European Research Council

Executive Agency

Established by the European Commission

Horizon 2020
European Union Funding
for Research & Innovation

European Research Council (ERC)

ERC Data Management Plan Template

Version 1.0
22 November 2021

ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

Project Acronym	Project Number
Sea2Cloud	

Template for the ERC Open Research Data Management Plan (DMP)[1]. The following sections should describe how you plan to make the project data Findable, Accessible, Interoperable and Reusable (FAIR).
Each of the following five issues should be addressed with a level of detail appropriate to the project.

SUMMARY *(dataset[2] reference and name; origin and expected size of the data generated/collected; data types and formats)*

For the Sea2Cloud project, a unique entry point will be provided to access to the different datasets, which is the Sea2Cloud data access catalog (https://sea2cloud.data-terra.org/en/catalogue/). The ocean and the atmospheric data are not necessarily archived in the same storage pool, but the objective is that the data catalog keeps it transparent for the user.
The project website (https://sea2cloud.data-terra.org/en) also contains the description of the campaign and the instruments.

The campaign results are composed of four different types of experiments, each potentially containing both seawater and atmospheric data (classically stored in different data bases, due to their production by different communities), and the dataset as a whole is named Sea2Cloud, with experiments named respectively Sea2Cloud-Tangaroa-ambient, Sea2Cloud-Tangaroa-chemical-fluxes, Sea2Cloud Tangaroa-sea spray-fluxes and Sea2Cloud-Baring Head.

The data formats are the ones used by the instrument PIs during acquisition and analysis. These are the native csv (ascii Comma Separated Values) and xlsx (Excel) files, eventually supplemented with quality flags added by the PIs.

The approximate size of the whole dataset is approximately 2 GB.

---

[1] *Based on 'Guidelines on FAIR Data Management in H2020', version 3.0. 26.07.2016, Annex1*
[2] *Several datasets may be included into a single DMP.*

1. MAKING DATA FINDABLE *(dataset description: metadata, persistent and unique identifiers e.g., DOI)*

The first data set was acquired during a ship campaign in March 2021 (11 days) and contains:
  (1) ambient air gaz-phase, aerosol particles and cloud properties from in situ (ship level) and remote sensing data. Besides raw data, different levels of data sets will be stored and made available following the ACTRIS types format requirements
  (2) underway seawater biogeochemical properties along the ship trajectory
  (3) meteorological (wind speed, direction, pressure temperature etc.) from in situ measurements and radio-soundings for profiles and ship data (position, roll and pitch etc.).

The second data set will be based on "incubation" experiments performed during the same ship campaign, they provide fluxes of gaz-phase compounds and particles at the sea-air interface. These data have never been acquired nor submitted to any data base in the past and will have no standards to rely on, but will probably submit time series of concentrations with a detailed meta data file.

The third data set is also dealing with fluxes, but measured using another device and adapted for sea spray fluxes. Sea spray fluxes are monitored for the size distribution of aerosol particles (sub and super-micron), CCN concentrations and online chemistry. These are continuously acquired data with a native 1 min resolution for some instruments (CPC total number concentrations) and 30 min averages for others and level 1 data. There will also be some discrete daily samples for this data set, very small in size.

The fourth data set is for ambient aerosol properties acquired from a fixed station (Baring Head, GAW regional station), for 8 months. These data include aerosol size distribution (10-500 nm), particle total concentrations from 1 nm, 2.5 nm and 10 nm, ions size distributions (2-40 nm), and ions chemical composition (APi ToF data).

All data sets will be made available at the rhythm of publication (there will be an embargo on individual data sets that have not yet been published, until they are). They will have their own DOI and their dataset description in the Sea2Cloud catalog as a landing. To link the 2 information, it will be registered on Datacite (https://datacite.org/) using its metadata scheme. Moreover, a dataset search tool is provided in the campaign catalog.

As an example, the first data set published is the BTEX fluxes from on deck experiments, which had the following DOI assigned https://dx.doi.org/10.25326/y3vz-

sp14. The landing page contains the abstract, the data access links, and the different points of contact.

2. MAKING DATA OPENLY ACCESSIBLE *(which data will be made openly available and if some datasets remain closed, the reasons for not giving access; where the data and associated metadata, documentation and code are deposited (repository?); how the data can be accessed (are relevant software tools/methods provided?)*

All datasets will be made available at the rhythm of publication (there will be an embargo on individual datasets that have not yet been published, until they are). Nevertheless, they all will be publically available after this phase, but with an accounting identification required (registration is free of charge).

As explained in the previous chapter, the documentation will be hosted on the campaign website (https://sea2cloud.data-terra.org/en), and the metadata will be available on the catalog (https://sea2cloud.data-terra.org/en/catalogue/). The data access links will point to 2 data repositories:
- for atmospheric data: the AERIS/ICARE Data and Services Center, which distributes the data as files using FTP and HTTPS protocols,
- for seawater data: the AERIS/SEDOO Data and Services Center, which distributes the data through REST API requests.

At the moment, no additional code is needed, but if needed, it could be hosted in the AERIS gitlab repository.

3. MAKING DATA INTEROPERABLE *(which standard or field-specific data and metadata vocabularies and methods will be used)*

The metadata scheme used to build the catalogue pages has been defined internally at AERIS but is compatible with the standard ISO 19115 scheme. A converter to ISO 19115 metadata is available, and could be used to feed geonetwork catalogues for instance. The instrument types and parameters vocabularies are based on the ones defined by the NASA's Global Change Master Directory (GCMD, https://gcmd.earthdata.nasa.gov), commonly used by the community.

4. INCREASE DATA RE-USE *(what data will remain re-usable and for how long, is embargo foreseen; how the data is licensed; data quality assurance procedures)*

All data will be hosted by AERIS, a data center strongly supported at national level. It guarantees data will remain available long after the campaign (see below).

The data are distributed using a CC-BY license (Creative Commons Attribution).
All data have been manually inspected and qualified by the Principal Investigators of the instruments before public release.

5. ALLOCATION OF RESOURCES and DATA SECURITY *(estimated costs for making the project data open access and potential value of long-term data preservation; procedures for data backup and recovery; transfer of sensitive data and secure storage in repositories for long term preservation and curation)*

The cluster for Atmospheric data AERIS (https://www.aeris-data.fr), part of the French Data Terra Research Infrastructure (RI), has the objective to facilitate and enhance the use of atmospheric data, whether from satellite, aircraft, balloon, or ground observations, or from laboratory experiments. It generates advanced products and provides services to facilitate data use, to prepare campaigns, and to interface with modeling activities. It consists of four Data and Service Centres (DSC) with more than 20 years experience and strong expertise in data curation, storage, preservation and dissemination: ICARE, ESPRI, SATMOS and SEDOO. Most of these data centres are involved in European initiatives and projects promoting the FAIR data principles and participating in the European Open Science Cloud (EOSC). AERIS is hosting the IAGOS data center and parts of the ACTRIS data center (trace gases remote sensing, atmospheric simulation chamber...).

The mid-term preservation (10 years) is guaranteed by AERIS. The question of long term preservation will be addressed soon in the framework of Data Terra RI. In case ICARE or SEDOO could not maintain this database anymore, the responsibility could be transferred to another data centre within AERIS. AERIS is strongly supported at national level (CNRS, CNES, Meteo-France…). In the highly unlikely event that AERIS will have to close operations, we guarantee that we will migrate all content to other suitable repositories, and since all datasets will have DOIs, all citations and links to datasets will not be affected.

The dataset does not contain sensitive data. The data are distributed by data centers, and so, by concept, the infrastructure and maintenance costs are mutualized, but for 2 GB, we can estimate a yearly cost of 1000 €, covered by public funding.

Upstream of the data distribution system, the data are stored on a NextCloud server hosted at AERIS/ICARE. It guarantees a local copy for client PCs and a centralized copy on the server. The data stored on this server are also backed up on LTO tapes on a daily basis.

For the atmosphere part, the data are then archived on a SAN (Storage Area Network) on hard drives, and the file systems are backed up on LTO tapes. In case of data loss on hard drives, the archive will be restored using these LTO backups.

For the seawater part, the data are then archived and backed up using a service

provided by the Observatoire Midi-Pyrénées that is hosting AERIS/SEDOO.

DISCLAIMER. Please note that the ERC Data Management Plan is not a part of the Ethics Review. It is the responsibility of the Principal Investigator to inform the ERCEA Ethics Team of any ethics issues/concerns regarding the collection, processing, sharing and storage of data in relation to the project. The Principal investigator can also be asked to submit an Ethics Data Management Plan (Ethics DMP).