



# DATA MANAGEMENT PLAN

## Project Information

### Project Name

BIOPOLE

### Project URL

<https://biopole.ac.uk/>

### Project Summary

BIOPOLE will address a fundamental aspect of the Earth System – how nutrients in polar waters drive the global carbon cycle and primary productivity. The oceans play a vital role in absorbing atmospheric CO<sub>2</sub>, mitigating large amounts of manmade carbon emissions. However, this part of the global carbon cycle relies on an adequate supply of nutrients to drive the carbon-absorbing marine biological processes. Much of these nutrients are exported from the polar regions. BIOPOLE will improve our ability to quantify this export and identify its sensitivity to climate change.

### Funding

#### Funder

NERC

#### Grant Number

NC-SM2, NE/W004933/1

#### Project Start Date

01/04/2022

#### Project End Date

01/04/2027

## Responsibilities

### Data Centres' Responsibilities

**The UK Polar Data Centre (UK PDC)** together with the PI are responsible for ensuring compliance to FAIR data principles and the NERC Data Policy, <https://www.ukri.org/about-us/nerc/our-policies-and-standards/nerc-data-policy/>, which includes the preservation of valuable observational data, model code and model output. The UK PDC are the lead NERC designated Data Centre and will offer support to the PI for any queries they have regarding the policy and managing their data.

We will also be guided by the UKRI open access policy, [UKRI-28072022-Final UKRI-Open-Access-Policy Version-1.5 July-2022.pdf](#).

Data will be preserved across the above nominated NERC Data Centres, as part of the NERC Environmental Data Service. The Data Centres will hold the data under embargo up to two years from the end of data collection if requested – in line with NERC policy – to allow researchers to work-up their data sets and publish their findings. When released, data will be published under the UK Open Government License, <https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>, unless agreed otherwise.

In order to keep all data generated by the BIOPOLE project as one data collection, the nominated Data Centres will:

1. Include the keyword 'BIOPOLE' in the abstract section of each data catalogue record
2. Create cross-reference links between relevant related data records, to ensure the whole collection can be discovered even if individual datasets are preserved across several Data Centres.

The UK PDC will make data available via the UK PDC Discovery Metadata System, <https://data.bas.ac.uk/>, and can generate a Digital Object Identifiers (DOI) for datasets of long-term value.

**The Environmental Information Data Centre (EIDC)** will work with UK Polar Data Centre (UK PDC) and the PI to identify data generated in the programme that fall within the terrestrial and freshwater science remit and agree a long-term plan for each that meets NERC data policy and [FAIR principles](#).

The EIDC will provide liaison guidance regarding data collection, within programme data dissemination between partners, requirements for long term data stewardship, dissemination to the wider scientific community and public, data access statements for publications and mechanisms for citation of datasets that promote feedback on impact.

The EIDC will provide named contacts to guide researchers through ingestion of data to the EIDC, to compliment the following web-accessible guidance:

Depositing data <https://eidc.ac.uk/deposit>

Acceptable formats <https://eidc.ac.uk/deposit/suitableFormats>

Metadata and supporting documentation  
<https://eidc.ac.uk/deposit/supportingDocumentation>

The EIDC will issue a DataCite Digital Object Identifier (DOI) for all data accepted, which can be used as a citation reference

The EIDC makes all data held findable and discoverable through records in its data catalogue <https://catalogue.ceh.ac.uk/eidc/documents>, the UK Govt data portal [www.data.gov.uk](http://www.data.gov.uk) and web-search tools e.g. Google DataSearch.

**The British Oceanographic Data Centre (BODC)** are responsible for defining standards for metadata and data supply, management of received data, guidance and ongoing liaison regarding data collection and processing, data dissemination between project partners, long term data stewardship and dissemination to wider scientific community and public on completion of project.

The following data deposit conditions apply when submitting data to BODC  
[https://www.bodc.ac.uk/submit\\_data/submission\\_guidelines/deposit\\_conditions/](https://www.bodc.ac.uk/submit_data/submission_guidelines/deposit_conditions/).

The BODC will make data available via the BODC All Data Series ([https://www.bodc.ac.uk/data/bodc\\_database/nodb/](https://www.bodc.ac.uk/data/bodc_database/nodb/)) and can publish Digital Object Identifiers (DOI) if requested which will be available via the BODC Published Data Library ([https://www.bodc.ac.uk/data/published\\_data\\_library/](https://www.bodc.ac.uk/data/published_data_library/)). In order to publish a DOI at BODC the data are required to be open access.

BODC are also the UK Data Archiving Centre (DAC) for glider data and are responsible for providing glider data in Near-Real-Time to the project and wider community.

## Contributors' Responsibilities

The BIOPOLE PI is responsible for ensuring that the data management plan is followed and that all data and accompanying metadata are submitted to the designated Data Centre by the end of the project for long-term curation.

Research data should be deposited to the NERC Data Centres soon after completion (requesting an embargo on access where required) rather than waiting until the end of the project. This ensures data and supporting documentation are archived in good time with a DOI issued that can be used to both cite the data as well as include in the required data access statement in any planned papers. Some journals are now requiring DOIs to be issued to datasets before papers are published.

All data, models and code generated during the research should meet the FAIR principles as far as possible to maximise the potential for collaborative research during and beyond the programme of work. Research teams should plan to generate and manage data, models and code so that they become Findable, Accessible, Interoperable and Re-usable.

BIOPOLE's data management activities should make maximum use of existing international standards and international collaborations providing these do not run counter to other elements of the NERC Data Policy.

NB: the decision whether a dataset, model or code are to be made 'openly available' or restricted due to legitimate reasons is a separate consideration. Resources should be made FAIR in order to maximise any re-use regardless of the level of openness.

All BIOPOLE data creators are encouraged to support sharing of data within the BIOPOLE community and are required to keep any 'embargo' period as short as possible. The original producer and owner(s), as set out in the BIOPOLE Data Management Plan, will have 'first right of use' in publishing research papers.

## Data Re-use

### Third-party Data and Models

These are data of interest to BIOPOLE generated by external collaborators, not produced by the BIOPOLE contributors. Third-party data may be a subject to a different IPR and licence and therefore both the IPR and licence issues will be specified for each third-party dataset individually in the BIOPOLE DMP table at the end of the document.

It is expected that BIOPOLE will use third-party data, for example for distribution data compilations. However, these data will come from open access resources, such as the Ocean Biodiversity Information System (OBIS) and the ArcticGRO programme.

If the third-party data are being used and reprocessed into a new dataset then the new dataset is treated as a BIOPOLE dataset and bears references to the original third-party data.

Third-party data and models identified so far are listed in table of new datasets at the end of the document.

If other third-party resources are identified or information about their re-analysis becomes available, the DMP will be updated accordingly.

## Data Collection

### Project datasets

A detailed list of the BIOPOLE generated datasets will be available from the *BIOPOLE Monitoring Tracker*.

The BIOPOLE Work Package leaders are responsible for ensuring that the Monitoring Tracker is up to date.

The BIOPOLE Data Manager is responsible for ensuring the Data Management Plan document is up to date with the Project Monitoring Tracker.

The DMP is a stand-alone document to be shared with the funder, project members and wider. Therefore, the DMP cannot link to an in-project internal tool but will be reviewed annually and updated based on information in the Tracker.

### Observational and Experiment Data:

The Work Package 1, 2 and 3 will generate biogeochemical, molecular and biodiversity data as well as model outputs as outlined in table of new datasets.

Parameters that will be measured are detailed in the BIOPOLE COOKBOOK, available from the BIOPOLE website, and include carbon and macronutrients, iron and other trace metals, tracers (oxygen isotopes, rare earth elements) and biological parameters. The sampling will cover a range of environments (rivers, fjords, field banks and ocean).

### Measurements and Instruments:

The main field seasons are planned for 2023, 2024 and 2025 although deployment of some moored marine instruments will take place in late 2022. Data processing will take place once datasets have been received from this fieldwork, with the majority of processing occurring in 2024, 2025 and 2026. We anticipate the transfer of processed field datasets will take place in the 2026 and 2027.

Measurements will be conducted by scientists from the British Antarctic Survey (BAS), National Oceanography Centre (NOC) and UK Centre for Ecology & Hydrology (UKCEH), British Geological Survey (BGS) and Centre for Polar Observation and Modelling (CPOM) using community-accepted standard operating procedures and protocols.

Freshwater data will be produced by laboratories at NOC, UKCEH, BGS, and BAS, with coordination being provided by BAS. Land-based field teams will collect samples and track them from field sample location to the analytical laboratory using a sample tracker spreadsheet. Data from the field and sample tracking will be coordinated by UKCEH.

Marine data will be collected using specialised oceanographic instrumentation on a variety of deployment platforms (RRS Sir David Attenborough or NERC vessels plus vessels of opportunity, moorings, and autonomous vehicles such as gliders and sailbuoys). This will include use of third party infrastructure e.g. the Davis Strait mooring where BIOPOLE water samplers will be installed.

## Observational Data Formats and Volume

All data suitable for long-term preservation will be made available in a non-proprietary format, following a guidance on open data formats, available e.g. at <https://www.bas.ac.uk/pdc-data-format-guidance>.

The anticipated data formats will be: Comma-Separated Values (CSV), NetCDF (NC) and plain text (TXT)

The anticipated total volume of observational data of each work package is expected not to exceed 0.5 TB.

## Time Scale

Detailed structure and management of the BIOPOLE programme is described in the *BIOPOLE Management Plan* document (available in the project OneDrive collaboration area).

Data capture will start in autumn 2022. Data processing will begin as soon as data is received. Delivery date for each dataset to the nominated NERC EDS is indicated in table of new datasets.

For WP3 data collection will begin with the deployment of water samplers on the Davis Strait mooring in September 2022. Water samplers and samples are due to be recovered in autumn 2024. It is hoped samples will be analysed and QC'd by spring 2025.

## Model Data

### Model Code

Model codes will be made available in an open source environment, governed by a source code management tool such as GIT. Where source code is not able to be shared because of licensing constraints it will be essential that documentation is available that describes what software systems and what version of those were used to create the data.

### Model Data Formats and Volumes

Model configuration files and input or forcing data will be preserved in standard formats (e.g. netCDF) in a versioned repository.

Due to high volume of model data, NERC Data Centres will only archive model data used in publications (so it can be used as a point of reference) or data that is perceived to have a high value to the wider scientific community (outside of the specific project).

An estimate of the total model data volume suitable for archiving shall be made based on the subsetting that does not affect underlying scientific principles. Options for data subsetting include climate averaging, reducing resolution, subsetting by variables, restricting spatial domain, reducing ensemble statistics or restricting to model realm.

The anticipated data formats will be: NetCDF (NC), Comma-Separated Values (CSV) or ASCII (e.g. in the copepod slab model).

The anticipated total volume of model data will not exceed 1 TB per a dataset in the majority cases. However, for very large datasets (over 1TB) the data would be stored in the CEDA Archive and linked to the metadata catalogue record in the designated Data Centre, as detailed in the table of new datasets.

## Time Scale

Detailed structure and management of the BIOPOLE programme is described in the *BIOPOLE Management Plan* document.

WP3 will be creating model output throughout the project, most actively between autumn 2022 and autumn 2026. A concise summary of the overall timescale of modelling data is in the table of new datasets.

## Data Quality

### Quality Assurance by Data Creators

It is expected that data being supplied by the PIs will have been suitably quality checked. All data analysis will be carried out using nationally or internationally recognised methodologies and international reference standard materials, where possible. The researcher will ensure that the methodology is documented, and monitor and maintain procedural and analytical reproducibility using replicates and standards where possible. Any corrections or processing of data must be documented in the dataset methodology.

Where appropriate, the CF standard names will be used in the NetCDF files, <https://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html>.

Data variable units will be specified for each measurable variable and units common for a specific type of measurement will be used. The Unified Code for Units of Measure will be consulted, <https://ucum.org/ucum.html#section-Grammar-of-Units-and-Unit-Terms>.

Missing values will be distinguished from true zero values, e.g. using the 'NaN' designation.

Where applicable, relevant terminologies will be used. For instance, marine taxonomic species identifications will be verified against the World Register of Marine Species, <https://www.marinespecies.org/>.

Other aspects of data quality assurance, such as validation of data integrity, handling of sample replicates, usage of community-specific standards or comments on data correction will be specific to each dataset and described as part of each dataset methodology.

All instruments will be calibrated by the manufacturers prior to deployment. Repeat sampling and instrument to instrument comparison will be used to reveal systematic errors. Each dataset will be quality controlled and suspect data points will be flagged.

### Quality Assurance by Data Centres

The above nominated Data Centres will perform their own quality assurance checks on the data and metadata to ensure compliance with the FAIR data principles.

## Metadata Collection

In general, the necessary metadata to enable data use and re-use will be recorded to the best of the scientist's abilities. The metadata should include details about how (with which instrument or technique), when and where the data have been collected, by whom (including the affiliation and

contact address or telephone number) and in the framework of which research project (BIOPOLE in this case). All involved Data Centres (BODC, EIDC and UK PDC) require the same essential discovery metadata and their metadata guidance is available from the links below.

For all submitted data, the data custodians need to know how the values were arrived at. The derivation process must be stated: all processing and calibration steps should be described and calibration values supplied, where applicable. The nature and units of the recorded variables are essential, as well as the grid or reference system. As much information as possible about fieldwork instrumentation should be included (e.g. serial number, copies of manufacturer's calibration sheets, and recent calibrations).

The PDC metadata guidance and metadata template (to be used when submitting data) is available at <https://www.bas.ac.uk/data/uk-pdc/metadata-guidance/>. UK PDC uses the DataCite Metadata Schema to issue DOIs for datasets.

The EIDC metadata guidance is available at <https://eidc.ac.uk/deposit/supportingDocumentation>

The BODC metadata guidance: Metadata and documentation will be supplied to the standards specified by BODC, please see the following link for a general guide

[https://www.bodc.ac.uk/submit\\_data/submission\\_guidelines/general\\_guide/](https://www.bodc.ac.uk/submit_data/submission_guidelines/general_guide/).

BODC advise to use the following metadata template

[https://www.bodc.ac.uk/submit\\_data/submission\\_templates/](https://www.bodc.ac.uk/submit_data/submission_templates/) to provide appropriate metadata accompanying a dataset.

Glider specific metadata guidance is as follows:

Metadata and documents for the glider deployments will be supplied to BODC at least 2-3 weeks prior to deployment to ensure the NRT data can be processed including:

- Names and serial numbers of the gliders deployed
- Dates and location of deployments
- Length of deployment
- Sensors (including serial numbers) that will be deployed on the glider
- Calibration sheets
- Trim sheets
- Example file from a test run.

BODC will create project and campaign records for the deployments, which includes project aims, participants, organisations, keywords and funders. These records ensure metadata are curated in a standardised way.

Metadata for NetCDF files should comply with the Climate Forecasting (CF) convention as far as possible. Global attributes shall be included, as described in the Attribute Convention for Data Discovery (ACDD), [https://wiki.esipfed.org/Category:Attribute\\_Conventions\\_Dataset\\_Discovery](https://wiki.esipfed.org/Category:Attribute_Conventions_Dataset_Discovery).

Metadata for model data should include what model, input data, and any pre- or post-processing software was used along with version information. It includes the date when the model output data were created and the persons and institutes responsible for running the model.

Metadata and any associated documentation will be submitted at the same time as the accompanying datasets.

Other metadata and/or documentation specific to this project, such as laboratory notebooks, field reports, equipment information or methodology reports will be available from the scientist responsible for each dataset.

The following documentation will be used:

BIPOLE\_COOKBOOK, GEOTRACERS\_COOKBOOK (<https://www.geotraces.org/methods-cookbook/>), PARAMETERS\_LIST, GO-SHIP\_RepeatHydrographyNutrientManual (<https://www.go-ship.org/HydroMan.html>), Nansen Legacy (<https://septentrio.uit.no/index.php/nansenlegacy/article/view/5793>).

Marine sampling will be described in the full cruise reports and submitted to the BODC following their guidelines.

Freshwater and terrestrial sampling will be described in field reports and stored locally on the UK CEH network being shared across the project through a dedicated OneDrive folder.

## Sample Collection

WP1 – Field sampling campaigns will be conducted to produce sediment and water samples across the four process stations at Ny-Alesund, the Tana River, Rothera and King Edward Point Research stations. Samples will be processed and temporarily stored in the laboratories of the field stations before being shipped to receiving analytical laboratories in the UK. A sample coding and tracking system will be implemented.

WP2 – Zooplankton, copepods and benthic species will be collected using a variety of towed instruments including pelagic nets such as Bongo nets and MOCNESS nets. Benthic samples will be collected by devices such as epibenthic sledges and Agassiz trawls. Phytoplankton and particulate material will be collected by Niskin bottles mounted on CTD frames. Some samples will be fixed and preserved in formaldehyde, ethanol or Lugols solution. Other samples will be frozen at either -20°C or -80°C. All samples will be given a unique label containing at least the cruise number, event number, collection device and depth of sample and bottle number for water samples. Sediment samples will be collected by a sediment trap and preserved in formaldehyde. Samples will be identified by mooring location, date of deployment and retrieval and sample tube number representing the period of sampling during the deployment.

WP3 – Existing data from the Davis Strait moorings provided by the BIOPOLE project partners are listed in the table of third-party data and their use will respect the partners' institutional obligations. Data from the Davis Strait field samples analysed by BIOPOLE will be preserved as listed in the table of physical samples below.

## In-project Data Management

### Data Storage

Early access and sharing data within the BIOPOLE programme members is crucial to deliver the project.

The BIOPOLE contributors are the programme members from:

- British Antarctic Survey (BAS)
- National Oceanography Centre (NOC)
- UK Centre for Ecology and Hydrology (UKCEH)
- British Geological Survey (BGS)
- Centre for Polar Observation and Modelling (CPOM)

The BIOPOLE contributors hold the IPR of BIOPOLE generate data, which are subject to the OGL, V3, as specified in the NERC Data Policy. Datasets created by BIOPOLE contributors are listed in the BIOPOLE DMP table of new datasets that specifies for each new dataset, when and by whom will each dataset be published. All BIOPOLE-generated data are freely available to all contributors.

BIOPOLE data can be used before they are published but the data originator must be informed and should be offered collaboration on the scientific analysis and offered co-authorship. The data originator has the right not to share the data if these were incomplete or not yet quality checked and this could lead to misinterpretation of the data.

In-project data storage mechanisms don't aim to change established practices but have to make sure that data are safely preserved. All project partners should store data they have generated securely in their institutional data storage system following the following best practice:

- a) for each dataset there is a systematic folder structure of 'original', 'processing' and 'final' directory and a 'metadata' document stored with the dataset
- b) access to each dataset is restricted to a defined list of users with an institutional account
- c) raw data are read only
- d) processing data have alongside them info on what cleaning/processing has been performed
- e) metadata document is as rich as possible and has at the minimum discovery metadata (where, when, who, how and why).
- f) large volume processing data can be stored on the JASMIN infrastructure using dedicated group workspaces. (Large-volume final data over 1TB associated with an article publication can be stored for long-term keep at the CEDA Archive.)
- g) Effort will be made to keep folder names unique to enable sharing between storage platforms. File names will as much as possible adhere to general recommendations, i.e. be consistent, enable easy sorting, without spaces and special characters.

Offsite backups and institutional IT security standards shall mitigate risks associated with data corruption due to malicious actors.

If required, a dedicated BIOPOLE volume can be created on the BAS Storage Area Network (SAN) where each dataset has the directories: 'original', 'processing', 'final' and a 'metadata' document. This SAN volume can be operational until all relevant high-value data have been permanently archived/published in a NERC-approved trusted repository as specified in the BIOPOLE DMP. Data can be delivered to the BAS SAN using a login-protected BAS FTP and distributed via a login-protected BAS webserver. Access would be limited to a defined list of users (BIOPOLE contributors).

## Data Sharing

Documents will be shared via a dedicated BIOPOLE OneDrive. There is 1TB storage limit and BIOPOLE contributors from BAS, BGS, CPOM, NOC and UKCEH should be able to access files from this dedicated drive.

It is expected that scientific datasets will not exceed the storage limit and will also be shared via the BIOPOLE OneDrive. If this turns out not to be sufficient, data can be shared using the above mentioned BIOPOLE collaboration volume on the BAS SAN.

Large files on JASMIN can be accessed by providing access to the relevant group workspaces managed by their administrator.

A version controlled repository, such as GitHub, will be used for sharing of a code which is in progress. A permanent repository, such as Zenodo, maybe used to publish the code final version. The final version of the code should have mutual links to the final version of the dataset.

## Data Availability and Re-use

### Data Ownership and Licence

NERC is the funder of the data generated by BIOPOLE and follows the UK Open Government Licence, V3.0, for data re-use. Intellectual Property Rights follow the NERC Data Policy, as described at <https://www.ukri.org/wp-content/uploads/2022/03/NERC-080322-PolicyGuidance-data-021219.pdf>.

Any other data licence or IPR to follow will be reviewed as the project proceeds and data sets are identified.

### Data Access

NERC Data Policy allows embargo on a dataset for up to two years from the end of data collection.

Data embargo will be applied to the datasets as specified in the table of new datasets below.

All research publications arising from NERC funding will follow the [UKRI open access policy](#) and will include a statement on how the supporting data and any other relevant research materials can be accessed. This should include formal data citation of the dataset in question. The data citation shall be included in the publication's list of references and a permanent unique DOI identifier assigned to the dataset.

### Data Preservation for long-term

Evaluation of data with potential value for long-term preservation considers the cost of preservation (including the cost of ensuring the data are usable), the likelihood and feasibility of reuse by a third party and the cost of re-creating the data.

Details on long-term preservation of individual datasets created by the project are provided in the table below.

## New Datasets Overview

### Digital Datasets

\*A brief description of the activities that will produce the data, observational or models.

Dataset Description*	Contact	Data Volume	Data Format	Issues	Delivery Date	Embargo Date	Preservation
<i>Dataset description</i>	<i>Dataset contact</i>			<i>Any issues with the data, .e.g. legal, access, ethics etc</i>	<i>Date expect to receive data</i>	<i>No more than 2 years after collection</i>	<i>e.g. Keep indefinitely, Do not keep etc, including destination Data Centre</i>
Coordinated land and sea based field pilot campaign in Loch Etive/West Coast Scotland	<b>WP1</b>	<1 Gb	TBD		Dec 2023	No more than 2 years after collection	To be kept indefinitely, EIDC
Coordinated land and sea based field campaign in Ny Alesund, Tana River, Rothera and KEP	<b>WP1</b>	<1 Gb	TBD		Dec 2025	No more than 2 years after collection	To be kept indefinitely, PDC
Bio-optics data existing + new (South Georgia ridge/seamounts)	<b>WP1</b>	<1 TB	TBD		Dec 2026	No more than 2 years after collection	To be kept indefinitely, PDC
Quality controlled dataset of T,S and bio-optics from field experiment in Scotia Sea	<b>WP1</b>	< 1 TB	TBD		Apr 2025	No more than 2 years after collection	To be kept indefinitely, PDC
Physics/bio-optics contrasts around the marginal ice zone from glider work	<b>WP1</b>	< 1 TB	TBD		Apr 2026	No more than 2 years after collection	To be kept indefinitely, BODC
Synthesis of existing shelf-basin/WAP glider data (AB)	<b>WP1</b>	<1 GB	TBD		Dec 2023	No more than 2 years after collection	To be kept indefinitely, BODC
3x Near real time glider deployments (includes T&S data)	<b>WP1</b>	<5 GB	.tbd/.sbd		During the glider deployment	None	To be kept indefinitely, BODC

3x Recovery glider deployments (i.e. the data from the glider SD card when the glider is recovered)	<b>WP1</b>	<5 GB	.ebd/.dbd		Sep 2024	Up to 2 years after glider recovery	To be kept indefinitely, BODC
3x Delayed mode glider deployments - collated and quality-controlled data from each glider deployment	<b>WP1</b>	<5 GB	netCDF		Mar 2025	Up to 2 years after glider recovery	To be kept indefinitely, BODC
South Georgia and South Orkneys 1992-2012 - lagrangian modelling - transport processes	<b>WP1</b>	<1 TB	TBD		Dec 2026	No more than 2 years after collection	To be kept indefinitely, PDC
Datasets (variety of spatial scales) on icebergs, ice sheets, sea ice runoff/freshwater fluxes (not all widely available)	<b>WP1</b>	<1 GB	TBD		Dec 2024	No more than 2 years after collection	To be kept indefinitely, PDC
Data at relevant scales on estimates of large-scale meltwater fluxes (approx. annual)	<b>WP1</b>	< 1 GB	TBD		Dec 2024	No more than 2 years after collection	To be kept indefinitely, PDC
Vertical profiles of N:P in particulate matter and particle size spectra: Data on N:P of particulates in Chukchi Sea	<b>WP2</b>	100 MB	.csv		Jul 2024	Jul 2026	To be kept indefinitely; BODC
Isotope pairing [pelagic denitrification] experiments: pelagic denitrification dataset	<b>WP2</b>	TBD	TBD		Jul 2024	Jul 2026	To be kept indefinitely; BODC
Database of culture experiments examining the body size changes, cellular PON and POP, primary production and respiration of eukaryotic plankton organisms at different temperature and N:P conditions	<b>WP2</b>	100MB	.csv		Dec 2024	Jul 2026	To be kept indefinitely; BODC
Gene sequencing analysis: Molecular dataset from Chukchi Sea samples	<b>WP2</b>	10 GB	TBD		Jul 2024	Jul 2026	To be kept indefinitely. uploaded to an online genomics repository with links retained

							by PDC and BODC
Database of distribution and abundance of lipid storing/diapausing mesozooplankton in Southern Ocean	<b>WP2</b>	200 MB	.csv		Jun 2025	Jun 2027	To be kept indefinitely; PDC
Database of distribution and abundance of benthic consumers of lipid rich diapausing zooplankton in Southern Ocean	<b>WP2</b>	200 MB	.csv		Jun 2025	Jun 2027	To be kept indefinitely; PDC
Database of depth-resolved copepod lipid content and lipid composition from literature, BIOPOLE analyses of partner cruises and BIOPOLE cruises	<b>WP2</b>	300 MB	.csv		Mar 2026	Mar 2028	To be kept indefinitely, PDC
Database of ETS activity in lipid storing mesozooplankton vs depth	<b>WP2</b>	300 MB	.csv		Mar 2026	Mar 2028	To be kept indefinitely, PDC
Copepod lipid sac measurements from BIOPOLE and partner cruises	<b>WP2</b>	100 MB	.csv		Jun 2025	Jun 2027	To be kept indefinitely, PDC
Copepod respiration rates from BIOPOLE and other cruises of opportunity	<b>WP2</b>	200 MB	.csv		Jun 2025	Jun 2027	To be kept indefinitely, PDC
Lipid pump model for polar calanoid copepods	<b>WP3</b>	1 GB	.csv		Nov 2024	Nov 2025	To be kept indefinitely, PDC
Wider Arctic coastal permafrost erosion model	<b>WP3</b>	1 TB	netcdf		Apr 2024		To be kept indefinitely, PDC
Processed Lagrangian trajectories for nutrient pathways and lipid pump	<b>WP3</b>	5 TB	netcdf		Dec 2025		To be kept indefinitely, PDC + CEDA Archive
Idealised model control runs (1o & 1/4o)	<b>WP3</b>	250 GB	netcdf		Oct 2023		To be kept indefinitely, BODC
Idealised model perturbation experiments (1o & 1/4o)	<b>WP3</b>	1 TB	netcdf		Oct 2024		To be kept indefinitely, BODC
Global DBPM simulations driven by UKESM1 to 2300	<b>WP3</b>	<1TB	TBD		Apr 2025		To be kept indefinitely, PDC

Global FISHMIP simulations (intention is to use a FISHMIP emulator) driven by UKESM for contrasting polar nutrient export scenarios	<b>WP3</b>	<1TB	TBD		Apr 2025		To be kept indefinitely, PDC
(Possible) use of coupled idealised-ecological model to explore influence of SO nutrient exports at Atlantic scale.	<b>WP3</b>	<1TB	TBD		Apr 2025		To be kept indefinitely, PDC
Adjoint sensitivities from the ECCOv4 setup of MITgcm	<b>WP3</b>	1 TB	netcdf		Sep 2025		To be kept indefinitely, PDC

## Hardcopy Records

*\*A brief description of the activities that will produce the records*

Dataset Description*	Contact	Data Volume	Data Format	Issues	Delivery Date	Preservation
<i>Dataset description</i>	<i>Dataset contact</i>			<i>Any issues with the data, .e.g. legal, access, ethics etc</i>	<i>Date expect to receive data</i>	<i>e.g. Keep indefinitely, Do not keep etc, including destination Data Centre</i>
Land based field sampling logs	<b>WP1</b>	<100 MB	Paper		Project end	Archived UKCEH, indefinitely

## Physical Samples

*\*A brief description of the activities that will produce the sample data*

Dataset Description*	Contact	Data Volume	Data Format	Issues	Delivery Date	Embargo Date	Preservation
<i>Dataset description</i>	<i>Dataset contact</i>			<i>Any issues with the data, .e.g. legal, access, ethics etc</i>	<i>Date expect to receive data</i>		<i>e.g. Keep indefinitely, Do not keep etc, including destination Data Centre</i>
River sediment samples from KEP, Rothera, Tana and Ny-AI	<b>WP1</b>	< 100 MB	.csv		2026		Archive UKCEH indefinitely

Particulate matter from Chukchi Sea cruises	<b>WP2</b>	100 MB	.csv		Sampling is planned for 2022 or 2023		Data resulting from analyses to be kept indefinitely; BODC
Zooplankton samples from BIOPOLE cruises and partner cruises	<b>WP2</b>	200 MB	.csv		Sampling is planned for 2022 or 2023		Data resulting from analyses to be kept indefinitely; PDC
Benthic samples from BIOPOLE cruises and partner cruises	<b>WP2</b>	200 MB	.CSV		Sampling is planned for 2022 or 2023		Data resulting from analyses to be kept indefinitely; PDC
Davis Strait Water Samples (Hydro) - DON/DOP	<b>WP3</b>	<10MB	.csv		June 2023	June 2026	Data resulting from analyses to be kept indefinitely; BODC
Davis Strait Water Samples (Hydro) - Ba/REE/Nd	<b>WP3</b>	<10MB	.csv		TBD		Data resulting from analyses to be kept indefinitely; BODC
Davis Strait RAS Samples - N/P/Si/DON/DOP	<b>WP3</b>	<10MB	.csv		June 2025	June 2026	Data resulting from analyses to be kept indefinitely; BODC

### Third-party Data

Dataset Name	Location	Contents	Estimated Size	Responsibility	Licence/ IPR Issues	Comments
<i>Name of dataset</i>	<i>Where is it stored</i>	<i>Brief Description</i>		<i>Who is responsible for sourcing the dataset</i>		<i>Any additional information (e.g. restrictions?)</i>
CMIP6	JASMIN	Model output for biogeochemical fields from multiple models contributing to CMIP6	2.5 TB	<b>WP3</b>	none	
NEMO-MEDUSA 1/4 and 1/12	HPC ACHER/JASMIN	Largangian forward and backward tracking 1950-2050 to determine	4 TB	<b>WP3</b>	none	Keep for the duration of the project on JASMIN;

		changes in the nutrients sources and pathways through the Arctic Ocean and into the North Atlantic.				commit to BODC
Davis Strait Mooring and Hydro data	NODC/CD C/CCHDO	Chemical, biological and physical data from Davis Strait cruises. Data from DS moorings relating to heat, freshwater and volume transports.	<100 GB	<b>WP3</b>	none	Not available until early 2025, possibly subject to embargo by partners. Will be kept indefinitely.