Data Management Plan

Revision history

Date	Author	Description of change
2020-04-01	S. Barbosa Y. Karimova	First draft
2020-06-05	S. Barbosa	Revised draft, including feedback from J. Moutinho (AIR Centre), M. Martins (Univ Minho), S. Ramos (CIIMAR)
2020-11-23	S. Barbosa	Version 1
2021-03-24	N. Dias G. Amaral S. Barbosa	Revised version (version 2), including update of Table 2 and addition of Table 3
2021-07-05	S. Barbosa Y Karimova	Version 3 (abbreviations list updated, added section 1.2, update of sections 2.3, 2.4 and 2.5)
2021-9-24	S. Barbosa Y Karimova	Update of sections 2.3, 2.4 and 2.5 (version 4)
2021-12-22	S. Barbosa	Update of sections 2.3, 2.4, 2.5, 2.6 (version 5)
2023-11-15	S. Barbosa	Update of sections 1, 2.4, 2.44, 2.5 (version 6)

Abbreviations

Abbreviation	Description
CTD	Conductivity-Temperature-Depth
dd	Day (01-31)
GNSS	Global Navigation Satellite System
HH	Hour (00-23)
MM	Minute (00-59)
mm	Month (01-12)
NMEA	National Marine Electronics Association Format
RINEX	Receiver Independent Exchange Format
SS	Second (00-59)
yyyy	Year

Table of contents

1. Introduction

The SAIL Data Management Plan (DMP) is a document detailing the management of all the data from Project SAIL - **S**pace-**A**tmosphere-Ocean **I**nteractions in the marine boundary **L**ayer.

The SAIL DMP will be monitored and updated during the whole project and periodically revised and updated according to project changes.

Project SAIL aims to improve the scientific understanding of the marine boundary layer by means of a unique monitoring campaign on board the iconic Portuguese tall ship NRP Sagres during its 2020 circumnavigation expedition, planned to start on January 5th 2020 (Fig. 1). The campaign will enable the measurement of the atmospheric electric field over the ocean, and to study space-driven interactions via the detailed monitoring of GNSS signals, cosmic radiation, environmental radioactivity and atmospheric ionization. The atmospheric measurements will be complemented by the collection of fish samples and by underwater monitoring of the ocean state (temperature, conductivity, dissolved oxygen, pH, spectral radiance), providing unique data for the detailed study of ocean-atmosphere fluxes and surface-atmosphere interactions.

Keywords: marine boundary layer; cosmic radiation; atmospheric ionisation; electric field; aerosols; climate; acoustic noise; ocean health;

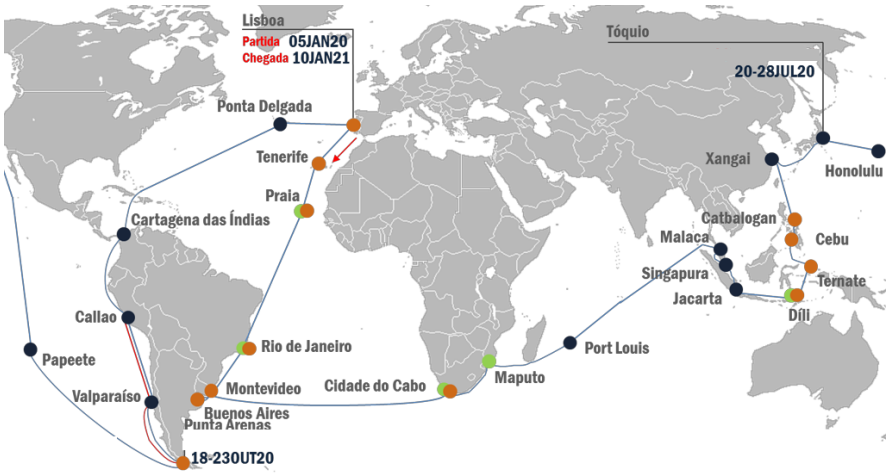

Figure 1: NRP Sagres planned circumnavigation expedition.

- Principal Investigator (PI): Susana Barbosa

ORCID iD: 0000-0003-2198-3715. Email: susana.a.barbosa@inesctec.pt

- Co-Principal Investigator (co-PI): António Manuel Maurício Camilo

- Chief Engineer (CE): Eduardo Alexandre Pereira da Silva

The SAIL management team is composed by the PI, the co-PI and the CE.

Project SAIL is divided into three main stages:

i) an initial step of planning and preparation of the campaign (October-December 2019);

ii) the monitoring campaign during the circumnavigation (January 2020 - May 2020);

iii) curation and analysis of the campaign data (May 2020 -).

In the **first stage** of the project a multi-parametric sensor infrastructure is installed onboard the NRP Sagres. A dedicated geo-referencing and temporal synchronization network, based on the Global Navigation Satellite System (GNSS) is also installed, along with a triple-antenna configuration capable of sensing the ship's movements in six-degrees-of-freedom. It provides an accurate time reference for synchronization and also allows the ship trajectory to be recorded for posterior analysis of the dynamic forces endured by the ship.

For each component of the monitoring system, specific procedures are detailed in an operation manual defining in detail eventual intervention from the crew on board for tasks such as tow-fish recovering, routine checks, and minor maintenance works like cleaning sensing surfaces or performing periodical visual checks on the equipment.

Software was developed for real-time interaction with the sensors in the monitoring system- Furthermore a diagnostic tool was developed and made available to the NRP Sagres crew, in order to enable routine brief checks on the status of the monitoring system and data collection, and eventually trigger human intervention and troubleshooting.

The planning stage also included the organisation of a training workshop aiming to enable the NRP crew to follow specific procedures for collection and storage of fish samples and to ensure preservation of the samples for further laboratory analysis after the trip.

The **second stage** comprises the collection of atmospheric and oceanographic measurements, as well as fish samples and ancillary information (navy records, system logs):

I. GNSS and atmospheric measurements (electric field, gamma radiation, visibility, ions, solar and cosmic radiation), every 1-second. All the sensors will transmit the data to the onboard PC with the corresponding time-stamp provided by the GNSS system.

II. underwater measurements from a tow-fish deployed by the Sagres crew for collection of underwater observations.

III. fish samples from opportunistic fish collection by the NRP Sagres crew,properly stored on board at -20°C for posterior laboratory analyses by CIIMAR

IV. meteorological information from the ship's records

Intermediate collection of measurements was performed whenever possible (e.g. after each leg of the trip, when the ship is docked) for a preliminary inspection and check of the data.

A summary of the measured parameters is displayed in Table 1.

Physical data	Instrument for collection	Observations
Atmospheric Electric Field	CS110 (Campbell Scientific)	sensor at the high position in the mast (starboard), denoted by E1
Atmospheric Electric Field	CS110 (Campbell Scientific)	sensor at the low position in the mast (port), denoted by E2
Total gamma radiation	NaI(Tl) scintillator (Scionix)	sensor next to E1
Cluster Ions	Cluster Ions Counter (CIC, Airel)	sensor next to E2
Shortwave radiation	SP510 (incoming) + SP610 (outgoing) , Apogee Instruments	sensors next to E1
Visibility	SWS050 (Biral)	sensor at the high position in the mast (port)
Cosmic radiation	SN005 (University of Oxford)	indoor sensor , next to the SAIL computer system
Global Navigation Satellite System	Antennas HX-CVX600A	antennas at the high position in the mast (port + starboard)
Ocean data (tow fish)	Tow fish system including Ocean Seven 310 multiparameter CTD (Idronaut) and Trilux multiparameter sensor (ctg Ltd)	occasional deployment (pressure, temperature, depth, O2, chlorophyll, pH, conductivity, salinity, turbidity acoustic noise, spectral radiance)
Fishing data	Fishing	fish samples
Meteorological data	Navy meteorological records	spreadsheet file created from meteorological records in paper forms ("Registo de missões")

Table 1. Measured parameters.

Workshops on selected topics related to the project theme were organised by the AIR Centre.

The second stage of the project was interrupted in May 2020 as the circumnavigation expedition of the ship was cancelled due and the ship called back to Lisbon due to the COVID pandemic. The NRP Sagres arrived at Cape Town (South Africa) on March 24th, as initially planned, but instead of continuing the circumnavigation from there, the ship returned immediately to Lisbon, with only a small stop in Mindelo (Cape Verde) for an emergency repair, arriving in Lisbon on May 9th. Figure 2 shows the trajectory of the ship. The voids in the ship's track correspond to data gaps due to system/instrumental malfunction and consequent data loss.

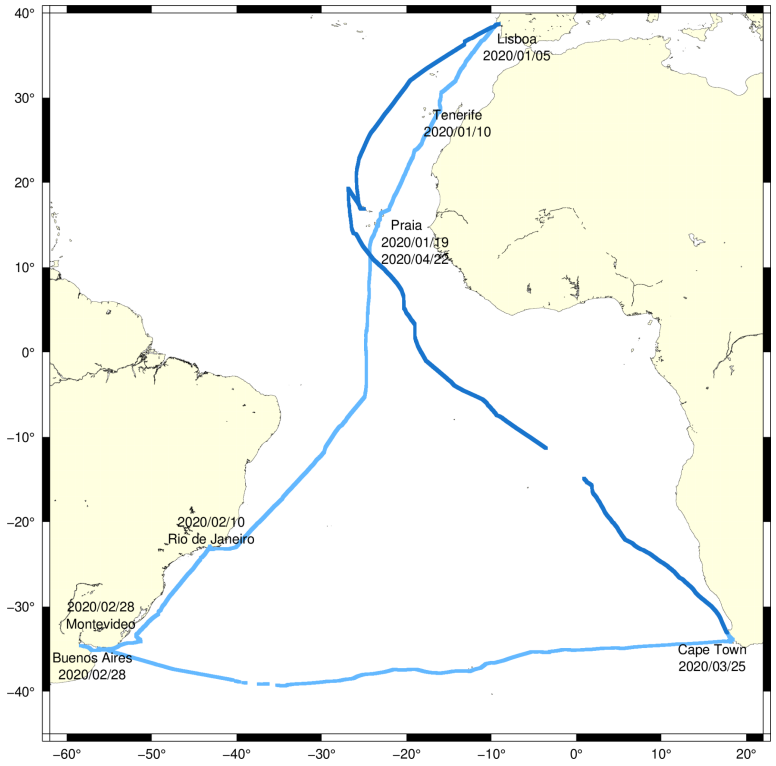

Figure 2. NRP Sagres expedition January-May 2020.

The SAIL project was naturally affected by the interruption of the circumnavigation expedition. For most of the rest of 2020, and also 2021, the ship remained docked at Base Naval de Lisboa, with only occasional short trips in Portuguese waters. Despite the obvious limitations, the activities corresponding to step I of the second stage of the SAIL project corresponding to the automatic acquisition of data, continued at the pier in Lisbon, and also whenever the ship set sail. Thus in addition to the data collected over the Atlantic in the initial part of the circumnavigation expedition (from January to May 2020), data were also collected when the ship was docked, in Lisbon, and during small trips near the Portuguese coast, including the 2021 summer trips between Portugal mainland and the Madeira and Azores islands. The data from these Sagres campaigns (other than the initial expedition) as well as the data collected onshore are handled separately.

The **third stage** of the project includes the final curation of all the data and its publication.

2. Data

2.1. Types of data

Different types of data will be collected, generated and processed in the project, including

- **Project management data**: data used internally in the management of the project, including meeting minutes, protocols, project reports, photos, data collected on public events organized by the project,...
- **Campaign data**: raw monitoring data, meteorological data collected by the ship crew, fish samples data;
- **Derived data**: data resulting from the processing and analysis of data collected in the project, such as processed data and curated databases, maps, plots,...

No data subjects are involved in the project.

Project management data are processed internally and are not passed on to third parties outside of the project except funding entities as may be required by any applicable reporting obligations.

All the other types of data do not include any sensitive or personal information.

2.2. Documentation and metadata

Project documents include:

- SAIL OpsCon – System Operational Concept (ISO/IEC 2011);

- Project proposal - FCT (April 2019);

- Project proposal - Fundo Ambiental (February 2020);

- SAIL communication plan;

- Manuals of instruments & sensors;

- Amaral, Guilherme, & Dias, Nuno. (2021, February 8). SAIL campaign - Technical report on Sensor Data correction. Zenodo. http://doi.org/10.5281/zenodo.4518865

- Ferreira, António. (2021, January 18). SAIL campaign - Technical report on GNSS Post-processing. Zenodo. http://doi.org/10.5281/zenodo.4447619

Metadata will be created based on generic (Dublin Core) or domain standardised metadata (Data Documentation Initiative, INSPIRE) when possible.

To improve the fit for re-use, detailed metadata, e.g. on the temporal and spatial resolutions, and the method used in the collection and analysis of the data will be described. Moreover, the metadata about geographic location will be obtained from the GNSS information.Only well-established and sustainable file formats are used for all the data in the project in order to enable interoperability of data and standardisation of data products.

2.3. Data collection

Data collection is performed in two distinct ways:

- by human observation and reading of sensors (thermometers, anemometers, rulers,...), requiring active human intervention.
- autonomously, in a continuous stand-alone mode - automatic data collection from sensors and instruments with no need of human intervention.

The data collected in a non-automatic way by the navy staff are of two distinct types:

- meteorological data, typically taken at intervals of 1 hour each day by the navy staff, and denoted with the infix **MT**
- fishing log and information on fish samples, taken sporadically (whenever a fish is caught), and denoted with the infix **FI**

The data collected without human intervention include diverse types of parameters. The naming conventions adopted for the data acquired continuously are displayed in Table 2.

infix	measurements
E1	Electric field from E1 sensor
E2	Electric field from E2 sensor
CI	Cluster ions
CR	Cosmic radiation
GA	Gamma radiation
NS	Global Navigation Satellite System (GNSS)
GNSS_NMEA	GNNS - NMEA standard
GNSS_RINEX	GNSS - RINEX format
SL	Solar radiation
TF_CTD	Tow fish measurements - CTD
VI	Visibility

Table 2: naming conventions for measured parameters.

The automatic data collection comprises 3 distinct collection activities, covering distinct periods (Table 3):

- Data collected in the 1st stage of the project of planning and set-up as well as during system maintenance interventions and testing activities The designation of these data includes the infix **TEST**.
- Data collected in the 2nd stage of the project, corresponding to the circumnavigation campaign. The designation of these data includes the infix **SAIL**.
- Data collected after May 10th 2020, which includes data from different Sagres campaigns on Portuguese waters as well as data collected when the ship was docked at base naval de Lisboa. The designation of these data includes the infix **SAGRES**.

data collection	period
TEST	before 2020-01-05 (installation stage) 2020-03-08 to 2020-03-09 (change of onboard computer) 2020-06-16 (calibration activities at pier - Lisboa BNL) 2021-06-21 to 2021-06-24 (change & update of onboard computer)
SAIL	2020-01-05 to 2020-05-10
SAGRES	2020-05-11 - 2021-03-02 2021-06-25 -

Table 3: data collection activities.

Each of these data collection activities is handled separately, but each comprises the following groups of data:

- **Ship data** - data obtained directly from the ship onboard system (see section 2.4).The designation of these data includes the infix **SHIP**.
- **Sensor data** - data derived from the ship data by correction of logging errors, as detailed in the technical report of Amaral & Dias (2021). The designation of these data includes the infix **SD**.
- **Geosensor data** - data derived from the sensor data by adding two additional columns corresponding to latitude and longitude. The geo-referencing of the files based on the GNSS data is described in the technical report of Ferreira (2021). The designation of these data includes the infix **GD**.
- **Preprocessed Data** - data derived from the geosensor data by applying variable-specific pre-processing and quality-control procedures. The designation of these data includes the infix **PD**.

Not all parameters are available for all types of data collection activities or for all types of data:

- GNSS measurements in NMEA and RINEX formats were not produced on board (before June 2021) and thus are only available for Sensor data, not for Ship data. Ship data in the RINEX format are available since 2021-06-25.
- Solar radiation measurements (SL) were collected in the same datalogger as E1 measurements, thus are available individually only on Sensor data, and not on ship data (in which those measurements are in the same files as E1 measurements).
- Cosmic radiation (CR) measurements are not available before July 2021 due to a failure in the sensor. The sensor was replaced by a new one in 2021-07-22.

Table 4 summarizes the naming conventions adopted for all the data from the project, which are formed as: type of activity + group of data infix + measured parameter infix.

	Testing	Circumnavigation	Sagres
Ship data	TEST_SHIP_*	SAIL_SHIP_*	SAGRES_SHIP_*
Sensor data	-	SAIL_SD_*	SAGRES_SD_*
Geosensor data	-	SAIL_GD_*	SAGRES_GD_*
Preprocessed data	-	SAIL_PD_*	SAGRES_PD_*

Table 4: dataset naming conventions; * denotes the parameter infix described in Table 2.

2.4. Data Storage and backup

The data from the set-up stage will be preserved on the PI's PC as well as on the INESC TEC institutional Drive (https://drive.inesctec.pt). Paper documents and documentation detailed in 2.1. will be preserved in the offices of the researchers responsible for the project. Selected documents (e.g. project protocols, instruments manuals,...) are preserved on the INESC TEC institutional drive https://drive.inesctec.pt/s/9ded5tFapW526BR.

Test data collected after the set-up stage – corresponding to data collected after changes in the monitoring system (e.g. change of the onboard computer hardware/software) or maintenance interventions, are also preserved both in the PI's PC and the INESC TEC institutional drive (https://drive.inesctec.pt/s/LaMx4nFyQjdndkB). Data collected for a short period from experimental sensors installed in the tow fish - hydrophone (University of Minho) and spectral radiance (INESC TEC / CAP) – are also preserved as test data.

Non-automatic data are collected by the Portuguese navy staff on pre-defined paper forms. The information collected in these records is later inserted by the navy staff into a computer spreadsheet. The paper records are preserved by the navy only, while the electronic spreadsheets are also stored at the INESC TEC repository for research data.

The automatic data from the SAIL monitoring system are stored in the onboard computer, organised in a separate folder for each day containing the individual hourly files for each sensor. Every day, at 01:00, the folder for each day is compressed (tar.gz format). The compressed data are kept for 2 days on the ship computer and copied automatically to a Network-attached storage (NAS) also on board the ship. A further copy of the data is stored in a SSD external disk (256 GB), which can be used by the INESC TEC technical team to access the data while on board or when the ship is docked. The monitoring data on the external disk (≲15 GB /day) are further transferred to an institutional server (sftp02.inesctec.pt) for backup and storage of all the data.

Problems on the onboard computer (hardware issues, possibly related to overheating due to air conditioning malfunction), occurred on 8 March 2020, 4 April 2020, and 3 March 2021, implying loss of all data for the following periods:
- 2020-03-08 to 2020-03-09
- 2020-04-04 to 2020-04-06
- 2021-03-03 to 2021-06-22

Security procedures are applied to all laptops and computers used during the project. All equipment is password secure, all software licensed and frequently updated to assure up to date security. All technical issues related to the software are controlled by each member of the project and in case support is needed the IT staff of the responsible entity will be contacted.

The storage structure for the different types of data is detailed below.

2.4.1. Ship data

Ship data are denoted with the infix **SHIP** (Table 4), preceded by the infix indicating the type of activity collection (Table 3) and followed by the infix relative to the specific parameter (Table 2). The data are stored in compressed files for each individual day, containing separate individual files for each hour of measurements. The storage structure of the ship data is summarised in Table 4, followed by the file structure for each measured parameter.

compressed files name	individual files
*_SHIP_CI_yyyymmdd.tgz	yyyymmdd-10s.records yyymmdd-1s.records yyyymmdd-block.records yyyymmdd.log
*_SHIP_E1_yyyymmdd.tgz	E1_yyyymmdd_HH.txt

*_SHIP_E2_yyyymmdd.tgz	E2_yyyymmdd_HH.txt
*_SHIP_GA_yyyymmdd.tgz	GA_yyyymmdd_HH.txt
*_SHIP_NS_yyyymmdd.tgz	GPS1_A1_oem4_yyyymmdd_HH GPS1_A2_oem4_yyyymmdd_HH GPS1_IMU_yyyymmdd_HH GPS1_NMEA_yyyymmdd_HH GPS2_A1_oem4_yyyymmdd_HH GPS2_A2_oem4_yyyymmdd_HH GPS2_IMU_yyyymmdd_HH GPS2_NMEA_yyyymmdd_HH GPS3_A1_oem4_yyyymmdd_HH GPS3_A2_oem4_yyyymmdd_HH rover_GPS1_A2_base_GPS1_A1_HH rover_GPS2_A1_base_GPS1_A2_HH rover_GPS2_A1_base_GPS1_A2_HH rover_GPS2_A2_base_GPS1_A1_HH rover_GPS2_A2_base_GPS2_A1_HH rover_GPS3_A1_base_GPS1_A2_HH rover_GPS3_A1_base_GPS2_A1_HH rover_GPS3_A1_base_GPS3_A2_HH
*_SHIP_TF_CTD_yyyymmdd.tgz	CTD_yyyymmdd_HH.txt
*_SHIP_VI_yyyymmdd.tgz	VI_yyyymmdd_HH.txt

Table 4: structure of ship data.

Electric field

The electric field hourly datafiles E1_yyyymmdd_HH.txt (upper mast sensor) contained in the compressed file *_SHIP_E1_yyyymmdd.tgz have the following structure:

col 1: timestamp (seconds.microseconds)
col 2: date (mm/dd/yyyy)
col 3: time (HH:MM:SS)
col 4: voltage (power) (V)
col 5: voltage (internal) (V)
col 6: Panel temperature (deg C)
col 7: Electric field (V/m)
col 8: Leakage current (nA)
col 9: CS110 status (numeric code)
col 10: Internal RH (%)
col 11: shortwave incoming radiation (W/m2)
col 12: shortwave outgoing radiation (W/m2)

The electric field hourly datafiles E2_yyyymmdd_HH.txt (lower mast sensor) contained in the compressed file *_SHIP_E2_yyyymmdd.tgz have the following structure:

col 1: timestamp (seconds.microseconds)

col 2: date (mm/dd/yyyy)
col 3: time (HH:MM:SS)
col 4: voltage (power) (V)
col 5: voltage (internal) (V)
col 6: Panel temperature (deg C)
col 7: Electric field (V/m)
col 8: Leakage current (nA)
col 9: CS110 status (numeric code)
col 10: Internal RH (%)

Gamma radiation

The gamma radiation hourly datafiles GA_yyyymmdd_HH.txt contained in the compressed file *_SHIP_GA_yyyymmdd.tgz have the following structure:

col 1: timestamp (seconds.microseconds)
col 2: date (mm/dd/yyyy)
col 3: time (HH:MM:SS)
col 4: gamma counts (counts per second)
col 5: battery voltage (kV)

Visibility

The visibility hourly datafiles VI_yyyymmdd_HH.txt contained in the compressed file *_SHIP_VI_yyyymmdd.tgz have the following structure:

col 1: timestamp (seconds.microseconds)
col 2: date (mm/dd/yyyy)
col 3: time (HH:MM:SS)
col 4: sensor ID string (SWS050)
col 5: sensor identification number (=001)
col 6: averaging time period in seconds (=060)
col 7: meteorological optical range (m)
col 8: obstruction to vision code
col 9: extinction coefficient
col 10: self-test diagnostics

Ions

The cluster ions counter (CIC) instrument measures produces raw measurement records about 30 times per second, which are automatically processed and turned into averaged records by the measurement program, yielding block average records (one average record from start to finish of each entry in the measurement cycle) and 1 second and 10 second averages. The compressed file *_SHIP_CI_yyyymmdd.tgz includes a text file of outputs from the sensor on the measurement process (yyyymmdd.log) and the hourly datafiles of block averages (yyyymmdd-block.records), 10-seconds measurements (files yyyymmdd-10s.records) and 1-second measurements (files yyymmdd-1s.records) with the following structure:

col1: date (yyyy-mm-dd)
col2: initial time (HH:MM:SS.microseconds)
col3: date (yyyy-mm-dd)
col4: final time (HH:MM:SS.microseconds)
col5: string of operation mode ("ions" / “offset”)
col 6 to col 11: cur_0 to cur_5 (electrometer currents)
col 12 to col 17: curvar_0 to curvar_5 (current variance estimates - noise level estimates)
col 18 to col 23: rawcur_0 to rawcur_5 (raw electrometer currents)
col 24 to col 29: volt_0 to volt_5 (electrometer voltages)
col 30: filter mean
col 31: filter stdev
col 32: Neg air conductivity mean (fS/m)
col 33: Neg air conductivity stdev (fS/m)
col 34: Neg avg cluster ion mobility (cm^2/V/s)
col 35: Neg avg cluster ion mobility stddev (cm^2/V/s)
col 36: Neg sample air pressure (hPa)
col 37: Neg sample air pressure stddev (hPa)
col 38: Neg sample air pressure sensor vtg (V)
col 39: Neg sample air pressure sensor vtg stddev (V)
col 40: Neg central electrode voltage (V)
col 41: Neg central electrode voltage stddev (V)
col 42: Neg central electrode voltage target (V)
col 43: Neg central electrode voltage target stddev (V)
col 44: Neg central electrode voltage sensor vtg (V)
col 45: Neg central electrode voltage sensor vtg stddev (V)
col 46: Neg central electrode voltage control vtg (V)
col 47: Neg central electrode voltage control vtg stddev (V)
col 48: Neg concentration 1 (#/cm^3)
col 49: Neg concentration 1 stddev (#/cm^3)
col 50: Neg concentration 2 (#/cm^3)
col 51: Neg concentration 2 stddev (#/cm^3)
col 52: Neg concentration 3 (#/cm^3)
col 53: Neg concentration 3 stddev (#/cm^3)
col 54: Neg filter voltage (V)
col 55: Neg filter voltage stddev (V)
col 56: Neg filter voltage enabled (V)
col 57: Neg filter voltage enabled stddev (V)
col 58: Neg filter voltage sensor vtg (V)
col 59: Neg filter voltage sensor vtg stddev (V)
col 60: Neg mobility 1 (cm^2/V/s)
col 61: Neg mobility 1 stddev (cm^2/V/s)
col 62: Neg mobility 2 (cm^2/V/s)
col 63: Neg mobility 2 stddev (cm^2/V/s)
col 64: Neg mobility 3 (cm^2/V/s)
col 65: Neg mobility 3 stddev (cm^2/V/s)
col 66: Neg sample flow (l/min)
col 67: Neg sample flow stddev (l/min)
col 68: Neg sample flow target (l/min)
col 69: Neg sample flow target stddev (l/min)
col 70: Neg sample flow sensor vtg (V)
col 71: Neg sample flow sensor vtg stddev (V)
col 72: Neg sample flow control vtg (V)

col 73: Neg sample flow control vtg stddev (V)
col 74: Neg sample flow target vtg (V)
col 75: Neg sample flow target vtg stddev (V)
col 76: Pos air conductivity (fS/m)
col 77: Pos air conductivity stddev (fS/m)
col 78: Pos avg cluster ion mobility (cm^2/V/s)
col 79: Pos avg cluster ion mobility stddev (cm^2/V/s)
col 80: Pos sample air pressure (hPa)
col 81: Pos sample air pressure stddev (hPa)
col 82: Pos sample air pressure sensor vtg (V)
col 83: Pos sample air pressure sensor vtg stddev (V)
col 84: Pos central electrode voltage (V)
col 85: Pos central electrode voltage stddev (V)
col 86: Pos central electrode voltage target (V)
col 87: Pos central electrode voltage target stddev (V)
col 88: Pos central electrode voltage sensor vtg (V)
col 89: Pos central electrode voltage sensor vtg stddev (V)
col 90: Pos central electrode voltage control vtg (V)
col 91: Pos central electrode voltage control vtg stddev (V)
col 92: Pos concentration 1 (#/cm^3)
col 93: Pos concentration 1 stddev (#/cm^3)
col 94: Pos concentration 2 (#/cm^3)
col 95: Pos concentration 2 stddev (#/cm^3)
col 96: Pos concentration 3 (#/cm^3)
col 97: Pos concentration 3 stddev (#/cm^3)
col 98: Pos filter voltage (V)
col 99: Pos filter voltage stddev (V)
col 100: Pos filter voltage enabled
col 101: Pos filter voltage enabled stddev
col 102: Pos filter voltage sensor vtg (V)
col 103: Pos filter voltage sensor vtg stddev (V)
col 104: Pos mobility 1 (cm^2/V/s)
col 105: Pos mobility 1 stddev (cm^2/V/s)
col 106: Pos mobility 2 (cm^2/V/s)
col 107: Pos mobility 2 stddev (cm^2/V/s)
col 108: Pos mobility 3 (cm^2/V/s)
col 109: Pos mobility 3 stddev (cm^2/V/s)
col 110: Pos sample flow
col 111: Pos sample flow stddev
col 112: Pos sample flow target
col 113: Pos sample flow target stddev
col 114: Pos sample flow sensor vtg
col 115: Pos sample flow sensor vtg stddev
col 116: Pos sample flow control vtg
col 117: Pos sample flow control vtg stddev
col 118: Pos sample flow target vtg
col 119: Pos sample flow target vtg stddev
col 120: diagnostic flags

Tow-fish

The hourly datafiles CTD_yyyymmdd_HH.txt of oceanographic measurements performed by the tow-fish, which are contained in the compressed files SAIL_SHIP_TF_CTD_yyyymmdd.tgz have the following structure:

col 1: timestamp (seconds.microseconds)
col 2: Pressure (dbar)
col 3: Temperature (degrees C)
col 4: Conductivity (mS/cm)
col 5: Salinity
col 6: Oxygen (%)
col 7: Oxygen (ppm)
col 8: pH
col 9: Chlorophil(a) (microgram/l)
col 10: Turbidity (FTU)
col 11: Phycoerythrin (microgram/l)
col 12: sensor time and memory

2.4.2. Sensor data

Sensor data are denoted with the infix **SD** (Table 4), preceded by the infix indicating the type of activity collection (Table 3) and followed by the infix relative to the specific parameter (Table 2). The data are stored in compressed files for each individual day, containing separate individual files for each hour of measurements. The storage structure of the sensor data is the same as for the ship data in Table 4.

Electric field

The electric field hourly datafiles E1_yyyymmdd_HH.txt (upper mast sensor) contained in the compressed file *_SD_E1_yyyymmdd.tar.gz have the following structure:

col 1: timestamp (seconds.microseconds)
col 2: Electric field (V/m)
col 3: Leakage current (nA)
col 4: CS110 status (numeric code)
col 5: Internal RH (%)

The electric field hourly datafiles E2_yyyymmdd_HH.txt (lower mast sensor) contained in the compressed file *_SD_E2_yyyymmdd.tar.gz have the following structure:

col 1: timestamp (seconds.microseconds)
col 2: Electric field (V/m)
col 3: Leakage current (nA)
col 4: CS110 status (numeric code)
col 5: Internal RH (%)

Gamma radiation

The gamma radiation hourly datafiles GA_yyyymmdd_HH.txt contained in the compressed file *_SD_GA_yyyymmdd.tar.gz have the following structure:

col 1: timestamp (seconds.microseconds)
col 2: gamma counts (counts per second)
col 3: battery voltage (kV)

Visibility

The visibility hourly datafiles VI_yyyymmdd_HH.txt contained in the compressed file *_SD_VI_yyyymmdd.tar.gz have same structure as the ship data (*_SHIP_VI_yyyymmdd.tgz).

Solar radiation

The solar radiation hourly datafiles SL_yyyymmdd_HH.txt contained in the compressed file *_SD_SL_yyyymmdd.tar.gz have the following structure:

col 1: timestamp (seconds.microseconds)
col 2: shortwave incoming radiation (W/m2)
col 3: shortwave outgoing radiation (W/m2)

Tow-fish

The hourly datafiles CTD_yyyymmdd_HH.txt of oceanographic measurements performed by the tow-fish, which are contained in the compressed file *_SD_CTD_yyyymmdd.tar.gz have the following structure:

col 1: timestamp (seconds.microseconds)
col 2: Pressure (dbar)
col 3: Temperature (degrees C)
col 4: Conductivity (mS/cm)
col 5: Salinity
col 6: Oxygen (%)
col 7: Oxygen (ppm)
col 8: pH
col 9: Chlorophil(a) (microgram/l)
col 10: Turbidity (FTU)
col 11: Phycoerythrin (microgram/l)

2.4.3. Geosensor data

Geosensor data are denoted with the infix **GD** (Table 4), preceded by the infix indicating the type of activity collection (Table 3) and followed by the infix relative to the specific parameter (Table 2). The data are stored in compressed files for each individual day, containing separate individual files for each hour of measurements. The storage structure of the geosensor data is the same as for the ship sensor data, with the exception of two additional columns, corresponding to latitude (degrees N) and longitude (degrees E).

2.4.4. Preprocessed data

Preprocessed data are denoted with the infix **PD**, preceded by the infix indicating the type of activity collection (Table 3) and followed by the infix relative to the specific parameter (Table 2).

Electric field

The preprocessed electric field data are obtained from the geosensor data files (SAIL_GD_E1.tar.gz and SAIL_GD_E2.tar.gz, respectively) by applying the following procedures:

- merging of hourly files into a single datafile;
- if sensor status code is not 1 (good instrument health) the corresponding measurement is set as missing (represented as NA);
- changing the sign of electric field measurements to comply with the sign convention denoting the potential gradient as positive under undisturbed atmospheric electrical conditions;
- averaging of 1-second electric field measurements into 1-min values;
- computation of the standard deviation every 1–minute from the 1-second measurements;
- averaging of coordinates to 1-minute averaged values (taking into account angularity).
- check record continuity (insert NA flag for missing observations)

The jupyter notebooks implementing these pre-processing steps are preserved in Zenodo (DOI: 10.5281/zenodo.10125809).

The resulting files SAIL_PD_E1.txt (upper mast) and SAIL_PD_E2.txt (lower mast) have the following structure:

col 1: date (yyyy-mm-dd)
col 2: time (HH:MM)
col 3: potential gradient (V/m)
col 4: standard deviation (V/m)
col 5: longitude (degrees E)
col 6: latitude (degrees N)

Gamma radiation

The preprocessed gamma radiation data are obtained from the geosensor data files SAIL_GD_GA.tar.gz by applying the following procedures:

- merging of hourly files into a single datafile;
- aggregation of 1-second measurements into 1-min values (sum).
- computation of the standard deviation every 1–minute from the 1-second measurements;
- averaging of coordinates to 1-minute averaged values (taking into account angularity);
- check record continuity (insert NA flag for missing observations).

The resulting file SAIL_PD_GA.txt has the following structure:

col 1: date (yyyy-mm-dd)
col 2: time (HH:MM)
col 3: gamma counts (counts/minute)
col 4: standard deviation (counts/minute)
col 4: longitude (degrees E)
col 5: latitude (degrees N)

The time series of 1-minute pre-processed gamma radiation counts displays several outliers, mostly corresponding to short (< 3 minutes) anomalous values, typically sharp spikes and anomalous low values before/after a data gap (these associated with the recovery of the sensor after power failure). These anomalous values are set as missing (replaced by the NA flag).

The jupyter notebooks implementing these pre-processing and quality-control procedures are preserved in Zenodo (DOI: 10.5281/zenodo.10134221).

Visibility

The preprocessed visibility data are obtained from the geosensor data files SAIL_GD_VI.tar.gz by applying the following procedures:

- check record continuity (insert NA flag for missing observations)
- extract meteorological optical range field

The resulting file SAIL_PD_VI.txt has the following structure:
col 1: date (yyyy-mm-dd)
col 2: time (HH:MM)
col 3: meteorological optical range (m)
col 4: longitude (degrees E)
col 5: latitude (degrees N)

The jupyter notebooks implementing these pre-processing steps are preserved in Zenodo (DOI: 10.5281/zenodo.10137012).

2.5. Data selection and preservation

The data acquired collected during the SAIL monitoring campaign are unique and thus very important for other researchers as well as for educational purposes. All the campaign data will be preserved to enable initially unforeseen uses of the data and to guarantee fully-documented and reproducible data from the project, ensuring the reuse of the data in multiple environmental domains and different applications.

All the campaign data will be preserved at the INESC TEC institutional repository (https://rdm.inesctec.pt) for at least 5 years after the end of the project completion (at least until 2027) The procedures enabling the transformation of Ship Data into Sensor Data and Geosensor Data are described in dedicated reports available on Zenodo (see section 2.2). The procedures describing the transformation of Geosensor Data into Preprocessed data will be preserved in the form of computational (Jupyter) notebooks and made available on Zenodo, linked to SAIL's Zenodo community.

The project PI will be responsible for any action related to the long-time preservation in accordance with the repository guarantees. The possibility of long-term preservation of the same data in another repository will be analysed during the project and the corresponding information added if necessary in a further version of the DMP.

Each preserved dataset (Table 5) will have Digital Object Identifiers (DOIs) assigned by the INESC TEC research data management repository.

Dataset	Availability	DOI
SAIL raw atmospheric data (ship data + sensor data + geosensor data)	Upon request	https://doi.org/10.25747/b2ff-kg31
SAIL data - electric field (preprocessed data)	Open	
SAIL data - gamma radiation (preprocessed data)	Open	
SAIL data - visibility data (preprocessed data)	Open	
SAIL data - solar radiation data (preprocessed data)	Open	
SAIL data - oceanographic data (preprocessed data)	Open	
SAIL data - meteorological observations	Upon request	https://doi.org/10.25747/rp31-kf14
SAIL data - fish logs	Upon request	
Sagres - raw data	Upon request	
Sagres - processed data	Open	

Table 5: preserved datasets.

2.6. Data sharing

Access to the **Test data** is restricted to the PI and co-workers designated by the PI. Access can be granted to others in case of a specific request to the PI.

The **raw data**, comprising the **Ship data, Sensor data** and **Geosensor data**, will not be, in general, publicly available - mainly due to its large size - but will be shared in case of a specific request to the PI.

Ship data acquired by the radiation sensor SN005 will be shared with the University of Bristol, according to the Memorandum of Understanding of 2019/12/19, and with the company Asquared from July 2021 onwards according to the Loan agreement of the instrument signed in June 2021.

The **Preprocessed Data** will be publicly available at https://rdm.inesctec.pt with licence Share-Alike, which allows to (https://creativecommons.org/licenses/by-sa/4.0/):

> Share — copy and redistribute the material in any medium or format;
>
> Adapt — remix, transform, and build upon the material for any purpose, even commercially.
>
> Under the following terms:
>
> - Attribution — You must give appropriate credit, provide a link to the licence, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
>
> - ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same licence as the original.
>
> - No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the licence permits".

Project outputs (presentations, posters, papers,...) will be shared on the Project SAIL community on ZENODO (https://zenodo.org/communities/sail/).

3. Responsibilities and resources

3.1. Responsibilities

3.1.1. Responsible entities

Principal (coordinating) institution:

INESC TEC - Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC); (Contact person: Susana Barbosa)

Participating Institutions:

- Marinha (Marinha/MDN) (Contact person: António Manuel Maurício Camilo)

- Atlantic International Research Centre (AIR Centre) (Contact person: José Luiz Moutinho)

- Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR) (Contact person: Sandra Ramos)

- Universidade do Minho (UM) (Contact person: Marcos Martins)

3.1.2. Responsible persons

Responsible for the collection of the data: Susana Barbosa (project PI) / Eduardo Alexandre Pereira da Silva (Engineering coordination)

Responsible for the processing and preservation of the data: Susana Barbosa (project PI)

Responsible for backups: Susana Barbosa (project PI); The system administrators of INESC TEC are responsible for the backups related to the INESC TEC research data repository;

Responsible for publishing and sharing data : Susana Barbosa (project PI)

Responsible for DMP creation : Susana Barbosa (project PI) / Yulia Karimova, (ORCID: http://orcid.org/0000-0002-1015-6709)

The Sensor Data from atmospheric sensors and tow fish measurements is owned by INESC TEC. All the Sensor Data, Geosensor Data and Preprocessed Data will be organised and preserved by INESC TEC.

Meteorological data resulting from onboard measurements performed by the navy are owned by Marinha, as well as data on fish captures. Specific sharing procedures will be specified by the data owner and detailed in a posterior version of the DMP.

Methodology for collection and storage of fish samples and data resulting from the laboratory processing of fish samples is owned by CIIMAR. Specific sharing procedures will be specified by the data owner and detailed in a posterior version of the DMP.

Data resulting from the processing of acoustic noise and turbidity measurements is owned by the University of Minho. Specific sharing procedures will be specified by the data owner and will be detailed in a posterior version of the DMP.

Data resulting from the on-board workshops (participants lists, workshop's programs, slides) is owned by AIR Centre. Specific sharing procedures will be specified by the data owner.

3.2. Resources

To deliver the data management plan it is necessary to have access to the INESC TEC research data repository (https://rdm.inesctec.pt/) for data deposit and access to the Project SAIL community on ZENODO (https://zenodo.org/communities/sail/?page=1&size=20).

Moreover, during the project the following assets will be used:

- Hardware/devices: on-board main computer and data storage, laptop computers for data acquisition and processing in mission.

- Institutional servers, USB flash drives, data storage disks, smart-phones, tablets.

- Software: all software will be open source running in the Linux operating system.

- Cloud services: https://drive.inesctec.pt, with access granted to a responsible from each participating institution according to point 2.6.

4. Ethics and legal compliance

The SAIL project complies with all existing requirements related to the Research Data Management and Protection of Personal Data. The diverse types of data that will be collected during the SAIL project are under no specific legal requirements and will be shared as detailed in section 2.6. No dataset containing confidential information, or any ethical or legal issues will be deposited.

All data collected during the SAIL project is a direct result of the project and thus copyrights and IPR belong to the SAIL members indicated on point 3. Data reusing should be made according to the established licence along with the dataset citation.