epoc explaining and predicting the ocean conveyor

DATA MANAGEMENT PLAN

Deliverable 7.1

EPOC – Explaining and Predicting the Ocean Conveyor

Project number: 101059547

Date: 2022/12/23

Prepared by: Rhein, M., Msadek, R., Putrasahan, D. et al.

Funded by the European Union

EPOC is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Document information sheet

EU Project No.	101059547	**Acronym:**	EPOC
Project title:	Explaining and Predicting the Ocean Conveyor	**Project website:**	www.epoc-eu.org

Deliverable:	D7.1: Data Management Plan
Work Package:	WP7: Coordination and Management

Date of delivery:	Contractual: 31 Dec 2022 (M06)	Actual: 23 Dec 2022
Dissemination level:	**PU** – Public	

Author(s) of this deliverable		
Name: Rhein, Monika	UniHB	mrhein@uni-bremen.de
Name: Msadek, Rym	CNRS	msadek@cerfacs.fr
Name: Putrasahan, Dian	MPIM	dian.putrasahan@mpimet.mpg.de
Name: Mertens, Christian	UniHB	cmertens@uni-bremen.de
Name: Frajka-Williams, Eleanor	UHAM	eleanor.frajka@uni-hamburg.de
Name: Robson, Jon	READING	j.i.robson@reading.ac.uk
Name: Desbruyeres, Damien	IFREMER	damien.desbruyeres@ifremer.fr
Name: Aksenov, Yevgeny	NOC	yka@noc.ac.uk
Name: Brown, Pete	NOC	peter.brown@noc.ac.uk
Name: Thornalley, David	NOC	d.thornalley@ucl.ac.uk

Name: de Steur, Laura	NPI	laura.de.steur@npolar.no

The Data Management Plan is accessible to all EPOC members at the internal portal https://bit.ly/3hy0MGv**.** It is a living document and regularly reviewed by the data manager (UniHB) and updated on a regular basis by the EPOC scientists.

Table of Contents

Data Management Plan

1 Data Summary

The project EPOC will generate a new conceptual framework for the Atlantic meridional overturning circulation (AMOC), to understand how it functions in the Earth system and impacts weather and climate. The data that will be produced and/or analysed in EPOC will allow to address five overarching scientific objectives:

1. Generate comprehensive records of AMOC transports across the whole Atlantic, to assess the timescales of transport variability and the degree to which the AMOC behaves as a conveyor belt.
2. Determine key processes that make or break meridional connectivity of ocean transports, and assess their representation in models, especially in high resolution coupled simulations.
3. Identify the processes and drivers of recent change in the AMOC and infer the likely roles of natural and anthropogenic forcings, and internal variability.
4. Assess the key processes of future AMOC changes and identify indicators of abrupt changes and AMOC-related climate impacts with societal relevance.
5. Design, and deploy elements of, a next generation observing system for monitoring the entire system of the AMOC.

In doing so, new observational data will be collected, and new model simulations will be carried out. Additionally, already existing observations and models will be used. The data produced by EPOC including new multi-observational data products, protocols, and output from coordinated high resolution experiments will be shared widely through the EPOC's open data policies, to enable further scientific progress beyond the lifetime of the project.

The **new observational data made in EPOC** encompass measurements of physical and biogeochemical parameters in the ocean from shipboard and from moored instruments and paleo proxy data from sediment cores, which have already been retrieved.

Provenance of existing observational data to be used: EN4 (UK Met Office), satellite altimetry (Copernicus), ARMOR3D (Copernicus), existing mooring data are available from long-term repositories (Arctic, 47°N, 53°N - PANGAEA, 26°N - BODC), air-sea fluxes (e.g., ECMWF ERA5), hydrographic data (CROSSROADS, BODC, PANGAEA), paleo proxy data (NOAA Paleoclimate, BODC, PANGAEA)

New observational data taken outside of EPOC will also be used, for instance the data from the new SWOT altimeter (launched 16 December 2022).

Provenance of existing model output: Model data for HighResMIP and CMIP6 are available through ESGF (https://esgf-index1.ceda.ac.uk/projects/esgf-ceda/) while data from PRIMAVERA are available on JASMIN and DOKU archive [DKRZ machine]

Types of data/research outputs: Observational (estimated size<1Tb), model (expected size: several hundreds of Tb) and text/images (<1Tb).

Data utility *EPOC observational data* are useful to a wide community of ocean and climate researchers from a paleo-perspective to today's climate and ocean state as well as the predictability of future climate and ocean change from decades to centuries. The data will support IPCC, IPBES, WOA IASC, and other regional and global activities, and will provide observational time series feasible to evaluate climate models.

Raw *model data* used in EPOC that are part of the highResMIP have been shared with the climate community on ESGF and will hence benefit other scientific studies in particular to assess the role of enhanced model resolution in other regions than the Atlantic.New model data that will be analysed in EPOC will produce results that can be used as a baseline to improve next generation of climate models, in particular to guide the development of new parameterizations or show where enhanced resolution is key.

Numerical data from four different coupled climate models will be generated and exploited in EPOC. Three out of the four models are employed for WP3 and WP4 to understand past and future changes of AMOC and their impacts. Past AMOC changes will be examined with so-called historical simulations as well as historical with fixed GHG experiments. Future AMOC changes will be estimated with so-called abrupt 4XCO2 experiments. Each experiment will be conducted with 3 models, namely HadGEM3-GC3.1 (READING/MetOffice), MPI-ESM-1.2 (MPIM), and CNRM-CM6-2 (CNRS-CERFACS), each in several configurations (i.e. different spatial resolutions). . A fourth model, ICON-ESM (MPIM), will be employed to produce a 30-year (transient 1993-2022) ultra-high resolution simulations designed within EPOC to study meridional connectivity of the AMOC for WP2. Table 1 summarises the model configurations and the simulations that will be the focus of EPOC WP3 and WP4. Table 1 also states whether the experiments are new i.e., produced in EPOC or whether they have been produced in previous multi-model projects (CMIP6/HighResMIP and PRIMAVERA) and will be used in EPOC. Note that while CNRS-CERFACS partner has contributed to PRIMAVERA and HighResMIP with the CMIP6 CNRM-CM6-1 model, all the simulations produced by this partner for

EPOC will be done with a new coupled model CNRM-CM6 that shares the same components as CNRM-CM6-1 except for the sea ice component. This model has been developed to allow an increase of resolution in the ocean/sea ice component that was not possible with the former model. Hence all the LR and HR simulations need to be done by CNRS-CERFACS partner unlike the other two modelling partners (MPIM and Reading).

Coupled model/ Configuration	**Control**	**Historical**	**Historical with fixed GHG**	**Abrupt 4xC02**
HadGEM3-GC3.1/LL	HighResMIP (control-1950;30 years)	HighResMIP (hist-1950, 70 years)	EPOC (70 years)	EPOC (30 years)
HadGEM3-GC3.1/HH	HighResMIP (control-1950;30 years)	HighResMIP (hist-1950, 70 years)	EPOC (70 years)	EPOC (30 years)
MPI-ESM-1.2/HR	HighResMIP(control-1950; 100 years)	HighResMIP (hist-1950, 70 years)	EPOC (70 years)	PRIMAVERA (100 years)
MPI-ESM-1.2/ER	HighResMIP (control-1950; 100 years)	HighResMIP (hist-1950, 70 years)	EPOC (70 years)	PRIMAVERA (100 years)
CNRM-CM6/LR	EPOC (control-1950; 30 years)	EPOC (hist-1950, 70 years)	EPOC (70 years)	EPOC (30 years)
CNRM-CM6/HR	EPOC (control-1950; 30 years)	EPOC (hist-1950, 70 years)	EPOC (70 years)	EPOC (30 years)

Table 1: Model configurations used in EPOC and simulations analysed or produced in WP3 and WP4. EPOC means the runs will be produced during the project. Other projects (e.g. PRIMAVERA or HighResMIP) means the simulations have been already produced in the framework of previous multi-model projects.

Type and format : Numerical simulations will result in gridded output, primarily on the model native grid but also possibly alternatively on an interpolated grid for usability or data volume reduction purposes. Modelling centres will also derive, from the primary model output, elaborated diagnostics and metrics. Both types of data (primary and derived) will be in the netCDF format and respect the CF conventions (Climate Forecast Metadata Conventions, http://cfconventions.org/).

Expected size: The amount of data resulting from the numerical runs is not yet known but likely to reach hundreds of Terabytes (TB). The modelling groups are currently working on a common data request. Once this data request is defined, estimates of the data volume can be made. Precise numbers will be provided in a next version of the DMP. It is to be noted that only a subset of the whole model's production is intended to be shared across the model members for analysis.

2 FAIR data

2.1 Making data findable, including provisions for metadata

Findability of data/research outputs: Observational products generated by EPOC will be deposited in appropriate long-term repositories (e.g. PANGAEA, http://pangaea.de, BODC - British Oceanographic Data Centre http://www.bodc.ac.uk) . PANGAEA assigns persistent DOIs to data and is cross-linked on data portals (including an EPOC external-facing data portal to be created and to exist for a minimum of five years beyond the project end). Pangaea also enables cross-links between data and literature, to connect scientific publications with their underlying data. We will seek to adhere to All-Atlantic Ocean Data Space recommendations for ocean data and metadata. Metadata are also gathered according to ISO19115 and INSPIRE. Data is labelled with uniform, internationally used variables and SI-units. Standard vocabularies are used. Assumptions made are clearly stated.

Regarding model products, the climate modelling community has long been committed to a FAIR approach, made necessary by the CMIP climate model intercomparison exercises led by the World Climate Research Programme (WCRP). The three modelling centres contributing to EPOC are all heavily involved in these exercises, in particular the recently achieved 6th phase (CMIP6). The data produced in PRIMAVERA and reused in EPOC was published on the Earth System Grid Federation (ESGF) as part of CMIP6 (HighResMIP). They are discoverable and accessible though the ESGF/CoG API (https://esgf-data.dkrz.de/search/cmip6-dkrz/). The data produced in EPOC will be

stored on CEDA and WDC archives, available there and discoverable through the search and download interfaces (https://data.ceda.ac.uk/ and https://www.wdc-climate.de/ui/q respectively). PRIMAVERA data can also be browsed and downloaded there.

In addition to data outputs, project meeting results and documents (e.g., slides) will be archived at ZENODO (http://zenodo.org).

Naming conventions and metadata standards: According to the object, naming conventions come with different standards. The CF conventions set general naming rules (e.g., the alphanumeric characters allowed or imposed), set their own standard (e.g., for standard name of variables) and rely on external specialised ones (e.g., units names must follow UDUNITS provided by unidata). File naming rules come with the CMIP6 DRS (Data Reference Syntax). The ontology for the file global attributes (infile metadata) and file name facets is the CMIP6 Controlled Vocabulary (https://wcrp-cmip.github.io/CMIP6_CVs/) setting the vocabulary for institutions, models, experiments, grids, realms, etc. Metadata for variables are gathered in the so-called “CMIP6 Tables” that provides its standard name, its numerical type, its units, the computation method (both spatial and temporal operations)

Versioning: For the sake of traceability, a version number will be associated with every model data set. A data set can potentially be made of several netCDF files, the version number could be included in each of them as a global netCDF attribute named “version_id” (note that the CMIP6 DRS convention already defined this metadata but only imposed as a facet of the data directory tree structure, not as file metadata). The version number choice is up to the data provider but must be unique for the same batch of data files (i.e., data obtained from the same simulation run). If a numerical experiment is rerun after an issue was discovered and fixed, the version of the data re-generated must be different from the original one.

Similarly, for observation data, if several versions of a data set are created, each needs an individual number and an extra entry. The numbers should enable us to discern which version is the most recent one. The name should also include EPOC.

Observational data in publications: In accordance with standards of good scientific practice, all observational data analysed in publications need to be stored together with the necessary metadata in appropriate long-term repositories (e.g., PANGAEA, http://pangaea.de). Additionally, the main products of the publications calculated from the data, including the software to calculate the product should be – where possible - also stored in the chosen repository and linked to the publication. All data sets used from outside EPOC need to be cited in the publication, providing the download website, the version of the data downloaded and the date of the download.

Models in publications: In accordance with standards of good scientific practice, all primary data need to be archived for a period of at least ten years. The institute where the model runs are carried out is responsible for the storage. Primary data is the information which is necessary for others to replicate the main ideas of a study. For numerical investigations the primary data is a documentation of the methods and tools used, in a way that would allow the ideas of a study to be reproduced. In many cases it may be more efficient, and less error-

prone, to meet this requirement by archiving the code or scripts (e.g., through Zenodo with links to GitHub) used in a study and documenting how they were executed. Model output is *not* primary data and need not be archived, although doing so can be helpful. Models are often modified, so code changes need to be documented. In cases where models are not maintained as part of an institute-maintained code repository, it is important to save the model code itself and the configuration specifications as part of the primary data archive, which would then be provided to the institute either as part of the electronic supplement, or in the form of a PID serving as link to an external repository where it is stored by a trusted source.

2.2 Making data openly accessible

Accessibility of data/research outputs: Primary observational products generated within EPOC will be made open access following time required for quality control and assurance. Key model outputs, including those needed to support reproducibility of research results, will be archived and openly available, while metadata will be permanently archived with persistent DOIs. All EPOC publications will also be open access.

Internal data exchange is ensured within EPOC before final data publication. Metadata and data are shared by the EPOC scientists. The responsible persons can be contacted for detailed questions and/or for data exchange.

EPOC scientists will be notified before an EPOC manuscript is submitted and the submitted version is made available.

Relevant data will be made openly available to EPOC partners throughout the project, in accordance with the intellectual property (IP) management plan for EPOC. Data that is archived in long-term storage will focus on data used within EPOC publications and will be openly available to the wider community. Inevitably some data will not be made available indefinitely due to the cost of storage. No data will deliberately be kept closed, except according to the conditions outlined in the Consortium Agreement (CA), due to data sensitivity, or if restricted by data protection law. If data is not available, it can be for storage space or manpower limitations. A responsible person in each of the 3 modelling institutes will be nominated so that users looking for data can contact him directly.

University of Reading data will be available through JASMIN for the EPOC partners and through CEDA LTA for the wider community. The MPIM will make its data available to EPOC partners via JASMIN or DOKU and to the wider community though DOKU or WDCC. CNRS-CERFACS will provide access to its model data through JASMIN for sharing across EPOC partners and CEDA or WDCC for a wider dissemination.

In order to ensure reproducibility of research results, the subsets of numerical simulation output associated with scientific publications (hereafter referred as “primary model data”) will be stored on long-term repositories with DOIs assigned (e.g., CEDA, https://archive.ceda.ac.uk/) for a period of at least 10 years. For non-primary data (whole model production and other derivative products of interest but not yet exploited in a scientific publication), the producer institute is responsible for their storage and preservation (or re-computation, the case being). The metadata storage and preservation is not a separate process in the case of model data—they are included in the data files and handled at the same time.

2.3 Making data interoperable

Interoperability of data: PANGAEA has a collaboration with the harvesting portal EMODnet (European Marine Observation and Data Network) to support the principle of “collect once, use many times”, promote the interoperability of datasets, and maximise benefit. We will work with GEOSS to create a portal for data products from EPOC and, where possible, through the wider AMOC observing community’s data distribution efforts. Specifically, the (i) CLIVAR AMOC Task Team has one aim to “improve data and product distribution from AMOC programs” to identify data products (i.e., beyond raw instrument data), types and formats preferred by end users, and (ii) building on AtlantOS efforts and OceanSITES’ mission to “collect, deliver and promote the use of high-quality data from long-term, high-frequency observations at fixed locations in the open ocean”, and (iii) the All-Atlantic Ocean Data Space is generating recommendations for ocean data and metadata standards and formats. We will work with these groups to create an externally facing EPOC AMOC-specific portal within GEOSS to improve interoperability and access to AMOC observing outputs. Making EPOC data available here will enable wider access to and exploitation of the outputs from EPOC and AMOC observing arrays.

The interoperability of model data is ensured by their strict conformance to the reference standard in the Earth System Modelling community. These standards are described in detail in section 2.1.Model outputs and gridded datasets will be stored as netCDF files. NetCDFs are machine-readable only and need particular libraries to be read; however, these libraries are freely available as stand-alone tools and included in many data manipulation software and programming languages, (Python, R, MATLAB/Octave, FORTRAN, etc.). These netCDF files will use CF conventions and are compliant with CMIP6 protocols for the new model runs completed in the project. Where possible, the subsets of the data from the existing model runs will be made CF and CMIP6 compliant. Naming rules and controlled vocabulary will be employed following CMIP6/HighResMIP. The netCDF file format is self-describing and machine actionable. Most programming languages support this format (Python, R, MATLAB/Octave, FORTRAN, etc.) and Python libraries dedicated to netCDF data access exist and are widely used (e.g., Xarray, Iris, etc). They facilitate data access and data array manipulation. Beyond the file format, a standardised data description is also crucial to ensure interoperability, to enable the automatic recognition of fundamental elements (netCDF attributes or fields)

such as coordinates, time axis and units, calendar attributes, time frequency, data units, missing values, etc. CMIP6 Tables ensures this, enforcing both the name of the attribute and its value for each output variable.

2.4 Increase data re-use (through clarifying licences)

Reusability of data/research outputs: Datasets generated within EPOC will be available under Creative Commons CC-BY licence. Software developed and used to generate research results will be available with an appropriate software licence to enable free access. The data behind peer-reviewed publications will be published and openly accessible at the time of publication or earlier, following the guidelines of H2020 and those of the journals. The remaining data will be made available at the end of the EPOC project. Data embargoes that extend beyond the time of the project will, as a general principle, not be permitted. If necessary, they will be reviewed in a case-by-case manner; an end date will be required and that the responsibility for data publication returns to the PI. In accordance with the common best practices and protocols of the community for data collection, quality control and assurance are provided in the metadata.

Valid reasons for an embargo period on data are primarily for educational reasons, allowing Ph.D. students to prepare and publish their work. Even if data is constrained during the embargo period, data will be shared internally within the project. Any disagreements on access to data or misuse of data internally are to be settled by the EPOC Steering Committee and General Assembly.

3 Allocation of resources

Curation and storage/preservation costs: UNIHB will be responsible for the data management plan, while individual task leaders will be responsible for quality assurance of the data generated in their task (see EPOC work plan Tables 3.1a). Storage costs for observational datasets generated within EPOC are included in purchase costs. Storage costs for model output generated by Reading are included within EPOC (JASMIN and CEDA), while model outputs generated by CNRS will be archived on Météo France machines, and by MPIM on DKRZ machines, at no additional cost. For sharing model data within the consortium, we will use a space within JASMIN that is accessible to EPOC partners for subsets of model output, or ftp for transfer of data between machines. Model outputs are exceedingly large (100s TB) and so for open access to model output, priority will be given to making key subsets available in order to support reproducibility of research results.

4 Data security

Immediately after observational data are recovered from moored instruments, the raw data are stored on at least two different storage devices for at least 10 years. The storage of the raw data is the responsibility of the institution that carries out the observations. Only authorised personnel have access to the data. This procedure is also applied to final data and products derived from them. The responsibility for submitting the final data and metadata to an appropriate repository (e.g., PANGAEA) is held by the PI the data originate from.

Model versions and related primary data are stored within each modelling centre and this archive will enable the recovery of data in case of partial loss or any other incident occurring in the data centres (e.g., CEDA) where the model outputs are shared among partners and made available to the wider community.

5 Ethical aspects

Ethical aspects will be covered in the context of the ethics review, the ethics section of the DoA and ethics deliverables. Informed consent for data sharing and long-term preservation will be included during data collection. Data that includes personal information will go through an anonymization process before being published. Informed consent will be requested from the participants of these data collections. Sensitive data will be separated as soon as possible and kept secure.

6 Other issues

No other issue identified.

7 Tables

The following tables provide an overview of observational data used in EPOC or related to EPOC.

General information about EPOC and EPOC-related expeditions are contained in Table 2. Observational data and data products to be produced by EPOC are in Table 3. Observational data and data products to be used from sources outside of EPOC are in Table 4.

Table 2. General information about EPOC and EPOC-related expeditions. At time of writing (Dec 2022), some details were still TBD (to be determined) or TBC (to be confirmed).

WP	Expedition name	Platform	Date	Port of Start/End	Research Area	Principal Investigator / Chief Scientist	Project	Organizing Institution
5	RAPID TBD	R/V Endeavour	Feb 2023	Florida, USA	Bahamas	Ben Moat (NOC)	RAPID	NOC
2	MSM 121	R/V Merian	Sep 2023	Nuuk/ Pt Delgada	Flemish Cap	Christian Mertens (UHB)	EPOC	UNIHB
2	CROSSROAD TBD	N/O Thalassa	summer 2024	TBD	Flemish Cap	Damien Desbruyères (IFREMER)	CROSSROADS	IFREMER
2	TBD	N/O Thalassa	summer 2025	TBD	Flemish Cap	Damien Desbruyères (IFREMER)	EPOC	IFREMER
5	RAPID TBD	TBD	Feb 2024	Canaries, Spain	Canary Islands	TBD	RAPID	NOC

Table 3. Planned data products to be produced within EPOC, named by Task number. Variable names are shorted where T – temperature, S – salinity, P – pressure, C – conductivity. WP – Work Package. Due dates are given in MXX where XX is the month of the project since start: e.g., M1=31 July 2022, M24 = 30 June 2024.

Name of the data x.x EPOC	**Data description** (e.g., expedition, measuring device, sample, method, analysis, model, ...)	**Parameter(s)**	**Responsible person**	**Internal data availability** Planned due date	**Data format** Estimated size and number of files	**Data publication**		**Archiving of used software** (optional)
						Planned due date	Planned repository (with URL), DOI, label, licensing, if applicable: embargoes and moratoria)	Planned due date, repository, label
1.1 EPOC	Multi-observational estimate of AMOC-related heat and freshwater transports from 35°S to 80°N, 1993-2022	Timeseries of heat and freshwater at seasonal (3-month) resolution.	Kiko Calafat (NOC) -Task 1.1	M24	netCDF < 1GB, 1 or 2 files	M36	BODC (https://www.bodc.ac.uk) and/or www.pangaea.de	
1.2.EPOC	Arctic Inverse model output	Gridded monthly mean T, S, velocity from the Arctic Gateways Additional products: Time Series of	Laura de Steur (NPI) - Task 1.2	M24	netCDF ~ 10 GB total in 120 files (monthly)	M30	www.pangaea.de and/or data.npolar.no	

		temperature and freshwater transports and overturning.						
2.1a EPOC	3D data set of Atlantic currents from altimetry and hydrography, 1993—2022	Gridded monthly volume transports	Christian Mertens (UniHB) - Task 2.1	M24	netCDF ~ 10 GB in ~340 files	M30	www.pangaea.de	
2.1b EPOC	AMOC transport time series from 47°N observations (1992—2022)	Monthly mean transports	Christian Mertens (UniHB) – Task 2.1	M24	netCDF ~< 1GB in 1 file	M30	www.pangaea.de	
2.3a EPOC	Mooring array in Flemish Cap region	Sensor data from T, C, P and velocity sensors Additional product: Travel time and pressure sensor data (PIES)	Damien Desbruyères (Ifremer)	M42	netCDF <1 GB, ~50 files	M54	www.pangaea.de	
2.3b EPOC	Paleoceanographic reconstruction from Flemish Cap region	210Pb, Sortable silt, foraminiferal abundances and geochemistry	David Thornalley (UCL)	M42	CSV format <1 GB total	M54	www.pangaea.de	
2.4 EPOC	PIES array in Flemish Cap region	Travel time and sensor data from P, T and velocity (optional) sensors Additional product: Time series of volume transport	Christian Mertens (UniHB)	M42	tab-delimited text <1 GB, ~50 files	M54	www.pangaea.de	

5.3 EPOC	Canaries region, RAPID-BGC moorings in eastern boundary system	Sensor data from T, S, P, oxygen, pCO_2, pH, alkalinity, nitrate, phosphate sensors and calibration water samples	Pete Brown (NOC) - Task 5.3	M18	netCDF / CSV / tab-delimited text <100 MB, ~100 files	M30	BODC (www.bodc.ac.uk) and/or www.pangaea.de	

Table 4. Planned data products to be used within EPOC, named in order of work package. Variable names are shorted where T – temperature, S – salinity, P – pressure, C – conductivity. Data products to be produced are linked (where specified) by the 'Products' column to Table 2. This table will be appended as additional datasets are identified to be used and version numbers or data versions stored for use within the project. Additional details including the date of download will be added in the data description column.

WP	**Data description** Including download website, which version and date of download	**Parameter** Which parameters are used in the EPOC analysis	**Products** Which EPOC products are calculated with these data, or what analysis planned	**Data format** Estimated size and number of files of the product	**Product published in publication and put in repository**	
					Storage planned due date	EPOC Publication planned due date
1	Gridded ocean bottom pressure data from GRACE gravimetry (RL06.1M.MSCNv03) https://podaac.jpl.nasa.gov/dataset/TELLUS_GRAC-GRFO_MASCON_CRI_GRID_RL06_V2	Ocean bottom pressure	1.1 EPOC (Table 2)	netCDF Size<100 MB	M12 Version used will be recorded.	WP1 – M36
1/2	Copernicus Marine Service , Global Ocean Gridded Sea	Sea Surface Heights	1.1 EPOC (Table 2)	netCDF Size 10 GB	M24	WP1 – M36 WP2 – M30

	Surface Heights, http://doi.org/10.48670/moi-00148		2.1 EPOC (Table 2)	340 files	Version u sed will be recorded	
1/2	Met Office, EN4: quality controlled subsurface ocean temperature and salinity, https://www.metoffice.gov.uk/hadobs/en4/	Ocean T and S	1.1 EPOC (Table 2) 2.1 EPOC (Table 2)	netCDF	M24 Version used will be recorded	WP1 – M36 WP2 – M30
1/3	Atlantic meridional overturning circulation observed by the RAPID-MOCHA-WBTS array at 26°N from 2004 to 2020 (v2020.2). doi:10.5285/e91b10af-6f0a-7fa7-e053-6c86abc05a09 http://www.rapid.ac.uk	Ocean T, S, density, P and transports at the 26°N RAPID-MOCHA section	1.1 EPOC (Table 2). RAPID Section properties (T, S, density, velocity and ocean heat/freshwater transports) to facilitate comparison of the models with the observations	netCDF Size<1 GB 3-4 files	Published via BODC on the 23 Sep 2022 Version used will be recorded.	WP1 – M36 WP3 - TBD
1/3	Overturning in the Subpolar North Atlantic Program (OSNAP) international program to provide a continuous record of the full-water column, trans-basin fluxes of heat, mass and freshwater in the subpolar North Atlantic at the	Gridded ocean potential temperature, S and velocity fields at OSNAP sections for 2014—2018	1.1 EPOC (Table 2). OSNAP Section properties (T, S, density, velocity and ocean heat/salt transport) to	netCDF Size<1 GB 3-4 files	Publication dates are Sep-Oct 2018. Data links to the individual datasets are published via the OSNAP website	WP1 – M36 WP3 - TBD

	OSNAP West and OSNAP East sections. https://www.o-snap.org/data-access/ Duke Digital Repository. https://doi.org/10.7924/r4z60gf0f		facilitate comparison of the models with the observations		Version used will be recorded.	
1	https://data.npolar.no/dataset/c4d80b64-25f6-4afd-b392-696430c3fd14 https://data.npolar.no/dataset/5b717274-2d85-4f13-a1b4-ff0517c78b4a https://doi.pangaea.de/10.1594/PANGAEA.900883 https://doi.pangaea.de/10.1594/PANGAEA.904565 http://psc.apl.washington.edu/HLD/Bstrait/Data/BeringStraitDownloadregister.html https://iop.apl.washington.edu/data/DavisStrait/cruises/index.php https://doi.org/10.21335/NMDC-1838527821 Mooring data for Fram, Bering, Davis straits and BSO are public up to 2020, 2020, 2016 and 2017	Ocean T, S and velocity from Fram Strait, Bering Strait, Davis Strait and Barents Sea Opening (BSO) Sea ice draft and/or sea ice volume transport time series from Fram Strait	1.2 EPOC (Table 2)	netCDF Size<10 GB ~50 files	Versions used will be recorded.	WP1 – M36

	resp. Updates to 2020 for BSO are in prep still (status Nov.2022)					
1	Data set to available from NMDC first half 2023. https://nmdc.no/nmdc Mooring data are public until 2022.	Atlantic water transport and hydrography. Moorings and repeated hydrographic sections	1.2 EPOC (Table 2)	netCDF Size<10 MB	Øystein Skagseth (MR)	WP1 - TBD
1	Paleoreconstruction of DWBC (David Thornalley, UCL)	Sortable silt grain size	Current strength	CSV format Size<1 GB	M42	Data are in prep and will be published by M42
2	Hydrography survey in Flemish Cap region	Ocean T, S, P, C, oxygen and velocity data (~100 full-depth profiles). Microstructure profiles (moored and shipboard)	2.3a EPOC (Table 2)	netCDF Size<1 GB	M32	WP2 - 2026