ACEAS
Australian Centre for Excellence in Antarctic Science

ACEAS Data Management Plan

Sample plan 1

Access the **ACEAS Data Management Planning User Guide**

DMP version	Date modified	Modified by
v1	7 December 2023	Prashasti Singh

Institution	University of Tasmania

People

Researcher 1				Researcher 2			
Name	Singh, Prashasti			Name	Armbrecht, Linda		
ORCiD	0000-0002-3543-607X			ORCiD	0000-0002-1213-1257		
Email	Prashasti.singh@utas.edu.au			Email	Linda.armbrecht@utas.edu.au		
Affiliation 1	IMAS, University of Tasmania			Affiliation 1	IMAS, University of Tasmania		
Affiliation 2	Click or tap here to enter text.			Affiliation 2	Click or tap here to enter text.		
Position	Primary investigator	DMP creator	☒	Position	Co-investigator	DMP creator	☐

Researcher 3				Researcher 4			
Name	Bolch, Christopher			Name	Teixeira, Joao		
ORCiD	0000-0001-9047-2391			ORCiD	0000-0001-6417-4702		
Email	chris.bolch@utas.edu.au			Email	joao.teixeira@anu.edu.au		
Affiliation 1	IMAS, University of Tasmania			Affiliation 1	Australian National University		
Affiliation 2	Click or tap here to enter text.			Affiliation 2	Centre for Interdisciplinary Studies, University of Coimbra, Coimbra, Portugal		
Position	Co-investigator	DMP creator	☐	Position	Co-investigator	DMP creator	☐

Researcher 5				Researcher 6			
Name	Click or tap here to enter text.			Name	Click or tap here to enter text.		
ORCiD				ORCiD			
Email				Email			
Affiliation 1	Click or tap here to enter text.			Affiliation 1	Click or tap here to enter text.		
Affiliation 2	Click or tap here to enter text.			Affiliation 2	Click or tap here to enter text.		
Position	Choose an item.	DMP creator	☐	Position	Choose an item.	DMP creator	☐

Project Details

Project title	The composition and evolution of diatoms around Antarctica inferred from marine sedimentary ancient DNA		
DMP ID number	Click or tap here to enter text.	Ethics ID number	Click or tap here to enter text.
Project start	1/12/2021	Project end	1/05/2025
FOR code(s)	04050, 050101, 060411	SEO code(s)	960306, 960801, 960502
Keywords	sedimentary ancient DNA, Antarctica, palaeoenvironmental reconstruction, Fragilariopsis cylindrus, evolution, adaptation		
Project description	Using ancient DNA preserved in the seafloor (sedimentary ancient DNA or *sed*aDNA), my research project investigates how Antarctic diatoms have adapted to climate change in Antarctica through time. Diatoms are a group of phytoplankton that are very useful in paleoenvironmental reconstructions because of their high sensitivity to changing environmental conditions. Sediment cores collected off East and West Antarctica are being used to extract *sed*aDNA (from the Holocene, ~11,700 years ago) and assess the genetic adaptations of the key polar diatom species *Fragilariopsis cylindrus* to variations in climate over time. As this is a pioneer work using *sed*aDNA to produce data from diatoms specifically, I am also developing optimised techniques for effective diatom *sed*aDNA extraction and enrichment by hybridisation capture to generate authentic *sed*aDNA signals. This research will generate significant new knowledge about the response of Antarctica's most important marine primary producers to past environmental change, leading to improved predictions about their future adaptation to the ongoing climate change, ultimately guiding ecosystem models and conservation efforts for Antarctica.		

Project methods	Data in this study is generated in the form of large metagenomic Next Generation Sequencing datasets. For data analysis, interpretation, and overall quick processing of the datasets I have used specialized bioinformatics softwares [ADAPTERREMOVAL v. 2.1.7-foss-2016a software (Schubert, Lindgreen, & Orlando, 2016), Komplexity (Clarke *et al*., 2019) BBMAP version 37.36, FASTQC (version 0.11.5-Java-1.8.0_101, Babraham Bioinformatics) and MULTIQC (version 1.0.dev0; Ewels, Magnusson, Lundin, & Käller, 2016), MALT (Herbig *et al*., 2016), seqtk (version 1.3, https://github.com/lh3/seqtk) Genious Prime and MEGAN6 (version 6_15_1; Huson *et al*., 2016)]. Additionally, I have used R Studio (2023.09.0 Build 463) for statistical analysis, data interpretation and visualisation. The evaluation of dN/dS ratio in my study will be performed through custom-made scripts and R Studio will be used again for statistical analysis, data interpretation and visualisation. The Rosalind Server at Menzies College, which is a part of the Tasmanian Partnership for Advanced Computing (TPAC) is being used for high-performance computing, which is needed for the running and processing of sequencing data.

Data Storage Requirements and Restrictions			
File size	1 to 10 terabytes	File types	Raw sequencing data, FASTQ files and PDF Files
Data storage service(s) to be used	Sequencing data is generated in the form of FASTQ files are processed on the Rosalind Server of Menzies. Raw sequencing data is saved in a raw data folder only my primary supervisor, Dr. Linda Armbrecht can access on the Rosalind Server. Rosalind server is also used for cloud computing, virtualisation and storage of raw data. Approximately, 2-5 TB of space will be utilized for storage of raw data (current quota = 8.0 TB, used = 689 GB). Post-analysis of processed data, including interpretation of results, data is saved in the form of PDF files in separate folders on a hard drive. Personally identifiable information is also retained in the form of Microsoft Word/Excel formats in separate folders on a hard drive. To ensure that important data is protected in the case of a hardware or system-level failure, I have backed up the data in a hard drive. This backup is updated at each step of data processing. Backups of experiment information (extraction and library preparation) are stored in a shared folder with my supervisor on OneDrive. The final data that will be used in the thesis, along with the raw data, will be either uploaded on the IMAS Data Portal or on the NCBI Sequence Read Archive. A backup copy is also saved on a separate hard drive with my primary supervisor.		
Collaborator access to data *If your project includes collaborators from other institutions, describe how each person will be able to access the data. If individual collaborators do not have access to the data storage services listed above, describe the method(s) of data provision and the precautionary process(es) that will be used to protect sensitive data.*	N/A		

Personal or potentially identifiable content *If your data identify individuals at any stage of the research, you will need to follow the relevant federal, state, and institutional privacy requirements. Refer to the User Guide for additional information.*	N/A
Confidentiality and contractual obligations	As both IMAS Data Portal or the NCBI Sequence Read Archive are open-access repositories, the data generated in this research will be made publicly accessible.
Other sensitive information *If your data contain other sensitive information, e.g. locations of threatened/ endangered species, include their details here and describe how your storage selection will safeguard this information.*	N/A
Intellectual Property	Prashasti Singh According to rule 2.2 under 4.3 of IP policy.

Data Publishing		
Anticipated research data outputs *List the types of data output that you intend to produce (e.g. CSV, audiovisual, interactive online resource, software), and their provisional titles. If your project includes the collection of physical samples that are to be deposited into a collection, relevant location/storage information can be noted here.*	• CSV • FASTQ • PDF	
Published/archived via IMAS Portal	☒	Note: IMAS datasets should NOT be published via RDP
Published/archived via other service	Files are archived in the form of FASTQ and PDF files which are stored on hard drives. Sequence Read Archive (SRA) will be used for archiving the final analysis data of each chapter in the future, which is available through multiple cloud providers and NCBI servers and is the largest publicly available repository of high throughput sequencing data. The archive accepts data from all branches of life as well as metagenomic and environmental surveys. SRA stores raw sequencing data and alignment information to enhance reproducibility and facilitate new discoveries through data analysis.	
Not publicly available	☐	Click or tap here to enter text.
Embargo prior to publication	☐	(Embargoes should be no longer than 12 months)
Creative Commons licence	Choose an item.	

ACEAS
Australian Centre for Excellence in Antarctic Science

ACEAS Data Management Plan

Sample plan 2

Access the **ACEAS Data Management Planning User Guide**

DMP version	Date modified	Modified by
V1.0	06 Dec 2023	Green, David B.

Save file versions following the format: ACEAS Program 2 DMP – Smith, Jane - 20230927.docx

Institution	University of Tasmania

People

Researcher 1				Researcher 2			
Name	Green, David B			Name	Bestley, Sophie		
ORCiD	http://orcid.org/0000-0002-0346-3129			ORCiD	https://orcid.org/0000-0001-9342-669X		
Email	david.green@utas.edu.au			Email	sophie.bestley@utas.edu.au		
Affiliation 1	Institute for Marine and Antarctic Studies, University of Tasmania			Affiliation 1	Institute for Marine and Antarctic Studies, University of Tasmania		
Affiliation 2	Australian Centre for Excellence in Antarctic Science (ACEAS), University of Tasmania			Affiliation 2	Australian Antarctic Program Partnership, University of Tasmania		
Affiliation 3	Click or tap here to enter text.			Affiliation 3	Australian Centre for Excellence in Antarctic Science (ACEAS), University of Tasmania		
Position	Primary investigator	DMP creator	☒	Position	Primary investigator	DMP creator	☐

Researcher 3				Researcher 4			
Name	McMahon, Clive R			Name	Hindell, Mark A		
ORCiD	https://orcid.org/0000-0001-5241-8917			ORCiD	https://orcid.org/0000-0002-7823-7185		
Email	Clive.McMahon@utas.edu.au			Email	mark.hindell@utas.edu.au		
Affiliation 1	IMOS Animal Tagging, Sydney Institute of Marine Science, Mosman, 2088, NSW, Aust.			Affiliation 1	Institute for Marine and Antarctic Studies, University of Tasmania		
Affiliation 2	Institute for Marine and Antarctic Studies, University of Tasmania			Affiliation 2	Click or tap here to enter text.		
Affiliation 3	Australian Centre for Excellence in Antarctic Science (ACEAS), University of Tasmania			Affiliation 3	Click or tap here to enter text.		
Position	Primary investigator	DMP creator	☐	Position	Co-investigator	DMP creator	☐

Researcher 5				Researcher 6			
Name	Lea, Mary-Anne			Name	Click or tap here to enter text.		
ORCiD	https://orcid.org/0000-0001-8318-9299			ORCiD			
Email	maryanne.lea@utas.edu.au			Email			
Affiliation 1	Institute for Marine and Antarctic Studies, University of Tasmania			Affiliation 1	Click or tap here to enter text.		
Affiliation 2	Click or tap here to enter text.			Affiliation 2	Click or tap here to enter text.		
Position	Co-investigator	DMP creator	☐	Position	Choose an item.	DMP creator	☐

Project Details

Project title	Monitoring phocid seals in the Southern Ocean		
DMP ID number	Click or tap here to enter text.	Ethics ID number	28182
Project start	2/05/2022	Project end	31/12/2025
FOR code(s)	3102, 3103, 4102	SEO code(s)	1804, 1099
Keywords			
Project description	Quantifying the relationships between Southern Ocean predators, their prey, and the underlying environment is a critical step in understanding how ecosystems could respond to future environmental change. To investigate these relationships, we need first to concurrently measure predator foraging behaviour, the distribution of their prey, and physical water properties of the environment in which they occur. Obtaining such measurements is difficult, particularly for prey (mid-trophic level organisms such as mesopelagic fish, cephalopods and crustaceans), and remains challenging in the remote Southern Ocean, particularly at scales relevant for resolving ecological interactions. Marine predator biotelemetry can help us overcome these challenges. Given predator foraging is strongly related to the distribution of prey, we can infer prey distribution from predator movements, and can obtain simultaneous measures of the environment (e.g. temperature and primary productivity) using integrated sensors onboard animal-borne loggers (conductivity-temperature-depth satellite relay devices, CTD-SRDLs). (Contd.)		

	(Contd.) This project links foraging behaviour and performance of Antarctic and sub-Antarctic phocids to *in situ* observations of prey (inferred from foraging behaviour), and the underlying biophysical environment. In so doing, this work aims to generate valuable insights into the structure and function of Southern Ocean ecosystems, and important understanding of how predator foraging is influenced by prey and environmental processes. In addition, this project aims to expand the spatial coverage of data in East Antarctica and the Indian Ocean, which have received relatively little attention in recent years, filling gaps in data poor regional ecosystems.
Project methods	**Data collection** Location and dive behaviour of seals, along with measurements of in situ biophysical properties are collected using (fluorometry-)Conductivity-Temperature-Depth Satellite Relay Device Logger ((f)CTD-SRDL) attached to the head of Antarctic and Subantarctic seals (primarily southern elephant seals and Weddell seals). (f)CTD-SRDLs are deployed at sites across the Indian and Pacific sectors of the Southern Ocean and provide valuable oceanographic and ecological data for poorly sampled regions of East Antarctica. *fCTD-SRDL deployment* Devices are deployed following standard operating procedures approved by the UTAS ethics committee (see above for ethics project number). For device attachment, seals are chemically sedated, weighed, and measured, and the tag glued to the pelage on the seal's head. The CTD-SRDLs remain on the seals until they either fall off at sea or during the annual moult. While at sea, CTD-SRDLs provide 2-15 ARGOS satellite location estimates per day and a random sample of dive profiles. Individual dives are summarized on board the device and aggregated into five time-depth segments, separated using a broken-stick algorithm which identified the four inflection points that best represent dive profile shape. Temperature, conductivity, and salinity profiles are also transmitted in an abstracted form with 17 inflection points determined in the same way as the dive profiles. Along with these summarised dive and CTD profiles, CTD-SRDLs also relay measurements of maximum dive depth, dive duration, and post-dive surface interval. **Data processing** • *Path analysis*: To estimate the most likely path and dive locations for individual seals, we fitted the ARGOS satellite location estimates and associated errors within a state-space model with a correlated random walk structure, using the R-package "aniMotum" (https://github.com/ianjonsen/aniMotum; Jonsen *et al*., 2022), using the "seaTracks" wrapper code (https://github.com/davo-b-green/seaTracks). • *Dive metric analysis*: Seal dive metrics (dive residual, surface residual, hunting time, ascent rate, descent rate, drift rate and delta drift rate) are computed using code available in the "seaTracks" GitHub repository (https://github.com/davo-b-green/seaTracks). • *CTD data processing*: All summarised CTD profiles are processed and QC controlled by the IMOS animal tagging facility following methods detailed and published in Roquet *et al*. 2011 (https://doi.org/10.1175/2010JTECHO801.1).

Data Storage Requirements and Restrictions			
File size	Up to 1 terabyte	File types	CSV, netCDF
Data storage service(s) to be used	All QC processed animal tracking data including movement and CTD profiles are stored and freely available on the AODN Data Portal. All data produced during specific research projects is to be stored on the IMAS Data Portal.		
Collaborator access to data *If your project includes collaborators from other institutions, describe how each person will be able to access the data. If individual collaborators do not have access to the data storage services listed above, describe the method(s) of data provision and the precautionary process(es) that will be used to protect sensitive data.*	Data outputs from model runs can be accessed by collaborators via the AODN or IMAS Data Portals.		
Personal or potentially identifiable content *If your data identify individuals at any stage of the research, you will need to follow the relevant federal, state, and institutional privacy requirements. Refer to the User Guide for additional information.*	No personal or sensitive content is produced or used.		
Confidentiality and contractual obligations	None		
Other sensitive information *If your data contain other sensitive information, e.g. locations of threatened/ endangered species, include their details here and describe how your storage selection will safeguard this information.*	None		
Intellectual Property	Click or tap here to enter text.		

Data Publishing		
Anticipated research data outputs *List the types of data output that you intend to produce (e.g. CSV, audiovisual, interactive online resource, software), and their provisional titles. If your project includes the collection of physical samples that are to be deposited into a collection, relevant location/storage information can be noted here.*	• Products made available on the AODN are in the form of NetCDFs Provisional titles: ○ Satellite Relay Tagging Program - Southern Ocean - Quality Controlled CTD Profiles ○ IMOS - Animal Tracking Facility - Satellite Relay Tagging Program - Near real-time data with quality-controlled locations. • Summarised datasets used in specific research studies will be made available as CSVs on the IMAS Data Portal Provisional title: ○ Daily summarised estimates of seal location, dive effort, and foraging success.	
Published/archived via IMAS Portal	☒	Note: IMAS datasets should NOT be published via RDP
Published/archived via other service	Click or tap here to enter text.	
Not publicly available	☐	Click or tap here to enter text.
Embargo prior to publication	☐	(Embargoes should be no longer than 12 months)
Creative Commons licence	Attribution 4.0 International (CC BY 4.0)	

ACEAS
Australian Centre for Excellence in Antarctic Science

ACEAS Data Management Plan

Sample plan 3

(for more complex project methodologies)

Access the ACEAS Data Management Planning User Guide

DMP version	Date modified	Modified by
Draft (20230601)	01 June 2023	Jane Smith
v1 (20230602)	02 June 2023	Jane Smith & Bob Jones
v2 (20230723)	23 August 2023	Jane Smith

Save file versions following the format: 'ACEAS Program 2 DMP – Smith, Jane - 20230927.docx'
Include version history using the example provided above.
Email completed DMP to: harko.werkman@utas.edu.au

Institution	University of Tasmania

People

Researcher 1				Researcher 2			
Name	Smith, Jane			Name	Jones, Bob		
ORCiD	0000-0001-2345-6789			ORCiD	0000-0002-4681-3579		
Email	jane.smith001@utas.edu.au			Email	bob.jones002@utas.edu.au		
Affiliation 1	IMAS, University of Tasmania			Affiliation 1	IMAS, University of Tasmania		
Affiliation 2	Geography & Spatial Studies, University of Tasmania			Affiliation 2	Research School of Earth Sciences, Australian National University		
Position	Primary investigator	DMP creator	☒	Position	Primary investigator	DMP creator	☒
Researcher 3				**Researcher 4**			
Name	Hutton, James			Name	Click or tap here to enter text.		
ORCiD	0000-0001-3579-2468			ORCiD			
Email	james.hutton03@anu.edu.au			Email			
Affiliation 1	Research School of Earth Sciences, Australian National University			Affiliation 1	Click or tap here to enter text.		
Affiliation 2	School of GeoSciences, University of Edinburgh			Affiliation 2	Click or tap here to enter text.		
Position	Co-investigator	DMP creator	☐	Position	Choose an item.	DMP creator	☐
Researcher 5				**Researcher 6**			
Name	Click or tap here to enter text.			Name	Click or tap here to enter text.		
ORCiD				ORCiD			
Email				Email			
Affiliation 1	Click or tap here to enter text.			Affiliation 1	Click or tap here to enter text.		
Affiliation 2	Click or tap here to enter text.			Affiliation 2	Click or tap here to enter text.		
Position	Choose an item.	DMP creator	☐	Position	Choose an item.	DMP creator	☐

Project Details

Project title	Investigation of soil formation and ecosystem development on the South Shetland Islands in response to glacial retreat		
DMP ID number	Not using DMPTool	Ethics ID number	31245
Project start	4/12/2023	Project end	31/12/2027
FOR code(s)	310206, 310302, 310509, 310703, 310704, 319902, 410102, 410603	SEO code(s)	180403, 180605, 180606, 190102, 190503, 280111
Keywords	climate change, glacial retreat, Antarctic ecosystems, soil formation, microbial communities, DNA barcoding, metagenomics, ecological succession, South Shetland Islands, plant colonization, animal colonization, biodiversity, environmental monitoring, soil properties, next-generation sequencing, polar research, ecosystem dynamics		
Project description	This project will investigate processes of soil formation on the South Shetland Islands, with an emphasis on changes resulting from glacial retreat driven by contemporary climate change, with a view to understanding how these changes influence the rates and diversities of plant and animal colonisation. The study will examine the connections between retreating glaciers and emerging terrestrial ecosystems, focusing on how newly exposed substrates evolve into fertile soils, and on the subsequent establishment of flora and fauna. By documenting these ecological developments, the research will provide crucial insights into the early stages of ecosystem succession in polar environments. The scientific significance of this project lies in its potential to enhance understanding of the fundamental biotic and abiotic processes governing soil development and biological colonisation in newly deglaciated areas. These insights are essential for predicting how Antarctic ecosystems will respond to ongoing and future climatic shifts. In addition, this research will contribute to broader ecological theories regarding succession and habitat formation in extreme environments. The findings will be beneficial in formulating conservation strategies, as they will inform efforts to protect emerging ecosystems and to manage existing biodiversity in the face of rapid environmental change.		

Project methods	**Data Collection** Field sampling will be conducted during the Antarctic summer when accessibility is optimal. Transects will be established from the current glacier termini to progressively older soils to capture a temporal gradient of soil development. Soil samples will be collected at varying depths (0-10 cm, 10-20 cm, and 20-30 cm) using a stainless steel corer. The number and locations of the transects will be determined during the first expeditions to the islands. Each sampling site will be documented with GPS coordinates so that sites can be monitored in future expeditions, and relevant environmental parameters (e.g., temperature, moisture, aspect, exposure, and vegetation cover) will be recorded. Plant and animal species present at each site will be catalogued to assess colonization rates and biodiversity. Soil samples will be analysed to determine physical, chemical, and biological properties. Particle size distribution will be measured using laser diffraction, and soil organic carbon, phosphorus, and nitrogen content will be quantified using elemental analysers. Soil pH and electrical conductivity will be measured using standard potentiometric and conductometric methods, respectively. Plant and animal samples will be identified using morphological and molecular techniques, including DNA barcoding for species verification. For plants, the chloroplast gene regions rbcL and matK will be targeted, and for animals the mitochondrial cytochrome C oxidase I (COI) gene will be used. An Illumina MiSeq high-throughput sequencing platform will be used to generate barcode sequences. **Data Processing and Analysis** Initial quality assessment and trimming of raw sequence reads will be performed using TRIMMOMATIC (version 0.39; Bolger, Lohse, & Usadel, 2014). The filtered sequences will then be aligned to reference databases using BLAST (Basic Local Alignment Search Tool; version 2.10.1; Altschul *et al*., 1990) to ascertain taxonomic identities. The resulting data will be compiled and visualized using MEGAN (version 6_19_9; Huson *et al*., 2016), allowing for the construction of detailed phylogenetic trees and biodiversity analyses. Microbial community composition will be determined through identification of the diversity and abundance of microbial taxa) will be assessed through large metagenomic Next Generation Sequencing of samples. For data analysis, interpretation, and processing of the datasets several specialized bioinformatics software packages will be used [ADAPTERREMOVAL v. 2.1.7-foss-2016a software (Schubert, *et al*., 2016), Komplexity (Clarke *et al*., 2019) BBMAP version 37.36, FASTQC (version 0.11.5-Java-1.8.0_101, Babraham Bioinformatics), MULTIQC (version 1.0.dev0; Ewels, Magnusson, Lundin, & Käller, 2016), seqtk (version 1.3, https://github.com/lh3/seqtk), MALT (Herbig *et al*., 2016), Genious Prime and MEGAN6 (version 6_15_1; Huson *et al*., 2016)]. Sequence clustering and Operational Taxonomic Unit (OTU) picking will be carried out using QIIME2 (version 2020.6; Bolyen *et al*., 2019), which facilitates the comparison of community compositions across different samples. To ensure robustness in taxonomic assignment, the BOLD Systems database (Barcode of Life Data System; Ratnasingham & Hebert, 2007) will be used as an additional reference for species identification. Data visualization and statistical analyses of species diversity and abundance will be conducted using R (4.0.3) with relevant packages such as 'phyloseq' and 'ggplot2'. Statistical analyses will be used to identify patterns and correlations between soil properties and ecological successions. Multivariate statistical techniques, including principal component analysis (PCA) and redundancy analysis (RDA), will be used to discern the main factors driving soil formation and biotic colonization. Generalized linear models (GLMs) and mixed-effects models will be employed to assess the influence of environmental variables on species diversity and abundance. The statistical software R will be used for most analyses, with packages such as ‘vegan’ for community ecology and ‘lme4’ for mixed models. Additionally, geographic information system (GIS) software, such as QGIS, will be utilized for spatial analysis and visualization of the data. The high-performance computing needed for the running and processing of sequencing data will be conducted on TPAC’s (Tasmanian Partnership for Advanced Computing) Rosalind Server at Menzies College. All other high-performance computing will be conducted on an Intel Xeon or an AMD Ryzen Threadripper (yet to be decided) with 64 GB of RAM and large-capacity solid-state drives.

Data Storage Requirements and Restrictions			
File size	1 to 10 terabytes	File types	.xlsx, .shp, .kml, netCDF, .jpg, .png, .tif, .mp4, .fastq/.fasta, .R/.py
Data storage service(s) to be used	Field data will be collected using instruments including GPS units, digital cameras, and data loggers. These will be equipped with SD cards and internal storage capacities typically ranging from 32 GB to 256 GB, sufficient for daily data collection, for up to several years if necessary. To ensure data safety, a multi-tiered backup protocol will be implemented. During expeditions data from the portable field instruments will, whenever possible, first be backed up daily onto two separate field-ready external solid-state drives (SSDs), for example the Samsung T7 (500 GB to 2 TB) or the SanDisk Extreme Pro (1 TB to 2 TB), with one drive held by each of the two primary investigators. For any initial data processing and preliminary analysis in the field, and to ensure redundancy in storage, high-end laptops with robust processing power and extended memory (e.g., Dell XPS or MacBook Pro with ~2 TB) will be used. For ease of location a spreadsheet cataloguing the sizes, locations, and names of files acquired in the field will be maintained. After the expeditions are complete this file will be extended to include files generated from processing, analyses, and publication of the data. At the conclusion of the expedition raw data collected from the field, as well as additional data subsequently derived from the processing/analyses of the field samples, will transferred to the UTAS OneDrive as the primary storage location.		
Collaborator access to data *If your project includes collaborators from other institutions, describe how each person will be able to access the data. If individual collaborators do not have access to the data storage services listed above, describe the method(s) of data provision and the precautionary process(es) that will be used to protect sensitive data.*	Most collaborator access with be mediated through the UTAS OneDrive system. Where large working datasets cannot be hosted on OneDrive we will discuss access to the Research Data Storage Infrastructure with the ACEAS data manager: the RDSI has the capacity to allow external collaborator access.		
Personal or potentially identifiable content *If your data identify individuals at any stage of the research, you will need to follow the relevant federal, state, and institutional privacy requirements. Refer to the User Guide for additional information.*	This project will produce no personal or person-identifying data.		
Confidentiality and contractual obligations	There are no confidentiality obligations associated with this project. There is a funder obligation attached to the ARC grant that stipulates that all data be made open access, and to this end the data will be published under a Creative Commons CC BY 4.0 licence.		
Other sensitive information *If your data contain other sensitive information, e.g. locations of threatened/ endangered species, include their details here and describe how your storage selection will safeguard this information.*	As a result of the remote nature and unique climatic conditions of the South Shetland Islands, there is significant likelihood of identifying new species that are likely to be rare, and endemic to the islands. Species ID and location data will be restricted to only the researchers in the initial stages of the project. Once a comprehensive understanding of the species present is established, this section of the DMP will be revisited and updated to reflect the international best practice in safeguarding habitat descriptions and location data of threatened/endangered species.		

Intellectual Property	IP belongs to the University of Tasmania as stipulated under Section 2.1 of the UTAS Intellectual Property Policy.

Data Publishing

Anticipated research data outputs *List the types of data output that you intend to produce (e.g. CSV, audiovisual, interactive online resource, software), and their provisional titles. If your project includes the collection of physical samples that are to be deposited into a collection, relevant location/storage information can be noted here.*	**Data Records** 1. Metagenomic sequencing data from soil samples of the South Shetland Islands File types: .fastq, .fasta, .txt (metadata), .csv (sample information) 2. DNA barcoding data of plant and animal species colonizing newly exposed soils in the Antarctic File types: .fastq, .fasta, .csv (sample metadata), .txt (species identification) 3. Environmental and soil property data from glacial retreat zones on the South Shetland Islands File types: .xlsx, .csv, .txt (metadata), .netCDF (environmental data) 4. Microbial community profiles and diversity indices from Antarctic soil samples File types: .csv (OTU tables), .biom (biological observation matrix), .xlsx, .txt (metadata) 5. Geospatial data of sampling sites on the South Shetland Islands File types: .shp (shapefiles), .kml (Keyhole Markup Language), .geojson (geospatial data), .csv (coordinates and site metadata). **Physical samples** Soil samples collected from the field will be barcoded for identification and tracking purposes, and stored in the -20 °C freezer store (room 216B) on level 2 of the Castray Esp building, or in the -80 °C freezer (#17) on the same floor. A .csv file cataloguing the sample IDs and locations will be included in the data record 'Geospatial data of sampling sites on the South Shetland Islands.'	
Published/archived external to IMAS	All genetic data generated from this study, including metagenomic sequences and DNA barcoding data, will be deposited in internationally recognised and accessible repositories to ensure broad availability and compliance with data sharing standards. Metagenomic sequence data will be submitted to the NCBI Sequence Read Archive (SRA). Additionally, assembled and annotated metagenomic datasets will be deposited in the MG-RAST (Metagenomics Rapid Annotations using Subsystems Technology) database, which facilitates community access and offers powerful analysis tools. DNA barcoding sequences will be archived in the Barcode of Life Data System (BOLD), which is specifically designed for the storage, analysis, and retrieval of barcode sequences. For supplementary datasets, including metadata and processed data, the Dryad Digital Repository will be used to ensure that all associated information is easily accessible and citable. Metadata records and links to these deposited genetic data will be submitted to the IMAS Data Portal. All non-genetic data will be published via the IMAS Data Portal.	
Published/archived via IMAS Portal	☒	Note: IMAS datasets should NOT be published via RDP
Embargo prior to publication	☐	(Embargoes should be no longer than 12 months)
Not publicly available	☐	Click or tap here to enter text.
Creative Commons licence	Attribution 4.0 International (CC BY 4.0)	