
TANGO - Estimating Tipping points in habitability of ANtarctic benthic ecosystems under GLObal future climate change scenarios

A Data Management Plan created using DMPonline

Creator: Anton Van de Putte

Affiliation: Other

Template: DCC Template

ORCID iD: 0000-0003-0502-8101

ORCID iD: 0000-0002-9037-7623

ORCID iD: 0000-0001-9269-6487

ORCID iD: 0000-0003-2983-9523

Project abstract:

TANGO is a project funded by the Belgian Science Policy (belspo) that aims to develop a mechanistic framework to predict tipping points leading to regime shifts in coastal Antarctic ecosystems, with a focus on the benthos (organisms living in contact with the seafloor). The project framework will consist of the integration of different biological/ecological levels, from (1) individual metabolic responses, (2) constraints of environmental changes on species interactions (trophic ecology and space competition), (3) the dynamics of the communities and its consequences for carbon fluxes and finally (4) responses upscaled to ecosystem levels. Regime shifts and threshold dynamics from individuals to ecosystems will be investigated by empirical evidence making use of long-term datasets, field observations along natural gradients and experimental and modelling approaches in Antarctic coastal

ecosystems. We will focus on selected taxa representing the main functional traits in a coastal benthic community. Different size classes (meio- to megafauna), life strategies (pioneer to climax species) as well as different trophic groups will be covered.

ID: 84809

Start date: 15-01-2021

End date: 15-04-2024

Last modified: 24-09-2021

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

TANGO - Estimating Tipping points in habitability of ANtarctic benthic ecosystems under GLObal future climate change scenarios

Data Collection

What data will you collect or create?

- Sampling events will be the core around which all the data generated by this project will be structured. Sampling events include the time and location of an event (e.g. the cruise, or collection of a specimen), and can be hierarchically structured (sub events nested in parents events) in the DarwinCore Event format as promoted by OBIS and GBIF.
- Specimen will be collected during three planned Antarctic field campaigns through SCUBA diving, benthic coring or bottom trawling. This will result in species occurrence data, as well as physical specimens for later use. The occurrence data will be formatted as CSV files that are several KB to MB in size and will be structured as an occurrence extension of the DarwinCore Event format. As such, each occurrence record will contain information about the species, such as geographical coordinates, ecosystem types / vegetation types, substrate, date and time of collection, technical metadata (expedition, equipment used, name of the observer,...), or abundance or number of individuals, as well as links to any other data that was derived from the organism (see further). Concerning the physical specimen: some will be kept alive for later experiments, while others will be long-term preserved (e.g. through freezing or fixing in ethanol) in institute collections (RBINS, UGENT, ULB).
- Non-invasive specimen sampling will be done by imaging during Remotely Operated Vehicle (ROV) transects. This will generate 2D images and 3D habitat maps, as well as sonar maps of the study sites. The visual image data include high resolution photographs of living specimen or in-situ species communities (estimated size around 150-600MB per image, format 24-bit BMP or TIFF), 3D-models of specimen or in-situ communities (up to 1GB, format FBX, OBJ, MAX, 3DS, C4D) or film from remotely operated vehicles or dives of in-situ communities (> 16GB, format; MP4, MOV, STL, OBJ, FBX).
- Numeric environmental measurements, such as depth, temperature, conductivity, chemical measurements, etc. will be taken for explanatory parameters in statistical analyses. These measurements will be taken from ice coring, seawater sampling and sediment sampling by sediment traps anchored to ice, as well as during each ROV cruise. This data is formatted as CSV tables and is several KB to MB in size. This data will be reformatted in a DarwinCore extended Measurement or Fact extension for data archiving.
- Experimental measurements will be taken from organisms in the lab. This includes fatty acid profiles, stable isotopes (carbon, nitrogen and sulphur), biological/physiological traits (respiration rates, reproduction condition, resource allocation strategies) and their dependencies on selected environmental parameters (temperature, food quality and food quantity). These data will be

recorded in CSV spreadsheet tables and will be several KB to MB in size. This data will be reformatted in a DarwinCore extended Measurement or Fact extension for data archiving. This can be linked to the sampling event structure.

- In some cases, genetic barcode sequencing will be performed on target species. The DNA data is formatted in FASTA or FASTQ, and is accompanied by metadata in the MIxS standard (CSV format), and is typically ranging from around one KB (single sequence) to several GB (a full genome) in size.

How will the data be collected or created?

This project will require collecting and generating data to answer ecological and climate-related research questions on the current state of the Southern Ocean environment. These data are also needed to feed mathematical models that will characterize potential responses and vulnerabilities of ecosystems from the Southern Ocean. The Project partners have extensive experience in biodiversity studies and biological data management.

Before each campaign sampling strategies and allocation of samples among the Principal Investigators (PI) and project partners will be discussed. Organizational schemes and approaches to data formatting, recording, storage, and integration, as well as sample handling protocols, will be discussed in detail.

During the expeditions, data and metadata management will be tightly coupled. A structured and standardized numbering scheme for samples and specimens will be coordinated among all partners, and will feed into a central database where all samples, events and sub-events will be recorded. For each event, a close record of anticipated metadata will be kept in a shared database during the expedition, and will include instrument details, sample and experiment characteristics. This is the primary means of identifying, ordering, and cataloguing sampling activities during an expedition, and ensures traceability of samples across activities. The investigators will share this responsibility and designate an individual to maintain the database.

Documentation and Metadata

What documentation and metadata will accompany the data?

All datasets will be made available according to the FAIR principles. For this, they will be fully described with metadata explaining their contents, origin, methods, licensing and citation. To the largest extent possible data will be made available under a creative commons CC-0 or a CC-BY license.

During each expedition, great care will be taken in documenting important metadata information, including sampling method and device, collector, taxon identifier and instrument type and parameter settings, and sample processing and preservation.

A major component to the data documentation will be the central database where all samples, events and sub-events will be recorded using a structured and standardized numbering scheme. This database will hold all necessary metadata, such as date, time and location information, instrument details, or sample and experiment characteristics, and will be published on GBIF and OBIS, accompanied by project-level documentation

using the Ecological Metadata Language (EML) format. The database will be structured in the DarwinCore Event standard, which structures each datapoint around an event (e.g. the collection of a specimen). Each event is linked to information about the specific time, place and context of the event, and can be embedded in parent events or contain sub-events. In this framework, species occurrences, as well as imaging-, sequence-, environmental- or experiment data can be linked to using various extensions to the Event Core, such as the occurrence extension or the Extended Measurement or Fact extension. This ensures scalable and complete documentation of each data point in a widely used format and using controlled and standardized terminology. As such, each data point will be fully documented through the DarwinCore archive that can be openly accessed through the GBIF website. Literature links will inform the user on the technical specifications of some data where additional information is necessary.

Species occurrence data will be documented in the DarwinCore Event format, which includes multiple fields that provide additional information on the initial measurement (e.g. observer, type of observation...), as well as a reference to a project-level EML file. Nucleotide sequence data will be documented by formatting metadata in the Minimum Information about any (x) Sequence (MIxS) format. This (meta-)data standard was developed by the Genomics Standards Consortium and has been adopted by the International Nucleotide Sequence Database Collaboration (INSDC). Sequences alongside the MIxS information will be archived on INSDC. Data will also be made available in the Barcode of Life System (BOLD).

The documentation of additional post-expedition numeric data (i.e. analyses, experiments and their outcomes) will be discussed among the PIs and collaborators, as no broadly used metadata standards exist yet. A standardized coordinated scheme will be developed for this project and will include all valuable information on instrument types and settings, experimenters, lab protocols and any other important parameters that could influence the experiment or labwork.

All this information will be openly accessible through the websites of GBIF (www.gbif.org) and OBIS (www.obis.org), as well as the Antarctic-specific biodiversity.aq data portal.

Event data and biological data will be documented with the DarwinCore standard and vocabulary for events and species occurrences, which was developed by the Taxonomic Databases Working Group (TDWG). There, metadata is documented in Ecological Metadata Language (EML) file, which is a structured XML-type file that describes the project metadata, and can be created with the Integrated Publishing Toolkit (IPT) from the Global Biodiversity Information Facility (GBIF).

The MIxS standard, developed by the Genomics Standards Consortium in 2011 will be used to document the nucleotide sequence data, and will contain a wide array of information, including technical details on the sequencing itself, as well as information on the biological material or associated biotic / abiotic measurements.

Presently, there is no commonly accepted metadata standard for 2D and 3D images or lab protocols, experimental setups and results. Documentation with this data will be done through

linking with the project EML files, events and species occurrences. Peer-reviewed publication of the experiment outcomes and set-up will also ensure detailed documentation on these data will be publicly available.

Ethics and Legal Compliance

How will you manage any ethical issues?

The project will only collect public domain personal information to identify people. This includes authors of scientific papers, collectors or constructors of material and copyright holders. This is done to ensure correct authorship attribution or identification authority.

For sampling of organisms, National and International ethical guidelines will be followed.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

All data will be made as publicly available as possible. All data providers will be asked to provide their data either as CC-0 or CC-BY.

Storage and Backup

How will the data be stored and backed up during the research?

In addition to the use of the institution network research storage facilities during the project, we aim to move data into the public domain as soon as possible. This is usually after publication of the associated results, but can be sooner for some types of information, like the EML metadata or the basic sampling event information.

This is to make the data compliant to the FAIR principles. Species occurrence and event data will be archived on GBIF, OBIS and biodiversity.aq, while nucleotide sequences and environmental data thereof on one of the INSDC databases, such as ENA, GenBank or SRA. Environmental and experimental measurements that are linked to events or occurrences can in case of complex (e.g. long-term experiment time series data) or high volume data (2D and 3D images) be stored on Zenodo (www.zenodo.org, managed by CERN).

We have carefully selected the GBIF, OBIS, INSDC and Zenodo databases because they are open to the public, easily accessible, long-term and internationally funded, secure, well managed and backed-up, and operate at no cost to the project.

The RBINS and UGENT institution networks implement a backup system for data stored in the institution cloud. All researchers will be responsible for backup of their data on these institution cloud systems or personal hard drives.

In addition, both GBIF, OBIS, INSDC and Zenodo implement a full backup and recovery system to protect data against a loss, preserving the ability to recover data at all time.

How will you manage access and security?

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

All in sit data collected will be preserved

The RBINS and UGENT institution networks implement a backup system for data stored in the institution cloud. All researchers will be responsible for backup of their data on these institution cloud systems or personal hard drives.

In addition, both GBIF, OBIS, INSDC and Zenodo implement a full backup and recovery system to protect data against a loss, preserving the ability to recover data at all time.

What is the long-term preservation plan for the dataset?

By the end of the project, all data will be stored in the open databases discussed above (i.e. GBIF, OBIS, INSDC or Zenodo). All these databases have stable international funding and long-term commitments to the maintenance and upkeep of the infrastructures. This will ensure the data will be preserved on longer time scales (>10 years).

Data Sharing

How will you share the data?

All datasets will made available according to the FAIR principles. To the largest extent possible data will be made available under a creative commons CC-0 or a CC-BY license.

This is to make the data compliant to the FAIR principles. Species occurrence and event data will be archived on GBIF, OBIS and biodiversity.aq, while nucleotide sequences and environmental data thereof on one of the INSDC databases, such as ENA, GenBank or SRA. Environmental and experimental measurements that are linked to events or occurrences can in case of complex (e.g. long-term experiment time series data) or high volume data (2D and 3D images) be stored on Zenodo (www.zenodo.org, managed by CERN).

Are any restrictions on data sharing required?

All datasets will be made available according to the FAIR principles. To the largest extent possible data will be made available under a creative commons CC-0 or a CC-BY license.

Responsibilities and Resources

Who will be responsible for data management?

Each partner will be responsible for the correctness and completeness of their own data documentation.

Each partner will be responsible for the storage and backup of their own data.

The project partner at RBINS can provide technical support, when necessary. Dr. Anton Van de Putte will act as data protection officer.

The Principal Investigator (PI) bears the overall responsibility for updating & implementing this DMP

What resources will you require to deliver your plan?