
Frederic Merceur, Ifremer (France), frederic.merceur@ifremer.fr



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL



Argo data system (simplified)

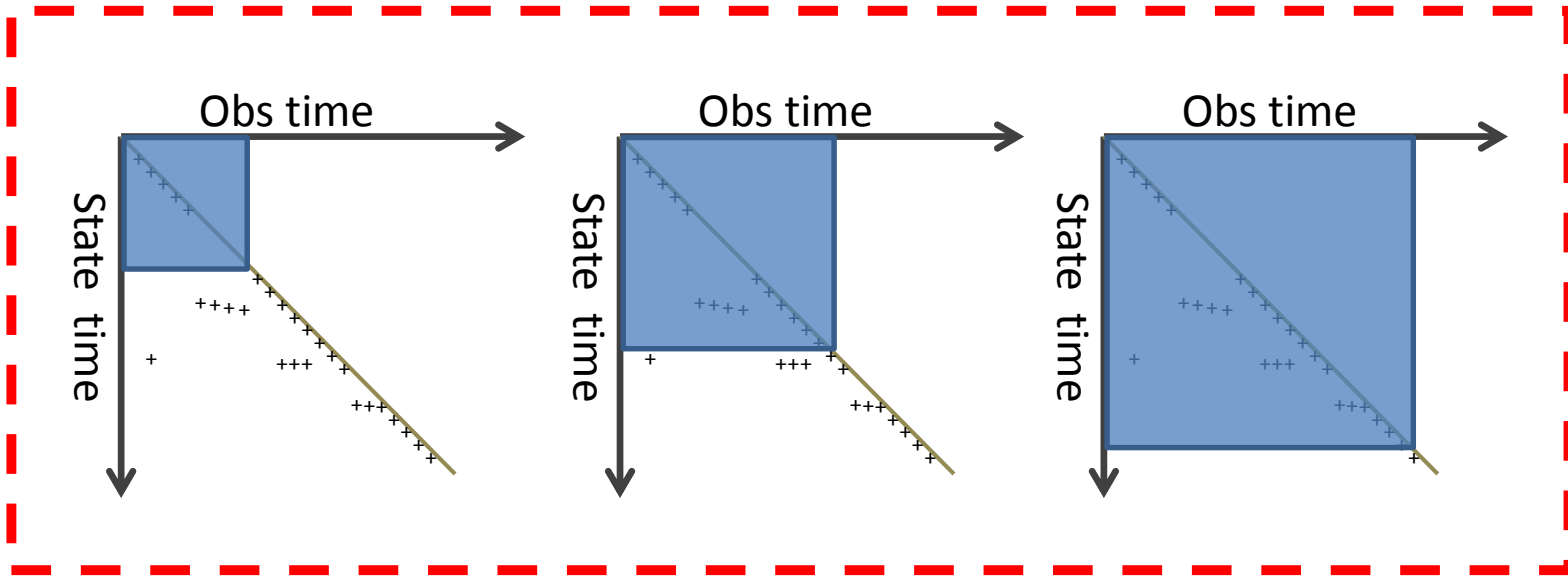


Key associated milestones

- 2014 – Introduction of dynamic data into the DataCite metadata schema
 - <https://schema.datacite.org/>
 - Snapshots were used as an interim solution for Argo
- 2015 – RDA recommendations on evolving data:
 - https://rd-alliance.org/system/files/RDA-DC-Recommendations_151020.pdf
 - Legacy GDAC architecture does not permit full implementation

How to apply a DOI for Argo

Archive (collection of snapshots/granules)



To cite a particular snapshot one can potentially cite a time slice of an archive i.e. the snapshot at a given point in time.

New single DOI

Argo (2000). **Argo float data and metadata from Global Data Assembly Centre (Argo GDAC)**. SEANOE. <http://doi.org/10.17882/42182>



Argo part of the integrated global observation strategy



Argo float data and metadata from Global Data Assembly Centre (Argo GDAC)

Publication date 2000-09-12

Author(s) Argo

DOI [10.17882/42182](http://doi.org/10.17882/42182)

Publisher SEANOE

Keyword(s) float, Argo, global ocean observing system, ocean circulation, in-situ, ocean pressure, sea water salinity, sea water temperature, multi-year, weather climate and seasonal observation, global-ocean, Installations de suivi environnemental

Abstract Argo is a global array of 3,000 free-drifting profiling floats that measures the temperature and salinity of the upper 2000 m of the ocean. This allows, for the first time, continuous monitoring of the temperature, salinity, and velocity of the upper ocean, with all data being relayed and made publicly available within hours after collection. The array provides 100,000 temperature/salinity profiles and velocity measurements per year distributed over the global oceans at an average of 3-

Click to download the data

DATA

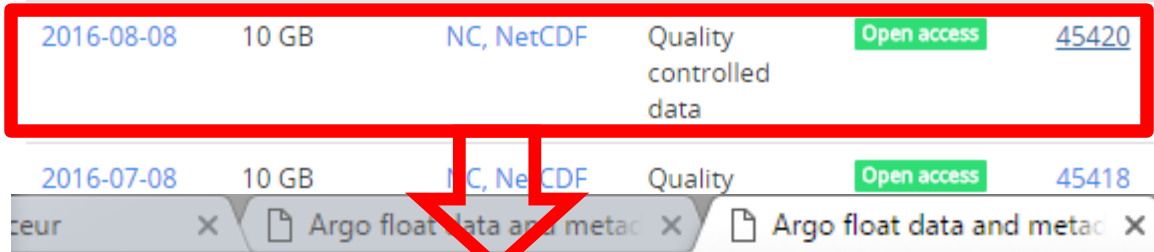


key used to identify snapshot

File	Size	Format	Processing	Access	Key
2016-09-08	10 GB	NC, NetCDF	Quality controlled data	Open access	45724
2016-08-08	10 GB	NC, NetCDF	Quality controlled data	Open access	45420
2016-07-08	10 GB	NC, NetCDF	Quality controlled data	Open access	45418

key used to identify snapshot

File	Size	Format	Processing	Access	Key
2016-09-08	10 GB	NC, NetCDF	Quality controlled data	Open access	45724
2016-08-08	10 GB	NC, NetCDF	Quality controlled data	Open access	45420
2016-07-08	10 GB	NC, NetCDF	Quality	Open access	45418



<http://www.seanoe.org/data/00311/42182/#45420>

Argo part of the integrated global observation strategy

Argo float data and metadata from Global Data Assembly Centre (Argo GDAC) - Snapshot of Argo GDAC of August, 8th 2016

Step towards RDA recommendation

Archive snapshots enables R1 and R2 at monthly granularity

R1 – Data Versioning

R2 – Timestamping

Argo Pilot effectively uses predetermined referencing of snapshots removing the need for requirements R3 to R7. # keys are PIDs for the snapshots and have associated citation texts.

R3 – Query Store Facilities

R4 – Query Uniqueness

R5 – Stable Sorting

R6 – Result Set Verification

R7 – Query Timestamping

R8 – Query PID

R9 – Store Query

R10 – Automated Citation Texts

SEANOE landing page architecture means R11 and R12 effectively met

R11 – Landing Page

R12 – Machine Actionability

Final two requirements untested at this stage

R13 – Technology Migration

R14 – Migration Verification

Summary

- There is now a single DOI for Argo
 - Takes account of legacy GDAC architecture
 - Monthly temporal granularity
 - Enables both reproducible research and simplifies the tracking of citations
 - ‘#’ rather than ‘?’ in identifier takes account of current DOI resolution architecture
- Extensible to other observing systems such as OceanSITES and EGO
- The concept allows for different subsets of Argo data e.g. ocean basins, Bio-Argo data