



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 869383

Initial Data Management Plan (D5.3)

Work package: WP5

Deliverable No. – Title: D5.3 - Initial Data Management Plan

Lead Beneficiary (P11): IO PAN

Lead responsible for the report (Name, Institution): Marcin Wichorowski & Artur Palacz (IO PAN)

Editors (Name, Institution): Andre Visser (DTU)

Contributors (Name, Institution): Rikke B. Jacobsen (AAU)

Type: Open Research Data Pilot

Dissemination level: Public

Submission Date: 30/11/2020

Project Number: 869383

Project Acronym: ECOTIP

Project title: Arctic biodiversity change and its consequences: Assessing, monitoring, and predicting the effects of ecosystem tipping cascades on marine ecosystem services and dependent human systems

DATA MANAGEMENT PLAN

Document Revision History

Date	Issue	Change Records	Authors
2020-11-30	1.0	Created document, structure, chapter headings, content version 1.0	Marcin Wichorowski, Artur Palacz

Table of contents

Document Revision History	2
Table of contents.....	3
1 Rationale of the project.....	4
2 DATA MANAGEMENT APPROACH	4
3 DATA IDENTIFICATION & DESCRIPTION.....	5
3.1 Data types and standards	5
3.2 Re-use of existing data.....	7
4 DATA ORGANISATION & EXCHANGE	9
5 DATA STORAGE AND BACK-UP	9
6 DATA SHARING, STANDARDS & METADATA.....	9
7 DATA RESTRICTIONS.....	10
8 DATA PUBLISHING & LICENSING.....	11
9 DATA ARCHIVING.....	12
10 REFERENCES.....	13

1 Rationale of the project

The ECOTIP project will improve our understanding of the causes, thresholds and consequences of the ecosystem tipping cascades on Arctic marine biodiversity and develop methods to assess, monitor and predict them. ECOTIP addresses marine ecosystems from microbes to marine mammals, by integrating biological observations into a traitbased approach to marine biodiversity, and by utilising the trait-based approach to link the changes in biological communities to ecosystem functions, and finally to the socio-economics of local communities and ecosystem services of global importance. To this end, ECOTIP will collect new observations, develop novel methodologies and tools, apply new concepts and process understanding, both on biological systems and on human systems, and integrate these for the optimal exploitation of the results.

The project has five main objectives:

- To map the current biodiversity of Arctic marine ecosystems and its past and present interaction with external drivers (multiple stressors), using traits as a measure of functional diversity
- To investigate the vulnerability of marine communities (with different trait compositions), functions and ecosystem services to multiple climatic and non-climatic stressors, and to determine their potential for ecosystem tipping points
- To use the analysis of functional (trait) diversity to predict a) changes in the local production and type of fisheries and b) in carbon sequestration by biological pump under multiple anthropogenic stressors
- To engage in dialogue and co-creation of alternative governance structures and adaptation strategies for the local and indigenous communities, as well as industries and regulatory authorities
- To ensure effective exploitation of the project results in international scientific assessments of Arctic biodiversity change and by policy-makers, to ensure dialogue, communication and dissemination to indigenous societies and European citizens, and to provide recommendations for optimizing the monitoring of Arctic biodiversity and ecosystem service

Successfully meeting these objectives relies on a strengthened and optimized flow of data and information both within the project consortium and beyond it. A Data Management Plan (DMP), meeting the requirements of an Open Research Data Pilot (ORDP), will support dissemination and exploitation of newly collected observations and model results, along with their analysis and synthesis.

The DMP will serve the needs of and be implemented by the entire project consortium consisting of 16 partners, listed in the ECOTIP Grant Agreement.

2 DATA MANAGEMENT APPROACH

The Data Management Plan of ECOTIP will lay out the means of making data available in a readily useable format, optimizing handling of data from discovery to publication, and facilitating data exploitation by the wider community. To this end, the project will adhere to FAIR (findable, accessible, interoperable and reusable) data principles. By providing timely, free and unrestricted access to relevant data and associated metadata wherever possible, it will also adhere to the IOC-UNESCO Strategic Plan for Data and Information Management.

ECOTIP agrees to participate in the Open Access to Data Pilot established by the European Commission on a voluntary basis. This is reflected in the Art.29.3 included in the Grant Agreement No. 869383 — ECOTIP.

ECOTIP conforms with H2020 requirements to make all peer-reviewed scientific publications generated by ECOTIP available in open-access. This means, at the very least, that we will ensure that these publications can be read online, downloaded and printed. The data management team will ensure that:

1. a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication is deposited in a repository (online archive) for scientific publications, as soon as possible and at the latest upon publication.
2. open access to publications is provided through:
 - a. self-archiving (green open access)
 - b. open access publishing (gold open access)
3. open access — via the repository — is provided to the bibliographic metadata that identify the deposited publication

DTU manages centrally allocated resources for at least 30 publications in green open access mode. Whenever possible, use of gold open access will be encouraged through one of many open access journals searchable at <https://doaj.org/> - an independent, community-curated online directory that indexes and provides access to high quality, open access, peer-reviewed journals. Consortium partners will be able to choose between self-archiving preprints in the project repository (see below), or at one of the open access reprint servers such as zenodo.org, arxiv.org, or more specifically for biological results: biorxiv.org. For details, and an open access checklist, please see section 4.3. and Appendix to the Plan for Exploitation and Dissemination of Project Results.

Consortium partners DTU and IO PAN constitute the body designated to coordinate the data management activities. The ECOTIP partners' IT Departments provide necessary expertise and capacity to assure proper data management.

A designated Data Manager, a data expert appointed by IO PAN has been incorporated into the project team. The Data Manager's main tasks are to coordinate the teams engaged in data acquisition and support them with tools for data transformation and upload. He is responsible for implementation of the Data Management Plan and coordination of data evaluation and overview of data management practices within the data life cycle frame.

3 DATA IDENTIFICATION & DESCRIPTION

3.1 Data types and standards

Ship-based measurements and fixed-point time-series observations will be the main means of collecting in situ ocean data for the project. ECOTIP will collect data on parameters spanning all ocean disciplines: physical, chemical, biological and geological, ranging from coastal to open ocean, from surface to deep ocean. In addition, socio-economic data collection and analysis will be conducted. The majority of measured parameters are considered as sub-variables of Essential Ocean Variables as defined by the Global Ocean Observing System, for many of which there exist guides, best practices or standard operating procedures on how to obtain comparable, high-quality measurements. Data and model-

derived products, such as indices or indicators, will also be generated according to the requirements established via stakeholder consultation workshops.

In general, OGC (Open Geospatial Consortium), ISO, NetCDF and the International Oceanographic Data and Information Exchange (IODE) standards will be used. The project will ensure the use of platform and parameter specific standards, if existing and as listed in the relevant EOV Specification Sheets (www.gooscean.org/eov). For ship-based hydrographic data (such as temperature, salinity, nutrients and carbonate system parameters), we will refer to standards established by the Global Ocean Ship-Based Hydrographic Investigations Program (GO-SHIP; www.go-ship.org/HydroMan.html) and the International Ocean Carbon Coordination Project (IOCCP; www.ioccp.org/documents/standards-and-methods). Fisheries data from acoustic surveys, as well as fish and plankton data collection will follow ICES metadata standards and sampling protocols (www.ices.dk/marinedata/guidelines-and-policy/). In all cases, the project will ensure the use of community best practices, e.g. as available at Ocean Best Practices repository (www.oceanbestpractices.org/), an activity coordinated by IOC-UNESCO.

Level 0 data (raw data) will be acquired and stored for further processing using variety formats. Data collected in situ depends on design of the measuring equipment’s firmware. In most cases measuring equipment either is supported by software tools provided by the vendor for data preprocessing, or provide text files with structured data. Samples collected in situ have to be processed together with associated metadata and deliver text files (.csv), numerical files (NetCDF .nc format) and multimedia material (photos and video sequences). These multimedia deliver large binary files.

ECOTIP will also generate model outputs which will form the basis for scientific publications. Outputs will be stored either as text files or numerical files. All model code developed during the project’s lifetime, together with supporting documentation, will be made publically available on GitHub (web-based hosting service for version control using Git) to foster further exploitation and innovation.

Variables or parameters measured, or derived from raw data, by organizations involved in ECOTIP are grouped according to [BODC P03 vocabulary](#), wherever applicable. A full catalogue with a detailed list of parameters under each group will be provided as more details about sampling campaigns and studies become available.

Table 1: Initial overview of data to be produced in ECOTIP.

Domain	Data level	Type/method	Parameter groups (BODC P03 vocabulary)
socio-economics	raw data	personal interviews	
	raw data	economic data	fisheries
	model	fisheries fleets and governance models	
	raw data	local ecological knowledge (LEK)	fish; other biological parameters
physical oceanography	raw data	seawater samples	water column temperature and salinity
	synthesis product	climate and stressor scenarios	water column temperature and salinity

biogeochemistry	raw data	seawater samples	dissolved gases; nutrients; carbonate system; suspended particulate material
	raw data	sediment traps	fluxes
	synthesis product	climate and stressor scenarios	dissolved gases; carbonate system
biology	raw data	laboratory experiments	zooplankton; rate measurements (including production, excretion and grazing); fatty acids
	raw data	net tows	zooplankton
	raw data	acoustic surveys	fish
	synthesis product	remote sensing merged with in situ	pigments; phytoplankton and microphytobenthos
	raw data	seawater samples	pigments; bacteria and viruses; rate measurements (including production, excretion and grazing)
	raw data	omics, proteomics	other biological measurements
	raw data	eDNA	other organic chemical measurements
	raw data	chemical analysis	biota composition
	model	statistical model	fish
	model	trait-based model	phytoplankton and microphytobenthos; zooplankton
palaeo oceanography	raw data	sediment cores	rock and sediment physical and chemical properties; rock and sediment biota; other organic chemical measurements

The storage allocated for the volume of data acquired during field expeditions and collected archive data used for this research is up to 30TB including numerical model outputs, satellite images and multimedia files. This storage will be extended according to growing demands of project development teams and optimized to comply technology advances. Data will be preserved for eternal use and availability for users.

Samples will be preserved at least 5 years after publication of data and termination of the project.

3.2 Re-use of existing data

To complete the work and meet project goals, extensive review of archive data and data available in repositories, especially dedicated for Arctic data will be conducted. Relevant data will be tagged and linked into ECOTIP Data Collection. For operational use data will be replicated locally, if necessary.

A preliminary list with examples of existing data which ECOTIP anticipates to use is provided in Table 2. An extensive review is being carried out under Tasks 1.1 and 2.1. The results of these reviews will be incorporated into the DMP in future revisions.

Table 2: Examples of existing data and products to be used in ECOTIP.

Data set description	Location	Time	Domain and BODC P03 parameter groups	Type	Source/partner contact
Surface Ocean CO ₂ Atlas (SOCAT)	global	1970-present	Physical and chemical: water column temperature and salinity; dissolved gases; carbonate system	underway/in-line observations; gridded product	www.socat.info
GLODAPv2.2020	global	1970-present	Physical and chemical: water column temperature and salinity; dissolved gases; carbonate system; nutrients	GO-SHIP repeat hydrography lines; gridded product	www.glodap.info
Earth System Model outputs (CMIP5 & 6)	global	present-2100	Physical and chemical: water column temperature and salinity; currents; dissolved gases; carbonate system	model outputs	H2020 project COMFORT; Excellence Cluster of the University of Hamburg; Copernicus Climate Data Store (CDS)
ESA Climate Change Initiative	global	1981/1998-2019	Physical, chemical and biological: water column temperature and salinity; Phytoplankton and microphytobenthos	remote sensing gridded products	climate.esa.int
Existing alien species datasets	Arctic partial		Biological data: biota abundance, biomass and diversity		EMODnet Arctic
Reconstructed hydrographic conditions	multiple sites in W Greenland, E Greenland and Iceland	The present interglacial and the preceding deglacial period	Paleoceanographic data: rock and sediment physical and chemical properties; rock and sediment biota; other organic chemical measurements	time series, sediment cores	AU
Greenland Ecosystem Monitoring	W Greenland (Nuuk, Disko Bay), E Greenland	2005 - present	Physical, chemical and biological data	station time series	https://data.g-e-m.dk

	(Zackenber Station)				
Off-shore and in- shore fishery surveys	Greenland	1991-present	Biological data: fish	longline, gillnet surveys	GINR
Acoustic survey data	Iceland-East Greenland- Jan Mayen	1995-present	Biological data: fish	acoustic surveys	data.hafro.is
Zooplankton data sets	Subpolar North Atlantic	1980's-present	Biological data: zooplankton	Time series and stations	DTU
Fisheries datasets	W and E Greenland	1991-present	Biological and economic: fisheries	logbooks and landing slips	GINR
LITR Observatory HAUSGARTEN	Fram Strait	1999-present	Physical, chemical and biological data	Time series	AWI/HZG
Arex cruises	Arctic Ocean	1995-2020	Biological data	ship-based transects	IO PAN

4 DATA ORGANISATION & EXCHANGE

Data will be extracted from raw files and stored in a database for further processing including quality control (QC) and metadata discovery. These data will be processed to elaborate complex products for internal and external use. Data derived from raw files is subject to maintenance by the data originators. They are responsible for enhancing datasets with metadata and data quality assurance. Quality controlled data is shared between project partners working on development of data products. Data products can be developed on different levels of complexity, to assure data interoperability and reusability by external users. External users have open access to all published data, featured with FAIR data principles (Wilkinson et al., 2016) of findability, accessibility, interoperability and reusability within the time frame of the project and after project termination.

5 DATA STORAGE AND BACK-UP

Data storage is provided by IO PAN Data Centre. Infrastructure of the Centre is HA enable, electronic equipment is powered by uninterruptible power sources, and data is replicated to PLGRID – an external private cloud developed for scientific purposes on national level in Poland. Local archive is supported by Barracuda Tape Library certified for two hours fire proof. This tape library is part of the distributed file system providing continuous backup activity.

6 DATA SHARING, STANDARDS & METADATA

To enhance FAIRness of the data and data products of ECOTIP, all data generated by project will be associated with metadata. Data findability is assured by the fact that each dataset or product released will have a DOI number assigned to it. This service will be offered both by the IO PAN Data Centre, and optionally also by Zenodo's ECOTIP community (<https://zenodo.org/communities/ecotip-arctic/>). The

most popular standards of metadata are these derived from ISO19115 and ISO 19136 including INSPIRE and SeaDataNet. All necessary metadata will be discovered by ETL tools and associated with data. To exchange data with external users (including automated systems) metadata will be attached to data requested by users on-the-fly according to demanded format.

The formats of data provided by system will be developed according to identified users needs (both internal and external), initially delivered as NetCDF, ODV and CSV (text) files.

7 DATA RESTRICTIONS

Each organization is responsible for ensuring that their collection, processing and sharing of personal data and / or special categories of personal data are in compliance with Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (the General Data Protection Regulation (GDPR)) and other applicant regulation on personal data. The Parties will therefore ensure the legal foundation for the data and ensure the contractual basis and take required security measures in accordance with GDPR before sharing any personal data and / or special categories of personal data. The Parties agree that any sharing of data to the widest extent possible will be in completely anonymized form not containing any personal identifiable data and therefore not subject to GDPR. If it is not possible for the Parties to share personal data in completely anonymized form, the Parties agree to use the EU provided templates for transfer of personal data and for processing of personal data cf. decision 2004/915/EC, decision 2001/497/EC and decision 2010/87/EU.

Each organization is data controller for the procession of these personal data, with reference to each Party's privacy policy and GDPR. The Consortium Agreement regulates the ownership and access to key knowledge and intellectual property, and developed issues with rights and confidentiality. DTU, as project manager, will advise the consortium on any issues.

Data will conform with Open Access policy and will be published as soon as quality assurance procedures are applied. Where possible, ECOTIP will promote the high standard of openness by deploying the Five Star Open Data Model. To score the maximum five stars, data must: (1) be available on the Web under an open licence, (2) be in the form of structured data, (3) be in a non-proprietary file format, (4) use URIs as its identifiers, (5) include links to other data sources.

ECOTIP is primarily generating and analysing environmental data. These data are neither sensitive/personal nor restricted data, however, the project intends to conduct sociological research as well. Interview data and transcriptions will contain names of informants and possibly also personal background and organizational or political affiliation. This information will not be made publicly accessible, but will be treated with confidentiality within the project. In the publically available analyses – e.g. reports and articles - ECOTIP will use one of the following two procedures:

- Anonymity as 'default' procedure for all interview data unless otherwise agreed (please see option 2 below). Anonymity will be secured in the (re)presentation of the interview persons' perspectives, opinions etc. This is particularly relevant in interviews on the subject of fishery governance, which can be politically sensitive. When citing or re-phrasing, there will not be reference to a specific name or title or location (by which the person could be easily identified),

but rather a broad professional category like fisher, hunter, manager, industry representative. Furthermore, personal information that could easily be used to identify the interview person will be omitted or even slightly revised (e.g. details on family relations, previous career etc.) (if inconsequential for the analysis itself). If names are to be included for the sake of writing style, a pseudonym will be provided.

- The interview persons will be offered an opportunity to claim credit (full name and community) for the knowledge they offer to ECOTIP. This is particularly relevant in Indigenous/Local Ecological Knowledge (LEK) studies in ECOTIP Task 4.4, where personal and collective intellectual ownership is a key issue. This option will be presented in the information sheet and is to be ticked-off as an additional item in the consent form.

8 DATA PUBLISHING & LICENSING

Data generated within the project will be published on the basis of scientific peer-reviewed papers with high probability of required open accessibility of underlying data. Additionally all data will be open to users on Creative Commons license after publication with no restriction.

ECOTIP will strongly commit to free and open data exchange and interoperability. First, data will immediately be exchanged within the project to enable performance of interdependent tasks. Second, the collected data and information will be made freely and openly available through portals operating both on the European (e.g. PANGAEA, SIOS, ICES) and North American (e.g. US NOAA NCEI, CCHDO) side of the Atlantic, thus ensuring they are validated and available for re-use by the wider community. All data can initially be deposited on the IO PAN Data Centre and replicated to other repositories. Each dataset or product released will have an assigned DOI number. This service will be offered both by the IO PAN Data Centre, and optionally also by Zenodo's ECOTIP community.

Data from hydrographic cruises conducted along regular GO-SHIP lines are recommended to be deposited in CCHDO (<https://cchdo.ucsd.edu/>). Surface ocean carbon data is typically submitted either to NOAA NCEI OCADS (<https://www.ncei.noaa.gov/products/ocean-carbon-data-system>) or directly to SOCAT (<https://www.socat.info/>) which now supports automatic data and metadata submissions. Majority of biological data on phytoplankton, zooplankton, biota composition and rate measurements, as well as paleoceanographic data will be deposited to PANGAEA. Data on fluxes and suspended particulate material from the Arctic Ocean is typically deposited in the NIRD Research Data Archive (<https://archive.norstore.no/>), available also through the Svalbard Integrated Arctic Earth Observing System (SIOS; <https://sios-svalbard.org/>). National Centre for Biotechnology Information (NCBI; <https://submit.ncbi.nlm.nih.gov/>) is most suited to accepting omics and proteomics type of data generated by the project. Other dissemination platforms might be used to provide free, open and FAIR access to other types of data, including fish surveys, and socio-economic data.

To maximize the possibility of data discovery and access to data and information using innovative and effective open science practices, metadata for all collected data will be provided in appropriate standards to key Pan-European and global ocean and marine data management infrastructures constituting the data and information product delivery backbone of the European Ocean Observing

System (EOOS) and GOOS. The domain of data integrators includes but is not limited to SeaDataNet (<https://www.seadatanet.org/>), EMODnet (<http://www.emodnet.eu/>) and Ocean Biogeographic Information System (OBIS; <http://iobis.org/>), the organizations which implement state-of-the-art cloud technology, and offer the use of accompanying software, products and reports to view and further analyse the data alongside other information found across the associated data repositories. Following the recommendations of the GEO Marine Biodiversity Observing Network (MBON), ECOTIP will engage with the European node of OBIS in particular to exploit the possibility of directly submitting or linking co-located biodiversity and environmental data and consistent metadata prepared according to one of the recommended standards.

Results of the project, according to Grant Agreement Article 26.1 are owned by the organization that generates them. However Project Partners are entitled to use them on the basis of joint ownership defined in GA Article 26.2. The dissemination of own and jointly owned results by one or several organizations, including but not restricted to publications and presentations, shall be governed by the procedure of Article 29.1 of the Grant Agreement subject to the following provisions. Prior notice of any planned publication shall be given to the other Parties at least 45 calendar days before the publication. Any other beneficiary may object within — unless agreed otherwise — 30 days of receiving notification, if it can show that its legitimate interests in relation to the results or background would be significantly harmed.

9 DATA ARCHIVING

Management of data will be performed by existing data centres in adherence to international accepted principles and practices, e.g. as defined by the ICSU World Data System (www.icsuwcs.org). Agreed quality control procedures will be used to curate the data, using flags and possible corrections. Long-term preservation of data will be ensured through the use of Global Earth Observation System of Systems common infrastructure as reference registry and mirroring data in different data infrastructures. Depending on the type of collected data, they can be permanently archived in national and international databases (e.g. PANGAEA, NCEI, ICES). For operational purposes, data collected and generated by the project can initially be hosted at IO PAN Data Centre and replicated to selected infrastructures. Costs of data curation and preservation have been included in the project budget.

Data remain on the same level of accessibility after project completion and termination. To assure this eternal stable state of data accessibility, project will use resources of IO PAN Data Centre and replicate data with other data systems, e.g. PANGAEA and SeaDataNet.

This far horizon strategy will assure long-term data accessibility, resilient for possible fault occurrences affecting one of the data systems, where data are stored. This accessibility is highly important for future environmental studies since environmental data, in particular data acquired through field campaigns, has a value which is hard to overestimate.

10 REFERENCES

ECOTIP Grant Agreement No 869383

ECOTIP Consortium Agreement, version 4.0.0, 2020-09-07

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>