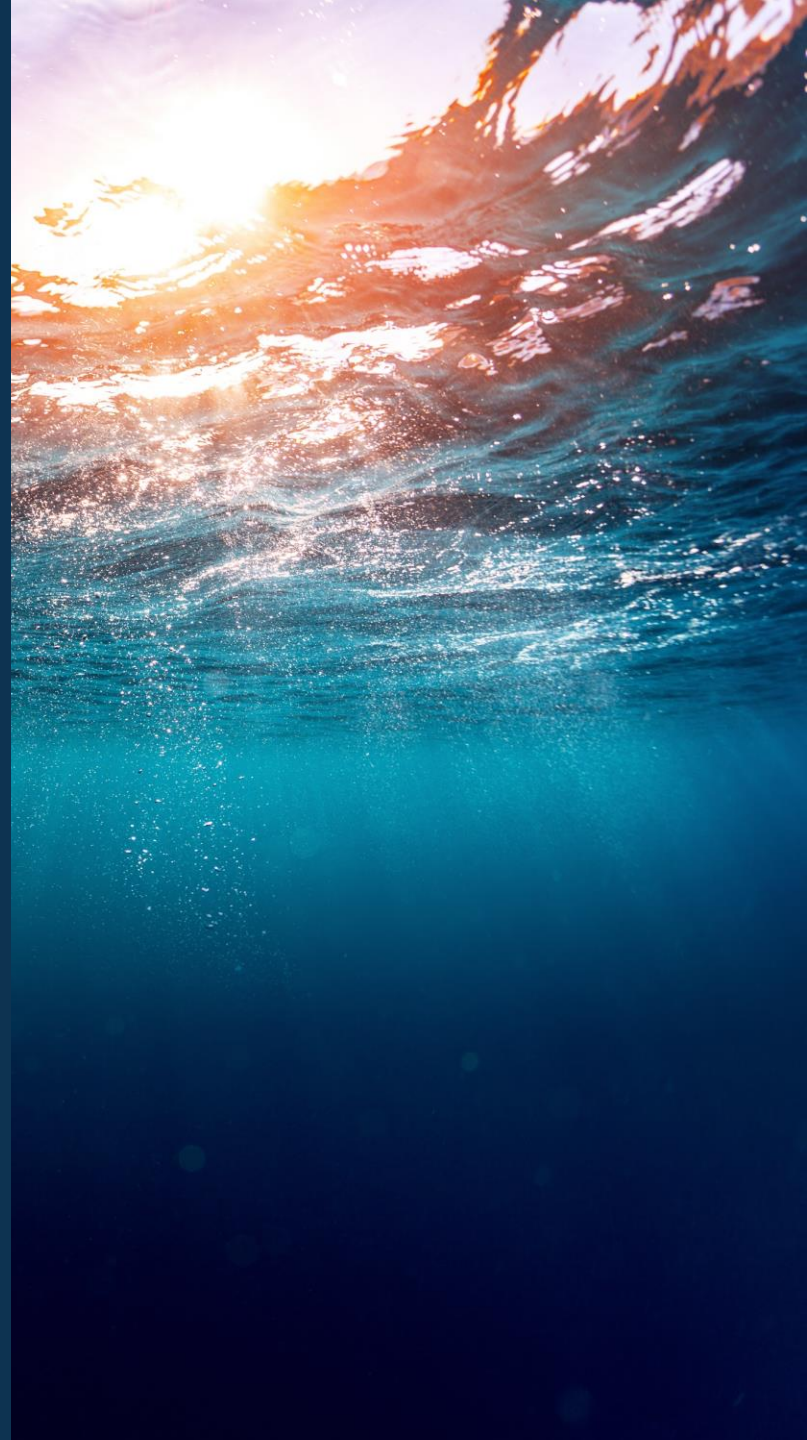
BACK TO BASICS: BRINGING NEW TECHNOLOGIES TO FORGOTTEN SYSTEMS

Monica Hanley

British Oceanographic Data Centre

BACKGROUND

- The British Oceanographic Data Centre (BODC) embarked on a journey to streamline it's workflows for archiving, ingesting, and publishing data in mid 2023
- This was in an effort to ensure robust, and reliable metadata is submitted alongside data, saving valuable data manager time
- Current codebase is convoluted and outdated with single points of failure
- Aim to move to modern technologies that are easier to maintain and knowledge share
- Aim to make workflows easier for both external submitters and data managers



PHASE 1 – DATA SUBMISSIONS

- Captures key metadata required for citations and ingestion with real time validation
- Guided process with simple interface
- Enhances transparency by exposing submitters to controlled vocabularies and discovery metadata
- Review mechanism for data managers to ensure all metadata is correct and captured
- File upload system creates holding area for data

Parameters

Header in File	Description in File	Units in File	Units	Medium	Parameter	Parameter Grouping		Instrument
temp	temperature	C	UPAA	S21S027	TEMPST01	TEMP		TOOL0035
cond	conductivity	S/m	UECA	S21S027	CNDCST01	CNDC		TOOL0035
depth	depth of water column	m	ULAA	S21S027	DEPHPR01	AHGT		TOOL0035

Start Date

04/11/2025 11:44:18

Beginning of temporal coverage of the dataset (e.g. beginning of data collection).

End Date

22/11/2025 11:44:22

End of temporal coverage of the dataset (e.g. end of data collection).

Spatial Coverage

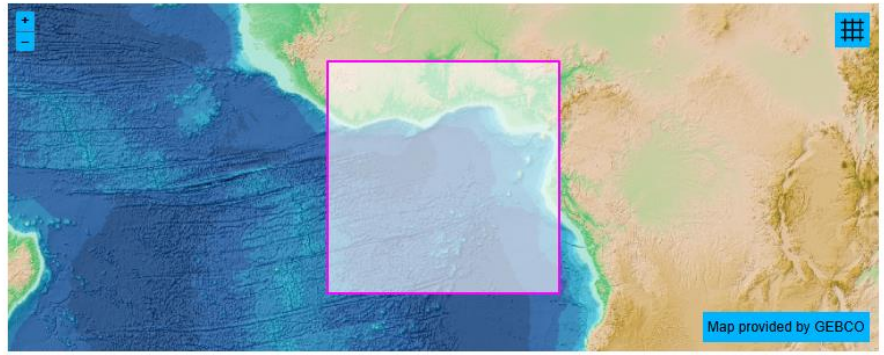
The coordinate values are expressed in decimal degrees with latitude ranging from -90 to 90 and longitude from -180 to 180. These should be provided as WGS84.
Please note values that are west of the prime meridian and south of the equator should be negative.

North: 10

West: -10

East: 10

South: -10



PHASE 1 – ARCHIVE PROCESS

Old process

- Various files emailed to data manager with little metadata – often gets lost
- 20+ year old scripts written in PERL and Matlab to create, open, and close archives
- Archives read only unless opened – must be closed!
- Reliance on crons to process archive jobs
- Upon closing archive, script scans through archive vs. Oracle to assess what's changed and incorporates change in archive monitoring table

New process

- App is a holding area where submitters can revisit their submissions and data managers can review them – saves all files
- Written in Python for easier maintenance
- No need to open archives, creates and adds to them while keeping them read only
- Inserts records straight into Oracle for archive monitoring
- Queue system for archiving with email confirmation when jobs finish

PHASE 2 – DATA CITATION

Old process

- 7 Oracle tables underpin Published Data Library
- Manually fill in tables via SQL
- Manually create DOI archive and copy files in (duplicates files across accession and DOI archives)
- Sign off form using Word doc emailed to originator
- Request publication via SQL command

New process

- Utilises metadata captured on submission and fills in Oracle tables using this information automatically
- DOI archive created by submission – does not duplicate file storage
- Webform to view and edit DOIs
- Sign off form in app where originator can view dummy DOI landing page
- Request publication in webform

PHASE 3 – DATA INGESTION - SAMPLES

Old process

- Sample data submitted in variety of formats with little metadata
- Various Oracle tables populated using metadata from many sources – lots of manual updates
- HTML metadata document written from scratch for each dataset
- Matlab script to populate data tables
- No data visualisation

New process

- Utilises metadata captured on submission and pulls it into ingestion
- Oracle tables populated in one insert by pressing one button in a webform
- Auto-creation of HTML document using metadata from ingestion
- New plots to visualise and flag data
- Ingestion of BTL sensor files

PHASE 3 – DATA INGESTION - SEA LEVEL

Old process

- Various Matlab scripts run to clean data files
- Data loaded into old plotting tool (20+ years old) – often crashed due to size of data
- Tidal statistics calculated and entered into Oracle
- Manual update of HTML files with statistics counted by hand

New process

- Files cleaned on upload
- Parameters auto-mapped to data channels
- New plotting tool to screen and flag data
- Tidal statistics calculated and output into both Oracle and .csv for update of HTML file
- Calculations now in line with the rest of NOC harmonising workflows

PHASE 3 – DATA INGESTION - MOORED SENSORS

Old process

- Data transferred to internal format through Matlab scripts – often need special formatter script
- HTML metadata files written from scratch
- Various Matlab, Linux, and SQL scripts needed to add all metadata to relevant Oracle tables
- QXF screened using old visualisation tool

New process

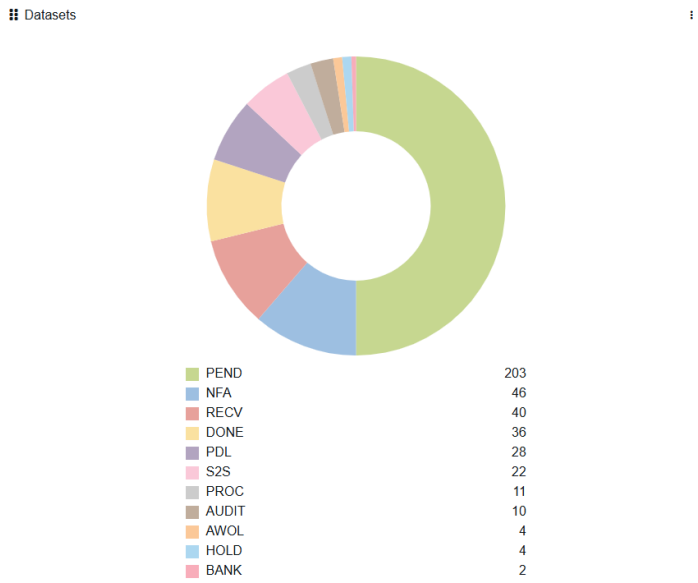
- Utilises metadata captured on submission and pulls it into ingestion
- Oracle tables populated in one insert by pressing one button in a webform
- Auto-creation of HTML document using metadata from ingestion
- New plots to visualise and flag data

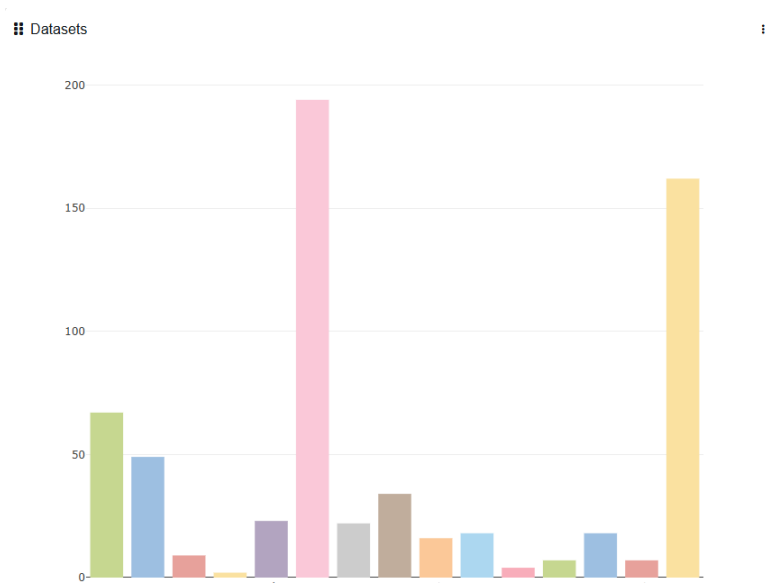
PHASE 4 – JOINING THE DOTS

- Parameter mapping – simple interface utilising NERC NVS Server allowing filtered searches on parameters
- Single interface starts to harmonise different workflows using the same mechanisms repeated in each workflow
- Import dataset list from data management plan
- Automatic updates through data lifecycle of project inventory record
- Dashboards sitting over all new systems to manage work

DOIs

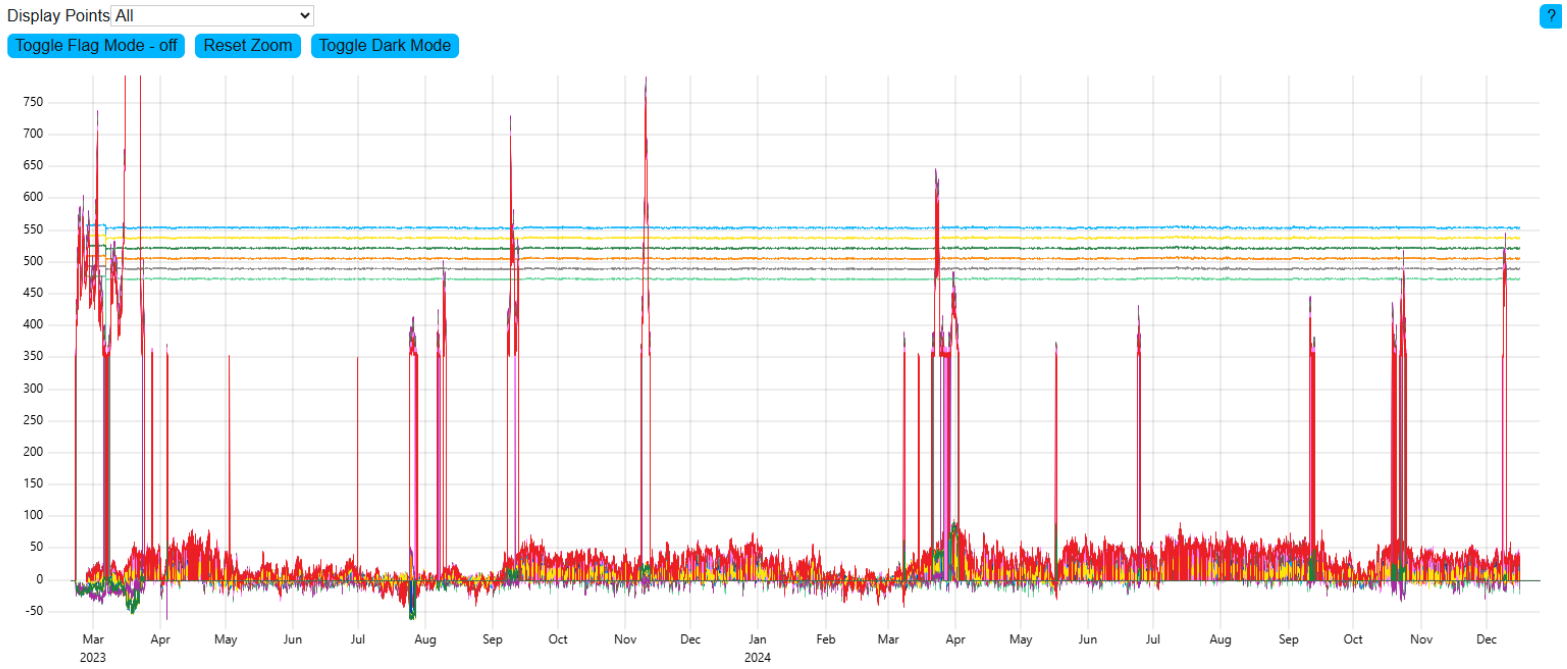
UUID	Title	Creator	Status
357FBCB6E15DE7A3E0637086ABC0F35C	Manx shearwater tracking data (GPS locations, accelerometer data and dive data) from Bardsey and Copeland in 2022 and 2023	MONHAN	creating
2EA9A7F7C47F2BFFE0637086ABC06652	Geochemical and mineralogical analysis of sulphide materials and seawater from oxidative dissolution experiments, from samples obtained from Wetar Island, Indonesia (~2002), from RV Celtic Explorer cruise CE11009 (2011) and from RRS James Cook cruises JC082 (2013) and JC138 (2016)	MONHAN	published
2EE601A2911C822DE0637086ABC0FE93	South Asian Nitrogen Hub Regional Marine Biogeochemistry Model under baseline scenario conditions (1993-2020)	MONHAN	published
2EE5FF1CDA6D81D4E0637086ABC0652A	South Asian Nitrogen Hub Regional Marine Biogeochemistry Model hindcast runs under three nitrogen production scenarios (2004-2020)	MONHAN	published
2E5A50104ABD2BEBE0637086ABC0B159	Highlight seabed images taken by ROV in the Ocean Minerals Company (OMCO) during expedition JC241 in the Clarion-Clipperton Zone (Pacific Ocean, 2023).	MONHAN	published
2DE087C9CEE387F4E0637086ABC0F9A9	Benthic images recorded by a Remotely Operated Vehicle video camera during cruise JC241 in the eastern Clarion-Clipperton Zone (Pacific Ocean, 2023)	MONHAN	published
05F21D1DBF9C5549E0636C86ABC0B846	ATOMIX shear probes benchmark data: a dissipation profile from the Faroe Bank Channel overflow obtained by a vertical microstructure profiler in June 2012.	MONHAN	published
031E7CA19711280DE0636C86ABC014A0	POLCOMS-WAM model run in the Irish Sea generating surface wave data for the CoFEE project for 11 years from 1996.	MONHAN	published
031E7CA19710280DE0636C86ABC014A0	POLCOMS-WAM model run in the Irish Sea generating depth-mean eastward and northward currents for the CoFEE project for 11 years from 1996.	MONHAN	published





IMPACT

- One codebase to maintain for all jobs
- Completely eliminated the need for 3 different programs
- Eliminated the need for multiple Matlab scripts
- Submission through to ingestion and citation time has been reduced by roughly 80%
- Eliminates the need to copy and paste the same information in multiple places



THIS WORK HAS BEEN CHALLENGING AND INNOVATIVE, AND SERVES AS A REMINDER THAT ESTABLISHED SYSTEMS NEED TO KEEP UP WITH MODERN TECHNOLOGIES AND ARE WORTH INVESTING IN!

THANK YOU

FOR ANY QUESTIONS CONTACT MONHAN@NOC.AC.UK

British
Oceanographic
Data Centre