ARCTIC PASSION

Deliverable 2.1

Data Management Plan

Version 0.1, 2022-01-13: First draft

Project
Arctic PASSION

EU Horizon 2020 grant agreement
101003472

Work package 2
Bringing the Arcrtic Data System into action

Lead beneficiary
1 - FMI

Lead author
Matias Takala (FMI)

Contributors
Øystein Godøy (MET)

Status
In preparation

Dissemination level
PU

Table of Contents

1. Data summary

The purpose of the data management plan is to document how the data generated by the project Arctic PASSION are handled during and after the project. It describes the basic principles for data management within the project. This includes standards for documentation at discovery and data levels as well as data sharing and preservation including life cycle management of datasets.

This document is a living document that will be updated during the project. Arctic PASSION is following the principles outlined by the Open Research Data Pilot and The FAIR Guiding Principles for scientific data management and stewardship (Wilkinson et al. 2016).

The Arctic PASSION project is described in more detail in the project website which is available at https://arcticpassion.eu/. The purpose of Arctic PASSION is creation and implementation of a coherent, integrated Arctic observing system. This implies integrating already existing observing systems in a systems of systems approach as well addressing gaps in the current observing system. Arctic PASSION will generate data through the project implementation, but equally important is to map and establish access to already existing datasets in support of the pilot services and other activities of Arctic PASSION.

Arctic PASSION will promote the use of self-explaining file formats (e.g. NetCDF, HDF/HDF5, DwCA) combined with semantic and structural standards like the Climate and Forecast Convention for data documentation. The default format for Arctic PASSION datasets in the geoscientific domain is NetCDF following the Climate and Forecast Convention (feature types grid, timeseries, profiles and trajectories if applicable). For data in the biological domain Darwin Core Archive is promoted. If none of these formats are suitable other formats can be used, but a detailed product manual following a template has to be prepared to ensure proper reuse of the data in the future.

Arctic PASSION will exploit existing data in the region. In particular operational meteorological data made available through WMO Global Telecommunication System (GTS) will be important for the model experiments. No full overview of third party data that will be used is currently available. An overview of the third party data to be used will be provided. Essentially this will at least, but limited to, include data from the World Meteorological Organisation, Copernicus services in Europe, data already generated by project partners and data found harvesting discovery metadata from relevant data centres.

If deemed necessary (required by the scientific community in Arctic PASSION) metadata describing relevant third-party observations will be harvested and ingested in the data management system and through this simplifying the data discovery process for Arctic PASSION scientists. If specifically needed by one of the pilot services of Arctic PASSION, data may also be cached to ensure interoperable data that can be used by the web based services of the pilot services[1].

Arctic PASSION will rely on both data generated by projhect partners during the duration of the project, legacy data and observing systems of the partners and third party data available through data centres not part of Arctic PASSION.

An overview of the data generated by the project will be made available in Appendix A, *Datasets*, more specifically in Table 1, “Overview of datasets generated by Arctic PASSION.”. This list serves as a reminder for datasets. Eventually all datasets should be discoverable through the Arctic PASSION data catalogue.

There is currently no estimate for the expected volume of the data. Such volume estimates only make

sense for the data actively managed by Arctic PASSION. These estimates will be generated when a better overview of the exact datasets is available. However it is expected that it will be in the order to several Terabytes.

Arctic PASSION aims to *bring the Arctic data into action*. Thus data can be relevant for many communities. Internally the primary purpose of the data is to serve the needs of the project's pilot services.

2. FAIR data

2.1. Making data findable, including provisions for metadata

2.2. Making data openly accessible

2.3. Making data interoperable

2.4. Making data reuseable

3. Allocation of resources

4. Data security

5. Ethical aspects

6. Other issues

Appendix A: Datasets

Table 1. Overview of datasets generated by Arctic PASSION.

#	Dataset	Description	Responsible	Generated	Published	Comment
1						
2						

[1] This could be necessary to establish an Arctic Window of Copernicus or when data are available through third party data centres but not in standardised and interoeprable form.