DATA MANAGEMENT PLAN

PROJECT	
Project number:	2024.13655.PEX
Project acronym:	ECHO
Project name:	Earthquakes, Currents, Hydroacoustics, and Oceanography with Distributed Acoustic Sensing

DATA MANAGEMENT PLAN	
Date:	2026-05-19
Version:	1.2

1. Executive Summary

This Data Management Plan (DMP) is based on the Horizon Europe Data Management Template (v1.1–01/04/2022), and the knowledge acquired during the SUBMERSE project (EU GA 101095055), regarding the specific challenges connected to Distributed Acoustic Sensing (DAS) data. In this document, we describe the framework for data acquisition, scrubbing (omission of security-sensitive spatial or acoustic signatures), storage, distribution, and publication. Dual-use is mitigated by spatial coordinates degradation, delayed release, and filtering procedures.

This document describes how project ECHO balances the importance of open and FAIR data with the potential dual-use of the raw data generated, while ensuring that scientific applications are impeded as little as possible by security restrictions.

2. Data Summary

Project ECHO will generate the following data types and formats:

1. Raw, full-resolution, continuous DAS datasets, in HDF5 format
2. Decimated and downsampled continuous DAS datasets in miniSEED format
3. Raw, full-resolution, windowed DAS datasets in both HDF5 and miniSEED formats
4. Decimated and downsampled real-time DAS streams in SeedLink format
5. Full-resolution seismic datasets from Ocean Bottom Seismometers (OBS) and land seismic stations, in miniSEED format
6. Full-resolution ADCP and CTD datasets, in CSV format
7. Metadata fully describing each dataset

All newly data generated will be directly related to the different project tasks.

It is estimated that ECHO will collect several hundred TB of raw DAS data (typically ~1 TB/day). After scrubbing and downsampling, the total volume to be permanently stored in a public repository will amount to several tens of TB. Seismic datasets from OBS and land station will amount to ~300 MB/day. ADCP and CTD data will amount to less than 6 GB.

Project ECHO will re-use data from previous experiments on the GeoLab fibre. They extend the temporal window beyond the short duration of an exploratory project such as ECHO. These datasets are or will become openly available during the project, and are as follows:

1. Raw full-resolution data in HDF5 format from the 3X temporary network (open)
2. Decimated and downsampled data in miniSEED format from the 3X temporary network (open)
3. Raw full-resolution data in HDF5 format from the 5V temporary network (planned release on June 2026)
4. Decimated and downsampled data in miniSEED format from the 5V temporary network (planned release on June 2026)

Project ECHO will also use existing data from other sources to complement the DAS, OBS, and ADCP datasets, in the following data types and formats:

5. EMODnet bathymetry (open)
6. Copernicus re-analysis data (open)
7. COAWST modelling data (open)
8. Buoy data from APRAM (open)
9. Marine mammal sightings lists (internal documents)

The newly collected datasets will extend well beyond the project's immediate scope. They will be valuable to researchers in oceanography, seismology, underwater acoustics, marine biology, and ocean soundscapes, as well as to stakeholders such as cable owners and operators, regional environmental and weather agencies, and port authorities. More broadly, these datasets can support societal applications, particularly through their integration into training frameworks for early-warning systems.

3. FAIR data

3.1. Making data findable, including provisions for metadata

All data made available at the conclusion of the ECHO project will be assigned a persistent identifier, and published with complete metadata in the appropriate formats. Coordinates for the GeoLab cable will be obfuscated for security purposes, degraded to a precision of around 200 metres, consistent with coordinates available from public documents and official cartography sources. Metadata will be generated according to the best practices of each field.

3.2. Making data accessible

Repository:

At the end of the project, data will be deposited in a trusted repository (e.g., EIDA, GEOFON, Zenodo). Real-time streams will be deposited at IPMA's EIDA node.

Each dataset generated by ECHO will be assigned a DOI through its hosting repository. DAS datasets will be assigned a temporary network code at FDSN.

Data:

All data collected during the ECHO project will be made openly available at the conclusion of the project, after an embargo period of three years after the first sample is collected, for exploitaition and security purposes. Specific portions of the datasets might only be available after legally-mandated scrubbing, according to future national or European legislation. Coordinates for the GeoLab cable will be obfuscated for security purposes, degraded to a precision of approximately 200 metres, consistent with coordinates available from public documents and official cartography sources.

Data will be stored in permanent repositories, and will be available indefinitely. Additionally, backup copies will be kept within the expected duration of the supporting media. Data will be available through free and standardized access protocols, and licensed under a Creative Commons 4.0 BY (Attribution to the creator of the dataset).

Metadata:

DAS datasets still don’t have consistent metadata standards, but the team members of the ECHO project have been closely working with GEOFON, FDSN, and the DAS-RCN Data Management Working Group to establish a metadata standard to increase reusability. ECHO will provide, at a minimum, a stationXML file that describes the experiment, the data, and the acquisition parameters in sufficient detail, including digital filter coefficients and searchable keywords. Additional metadata formats will be made available according to any updated guidelines emanating from the DAS-RCN Data Management Working Group and FDSN.

For all other datasets, the appropriate metadata standards will be used.

Metadata will remain available even after data is no longer available.

Documentation on how to access the data, including software, will be included with the metadata. Source code will be released as open-source.

3.3. Making data interoperable

To increase the data interoperability, the ECHO project will follow community-endorsed data and metadata formats. Full-resolution, raw DAS data will be released in HDF5 format, accompanied by downsampled and decimated subsets in miniSEED format. DAS metadata will be provided in StationXML or JSON formats, following the scheme proposed by FDSN.

Oceanographic data, including ADCP and CTD observations, will be stored in open, used formats such as CSV and NetCDF, following the CF (Climate and Forecast) metadata conventions.

Marine mammal acoustic detection datasets will be stored in CSV with accompanying metadata. The data structure will follow bioacoustic metadata practices described by Roch et al. (2016). Each detection will include project, site, sensor and/or channel identifiers, input data source or file reference, detection time (UTC), species or taxonomic group code, call type where applicable, signal frequency range (Hz) and duration (s) where available, detector or classifier used, software version, relevant detector parameters or thresholds where applicable, quality or validation notes, effort type, and analyst name. Accompanying metadata will document information about the recording and files used, including sampling rate, analysed periods and channels, and detection workflow applied.

Marine mammal sightings data from whale-watching companies operating on the south coast of Madeira Island will be provided in CSV or XLS format, for comparison with acoustic detections. Each sighting provided will include the name of the company, date, time, GPS coordinates, species identification (to the species level whenever possible), and group size.

Coordinates will be provided in latitude/longitude pairs in WGS84, with depths indicated as negative altitudes. Timing information will be in UTC.

Any data collected outside the ECHO project scope will be referenced using persistent identifiers and, whenever possible, openly-accessible repositories.

3.4. Increase data re-use

Beyond the datasets, a comprehensive data report will be published in an open-access, peer-reviewed journal, detailing not only the acquisition parameters, but also examples of processed data and analysis.

After the embargo period, data will be released under a CC 4.0 BY license, allowing use by third parties after the end of the project. Provenance of the data, acquisition parameters, formats, and repository addresses will be thoroughly documented using the appropriate standards.

During acquisition, and especially before being uploaded to the permanent repositories, all datasets will be verified to ensure quality standards are followed. Major checks will be focused on spatial and temporal data completeness, noise levels, and metadata completeness.

Research outputs, such as transfer functions, and data catalogues, will be subject to the same checks.

Citation requirements will be added to each output and dataset made available.

The contact points for reuse after the end of the project will be the PI or ARDITI

4. Other research outputs

Software developed during the duration of the ECHO project will be published as open-source in public permanent repositories (e.g., GitHub + Zenodo archiving), with persistent identifiers assigned to each output, and licensed under MIT or GPL terms.

5. Allocation of resources

Making data and research outputs FAIR has direct and indirect cost related to storage, bandwidth, and archival. The ECHO project will be responsible for bearing the storage costs until the end of the project (1 k€/ month of collected data). After that date, permanent storage costs will fall on the permanent repository structures.

The Principal Investigator (PI) of the ECHO project will be responsible for the Strategic oversight and Compliance with DMP and funder requirements. Task Leaders will be responsible for datasets generated within their tasks, ensuring metadata completeness and quality control.

After the end of the project, long term preservation of data and outputs will rely on the existing public repositories for scientific data.

6. Data security

After acquisition, datasets will be copied to multiple locations and on different storage media, ensuring a robust back-up strategy. DAS data will only be moved from the interrogator to storage over a segregated point-to-point fibre link, not publicly accessible.

Only project members or associated researchers will have access to raw data streams before the end of the embargo period. Processed data (levels 2 and 3) can be shared with external researchers and entities.

To prevent misuse of infrastructure data, processing of data will rely on the obfuscated coordinates to be made available publicly when data is released. No tap tests to accurately locate DAS channels will be performed.

Datasets and outputs will be safely stored in trusted repositories for long term preservation and curation.

7. Ethics

Barring legal framework changes, no ethics or legal issues are expected that could have an impact on data sharing.

To avoid impacts on marine life, only passive data acquisition techniques will be used. No biological samples will be collected. No personal data will be collected. This is done in compliance with EU directives.

Incidental handling of sensitive data (e.g., military vessel traffic inferred from DAS) will be done in cooperation with Portuguese authorities.

8. Other issues

This Data Management Plan does not replace or supersede applicable national or European data policies or regulations; in the event of any conflict, those policies shall prevail.

References

Roch, M.A., Batchelor, H., Baumann-Pickering, S., Berchok, C.L., Cholewiak, D., Fujioka, E., Garland, E.C., Herbert, S., Hildebrand, J.A., Oleson, E.M., Van Parijs, S., Risch, D., Širović, A., & Soldevilla, M.S. (2016). Management of acoustic metadata for bioacoustics. Ecological Informatics 31, 122–136. https://doi.org/10.1016/j.ecoinf.2015.12.002

HISTORY OF CHANGES		
VERSION	PUBLICATION DATE	CHANGE
1.0	2026-04-02	Initial version
1.1	2026-05-14	Explicitly set forth data security measures, ethics and other issues (sections 6, 7, and 8)
1.2	2026-05-19	Clarified the responsibility for data management Added oceanography datasets description Added marine mammals datasets description