3.3 Data Management Plan

Deliverable ID:	D3.3
Dissemination Level:	PU
Project Acronym:	SINOPTICA
Grant:	892362
Call:	H2020-SESAR-2019-2
Topic:	SESAR-ER4-05-2019 - Environment and Meteorology for ATM
Consortium Coordinator:	CIMA Research Foundation
Edition date:	30 November 2020
Edition:	00.01.00
Template Edition:	02.00.02

Founding Members

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Riccardo Biondi/UNIPD	Project Communication Leader/Dr.	30/11/2020

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
Antonio Parodi/CIMA	Project Manager/Dr.	25/11/2020
Eugenio Realini/ GRED	Project Quality Leader/Dr.	25/11/2020

Approved for submission to the SJU By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date
Riccardo Biondi/UNIPD	Project Communication Leader/Dr.	30/11/2020

Rejected By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date

Document History

Edition	Date	Status	Author	Justification
00.00.01	30/10/2020	First version	Riccardo Biondi	
00.00.02	24/11/2020	Second version	Antonio Parodi	Inputs on data chain and archiving
00.00.03	25/11/2020	Third version	Eugenio Realini	Inputs on data sharing
00.00.04	27/11/2020	Fourth version	All partners	Overall review of the document
00.00.05	29/11/2020	Fifth version	Isabel Gomes	Overall review of the document
00.01.00	30/11/2020	Final version	Riccardo Biondi	

SINOPTICA

SATELLITE-BORNE AND IN-SITU OBSERVATIONS TO PREDICT THE INITIATION OF CONVECTION FOR ATM

Abstract

This document describes the Data Management Plan (DMP) of SINOPTICA project according to the Guidelines on Data Management in Horizon 2020 document [1]. This DMP describes the data management life cycle for all datasets to be collected, processed and generated by the SINOPTICA project during the research activity. In detail, the document identifies the data collected and generated, how it will be exploited and/or made accessible for verification and re-use, and how it will be curated and preserved.

Table of Contents

List of Figures

Executive Summary

The SINOPTICA Data Management Plan (DMP) describes the data management life cycle for all datasets to be collected, processed and generated by the SINOPTICA project during the research activity. In detail, the document identifies the data collected and generated, how it will be exploited and/or made accessible for verification and re-use, and how it will be curated and preserved.

According to the Guidelines and Data Management in Horizon 2020 recommendations, a comprehensive view of the data life cycle should be provided with appropriate definitions of terms being used and assumptions under which the plan has been devised, with a description of stakeholders groups. In this context, the present DMP provides information on the data sources in use and alternative data sources that may be used during the project. The list of data and data sources will be revised during the project lifetime. Other possible updates will be reflected in the iterated process for data provision, clarified in section 3 of this document.

Within the present DMP, datasets are considered to be collections of data that are generated, constructed or provided to the SINOPTICA consortium in order to achieve project's objectives and may be revised by exploiting data from a single or multiple data sources. Each dataset has a specific coverage, as well as specific spatial and temporal resolutions and may contain other features per data source. The DMP describes the data sources which will be potentially exploited at high level, also specifying information on already known metadata, scale, modality, and provision methods. Data sources will be described in detail accompanying every dataset to be generated, constructed or provided.

The present DMP also specifies a generic policy and methodology for associating data sources being used and datasets devised with metadata, together with a specific initial policy for version control. The specific aspects related to data integration for every dataset iteration - addressing the temporal and geographical coherence of each dataset - will be described in the document linked to every dataset iteration. Additionally, this DMP provides information on data sharing, taking into account partner's limitations, IPR, legal issues, and licensing. No ethical issues have been identified in this regard for the SINOPTICA project. Finally, the DMP concludes with issues concerning archiving and preservation of the data.

1 Introduction

1.1 Purpose of the Document

This document describes the data management processes that SINOPTICA project will implement in order to achieve its objectives, ensuring both the achievement of the project goals in terms of data usability and availability and in terms of research data quality, sharing and security.

The DMP explains how the data will be handled within the project frame, during the research and development phases and shows the intentions for archiving the data and making the data available once the project is completed. The DMP describes the overall context of SINOPTICA project research for what concerns data collection and assimilation for nowcasting purposes.

1.2 Relation to Work Packages and Deliverables

The deliverable is related to the WP3, WP4, WP5, WP6, WP7 and WP8. This section refers to three different data levels: Level S1 (data coming from third parties), Level S2 (intermediate products) and Level S3 (final products). For the details of the Level S1, S2 and S3 data, please check section 3 of this document.

- The WP3 (Dissemination and Communication) is responsible for the DMP.

- The WP4 (Users requirements analysis and KPIs) will use the Level S1 data to define the requirements of the SINOPTICA system from the end-user point of view.

- The WP5 (Forecast system demonstration and validation) will use the Level S2 data as an input to assess and validate the forecasting and nowcasting system.

- The WP6 (ATM integration) will use the Level S3 data created within the WP7, as an input to demonstrate their usefulness in different air traffic and weather scenarios available.

- The WP7 (Data assimilation system development) will get in input the Level S1 data and create the Level S2 and Level S3 data. The Level S2 are intermediate products, reusable high quality data and are themselves a valuable outcome of the project. These datasets will be accompanied by detailed data and metadata descriptions, easy to be read and visualized, shared with all the partners for inspection, evaluation and exploitation. The level S3 data are the final outcome of the project and an input for the WP4 and WP6.

- The WP8 (Data provision system development) will use the Level S1 data, downloading the datasets provided by third parties, archiving, reformatting and - when needed - processing the datasets to be provided as an input for the WP7. This WP will also check the quality of the data and homogenize the same type of data coming from different sources/networks.

1.3 Approach Taken and Structure of the Deliverable

The present DMP describes in detail the data sources to be exploited within the SINOPTICA project, the management of datasets to be constructed and the lifecycle of the SINOPTICA's generated data.

The consortium defines a *dataset* as being any coherent set of data from single or multiple sources, and newly generated data created within the SINOPTICA project. Datasets can be used for analysis, validation, reusability, dissemination, communication and demonstration purposes. Data sources provide input data to the SINOPTICA project enabling it to realize their functionalities and computing results.

During the project lifetimes, the SINOPTICA partners may decide to incorporate new data from not already known sources (or not taken into account while writing the present DMP), to enhance some of the existing datasets or to remove datasets from the data chain. The provided datasets will be continuously refined according to the results and to the partners needs and requirements. This iterative process is not intended as an update of the DMP, however, in case an updated version of the DMP will be considered necessary due to substantial changes, this deliverable will be updated accordingly.

Data sources provided from SINOPTICA partners or from third parties, satisfy specific properties and are subject to constraints and limitations for access and exploitation. Thus, they should follow specific IPR and legal restrictions which have been specified on the SINOPTICA Consortium Agreement, already signed by the partners, and subject to further refinement during the project duration, if necessary.

The SINOPTICA project aims at the development of research processes, an advanced forecasting of severe weather events affecting the air traffic and consequent provision of accurate risk polygons to the Air Traffic Management (ATM). These aspects will be rigorously tested and the results will be disseminated, communicated and exploited using specific datasets made available to the research community via widely-used and easily accessible data repositories. All the datasets used or research results produced by the project, will be described using specific meta-data.

This documents aims to describe the data used, the relation to the project Work Packages, and to explain how the data produced by the project will be made available to the different communities. After a general introduction, section 2 introduces the data life cycle, while section 3 will provide details of the data sources. Finally, section 4 will report the fair logic and section 5 will describe how the datasets will be shared, archived and preserved with the aim of making the data publicly available during and after the project.

2 Data life cycle

2.1 SINOPTICA data chain

The SINOPTICA project workflow is described in the following figure and consists of different tasks. The Weather Research and Forecasting (WRF) model is executed on high performance computing facilities after the preparation of initial and boundary conditions provided by the WRF Preprocessing System (WPS). The WPS task is usually executed on cloud computing facilities and it allows for the processing at the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) and the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) global circulation model data to generate input fields for the WRF model itself. The WRF Data Assimilation (DA) task is a flexible, state-of-the-art atmospheric data assimilation system that is portable and efficient on available parallel computing platforms: WRFDA is a task executed on cloud/high performance computing facilities and it allows to ingest observational data such local authoritative and personal weather observations, Global Navigation Satellite System (GNSS), GNSS-radio occultation and radar reflectivity. The WRF model output including vertically integrated liquid (VIL), vertical maximum intensity (VMI), echo top height, and rainfall intensity maps are blended with radar nowcasting fields and subsequently passed to Air Traffic Management (ATM) integration and Arrival MANager (AMAN) tasks.
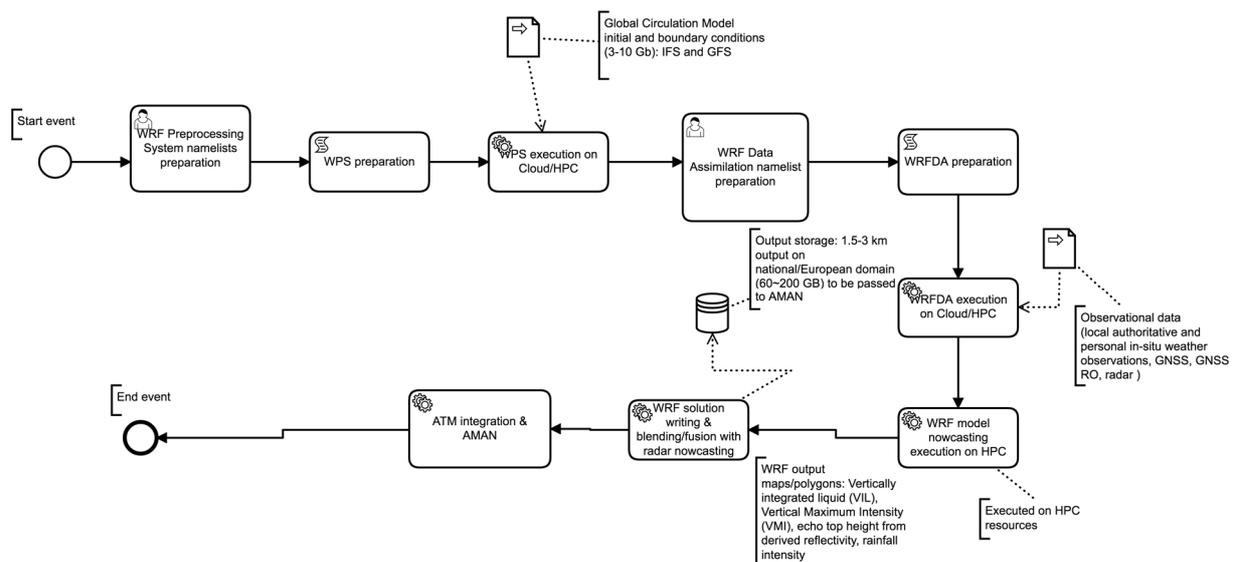

Figure 1. SINOPTICA data chain

In general, the Level S1 data are collected archived and reformatted by the WP8. The WP7 gets in input the Level S1 data provided by the WP8 and produces the Level S2 data and an intermediate product and the Level S3 data as a final product.

2.2 Stakeholders and constraints

Key stakeholders have been identified as having influence on the data management plan decisions, specific interests in data sources, research results and domain specific results.

Founding Members

Data providers. These represent the organizations which provide the datasets to be used within the SINOPTICA project. These parties are represented in the SINOPTICA project by the Italian Department of Civil Protection, WeatherUnderground, the International Geodetic Service (IGS), the University of Bern, the Istituto Nazionale di Geofisica e Vulcanologia (INGV), the National Aeronautics and Space Administration (NASA), the Crustal Dynamics Data Information System (CDDIS), the University Corporation for Atmospheric Research (UCAR) and the Radio Occultation Meteorology Satellite Application Facility (ROM SAF) as providers of the Level S1 datasets. The SINOPTICA partners play significant roles, by specifying the data sources to be used, defining the requirements from the data management components and reformatting the datasets according to the model requirements.

SINOPTICA partners. The SINOPTICA partners collaborate to specify the requirements for data management and analytics according to the project research objectives. The quality and the characteristics of the data provided by different sources is an important aspect for the overall success of the project. Thus, partners are key figures in evaluating processes quality and to produce intermediate and final products.

End users. These represent user-interest groups that focus on real-life, industrial and user-defined challenges concerning weather forecast and operations regarding moving entities in air. These are identified as Austro control (already partner of the SINOPTICA project) and other entities which will be identified during the project's life cycle.

3 SINOPTICA data sources

3.1 Target research context

The data used and produced by the SINOPTICA project can be distinguished in 3 different levels:

- **Level S1**: data acquired/downloaded from third party sources;

- **Level S2**: intermediate products which are produced by the project but they are also an input for the another stage of the SINOPTICA data chain;

- **Level S3**: Final results and output of the SINOPTICA project.

3.2 Atmospheric data sources

- **Dataset name: Ground based GNSS**

Project level data: Level S1

Data type, general data use and description
Observations including:
GNSS time, code, phase observations;
Orbits - GNSS satellites coordinates (IGS);
Clocks - GNSS satellites clock offsets (IGS);
Ionospheric maps - Predicted ionospheric delay maps (CODE/ESA);
Earth Rotation Parameters - Predicted Earth rotation parameters (CODE);
CRX files - GNSS satellites health status (AIUB);
Tropospheric mapping functions - Predicted Vienna Mapping Function grids (1deg x 1deg, TU Wien).

Metadata and format if available
Metadata: Not available.
Format: RINEX, SP3c, IONEX, ERP, CRX and GRID.

Scale
The data are 1D, provided as temporal series for each station. One year of all the mentioned products in the SINOPTICA project area of interest requires about 9GB of memory.

Provision method
Depending on the parameter they are freely available via FTP protocol, NTRIP protocol or HTTPS.

Available temporal and geographical coverage

Global coverage in the period 2000-present (varying significantly depending on the various networks archiving policies).

WP in which the data are used and connection to the other WPs

WP7 Data assimilation system development

WP8 Data provision system development

Contingency plan for non-availability

The data are retrieved from different networks of public domain such as International Geodetic Service (IGS), University of Bern, Istituto Nazionale di Geofisica e Vulcanologia (INGV), National Aeronautics and Space Administration (NASA), Crustal Dynamics Data Information System (CDDIS), etc. and GReD srl already has a backup of the relevant data for the project.

- **Dataset name: GNSS Radio Occultations**

Project level data: Level S1

Data type, general data use and description

Vertical profiles of bending angle, refractivity, temperature, and pressure from the missions COSMIC, COSMIC-2, PAZ, Kompsat, TerraSar, MetOp-A, MetOp-B and MetOp-C.

Metadata and format if available

Metadata information available at

https://cdaac-www.cosmic.ucar.edu/cdaac/cgi_bin/fileFormats.cgi?type=atmPrf

Format: Level 2 atmPrf data in Network Common Data Form (NetCDF) format.

Scale

The data are 4D, provided as temporal series of acquisitions with horizontal and vertical structure. One year of all the mentioned products in the SINOPTICA area of interest requires about 4GB of memory.

Provision method

All the data are provided via HTTPS at

https://data.cosmic.ucar.edu/gnss-ro/cosmic2/nrt/
https://cdaac-www.cosmic.ucar.edu/cdaac/rest/tarservice/data/

https://www.romsaf.org/pub/nrt/

Available temporal and geographical coverage

Global coverage in the period 2006-present.

WP in which the data are used and connection to the other WPs

WP7 Data assimilation system development

WP8 Data provision system development

Founding Members

Contingency plan for non-availability

There are seven data providers of GNSS RO data. The University of Padova has already established an agreement with the Wegener Center for Climate and Global Change (Austria) and with the Danish Meteorological Institute (Denmark) to get access to their RO archives.

- **Dataset name: GNSS Radio Occultations Anomalies**

Project level data: Level S2

Data type, general data use and description

Vertical profiles of bending angle anomalies. A first computation is made on the seasonal climatology of the bending angle in airport area. The next step consists to subtract the climatology to the actual profiles collocated with the study case and to observe the anomaly. The observed anomaly highlights the impact of the severe weather on the atmospheric structure provided to the model wich represents a very insightful information to improve the nowcast.

Metadata and format if available

Metadata:

```
 int altitude(altitude);
:standard_name = "height";
:long_name = "Altitude above geoid";
:units = "m";
:_CoordinateAxisType = "Height";

 double bending_angle(altitude, datetimeTC, numRO);
:long_name = "Ionosphere corrected non-optimized bending angle";
:units = "rad";
:_FillValue = -9999.0; // double

 double bending_angle_climatology(altitude, datetimeTC, numRO);
:long_name = "Climatological ionosphere corrected non-optimized bending angle";
:units = "rad";
:_FillValue = -9999.0; // double

 double bending_angle_anomaly(altitude, datetimeTC, numRO);
:long_name = "bending angle anomaly";
:units = "%";
:_FillValue = -9999.0; // double

 float latRO(datetimeTC, numRO);
:standard_name = "latitude";
```

```
:long_name = "Latitude of RO profile";
:units = "degrees_north";
:_FillValue = -9999.0f; // float


 float lonRO(datetimeTC, numRO);
:standard_name = "longitude";
:long_name = "Longitude of RO profile";
:units = "degrees_east";
:_FillValue = -9999.0f; // float


 int RO_datetime(datetimeSevereEvent, numRO);
:standard_name = "time";
:calendar = "standard";
:long_name = "Datetime of RO profile";
:units = "seconds since 1970-01-01 00:00:0.0";
:_FillValue = -9999; // int


char RO_ID(datetimeTC, charr=64, numRO);
:long_name = "ID of collocated Radio Occultation profile";
:standard_name = "RO_ID";


 int satellite(datetimeTC, numRO);
:long_name = "RO satellite name";
:flag_values = 1, 2, 3, 4, 5, 6, 7, 8; // int
:flag_meanings = "COSMIC, COSMIC-2, PAZ,  Kompsat, TerraSar, METOP-A, METOP-B, METOP-C";
```

Format: The data will be provided in Network Common Data Form (NetCDF) format.

Scale

The data are 4D, provided as temporal series of acquisitions with horizontal and vertical structure. One year of all the mentioned products in the SINOPTICA area of interest requires about 10MB of memory.

Provision method

The products are provided to the partners via sftp protocol.

Available temporal and geographical coverage

Global coverage in the period 2006-present

WP in which the data are used and connection to the other WPs

WP7 Data assimilation system development

WP8 Data provision system development

Contingency plan for non-availability

The dataset is produced within the SINOPTICA project.

- **Dataset name: Copernicus Sentinel EO**

Project level data: Level S1

Data type, general data use and description

Ocean Wind Field - Sentinel-1 Level 2 OCN Ocean Wind Field (OWI) Component

Land Surface Temperature - Sentinel-3 Level 2 LST product

Sea Surface Temperature - Sentinel-3 Level 2 WST product

Ground Range Detected (GRD) - Sentinel-1 Level 1 dual polarization (VV+VH) GRD product

Metadata and format if available

Metadata: Not available.

Format: Data provided in Network Common Data Form (NetCDF) and GeoTIFF formats.

Scale

The data are provided as temporal series of 2D acquisitions. One year of all the mentioned products in the SINOPTICA area of interest requires about 3.3TB of memory.

Provision method

All the data are provided via HTTPS at

https://scihub.copernicus.eu/

Available temporal and geographical coverage

Sentinel-1: Global coverage in the period 2014-present

Sentinel-3: Global coverage in the period 2016-present

WP in which the data are used and connection to the other WPs

WP7 Data assimilation system development

WP8 Data provision system development

Contingency plan for non-availability

Also Eumetsat provides the Sentinel-1 and Sentinel-3 products used for the SINOPTICA project.

3.3 Weather data sources

- **Dataset name: CAPPI (Constant Altitude Plan Position Indicator)**

Project level data: Level S2

Data type, general data use and description

The map is generated by combining the data contained in the vicinity of the intersection between the radar beams and the altitude set for the CAPPI. Each CAPPI value is obtained from the weighted average between the values corresponding to the two consecutive elevations immediately above and below the H quota of interest. CAPPI are available at 2000, 3000, 4000 and 5000 m.

Metadata and format

Metadata:

Format: 'tif'

Height: 2466

Width: 2681

BitDepth: 32

ColorType: 'grayscale'

ModelType: 'ModelTypeGeographic'

GCS: 'WGS 84'

Datum: 'World Geodetic System 1984'

Ellipsoid: 'WGS 84'

SemiMajor: 6378137

SemiMinor: 6.3568e+06

PM: 'Greenwich'

PMLongToGreenwich: 0

UOMAngle: 'degree'

UOMAngleInDegrees: 1

Format: The data are provided in geotiff.

Scale

Each file is about 300 Kb (500 m grid spacing)

Provision method

This is an intermediate product produced by the project and it is provided to the partners via sftp protocol.

Available temporal and geographical coverage

The data are available since july 2018 at 10 minutes temporal resolution.

Geographical coverage:

Longitude bounds [5.6 19.0]

Latitude bounds [35.25 47.58]

Founding Members

WP in which the data are used and connection to the other WPs

WP4 User requirements analysis & KPIs definition

WP5 Forecast system demonstration & validation

WP7 Data assimilation system development

WP8 Data provision system development

Contingency plan for non-availability

The dataset is produced within the SINOPTICA project.

- **Dataset name: VMI (Maximum vertical intensity)**

Project level data: Level S1

Data type, general data use and description

The VMI variable is the radar-detected maximum vertical reflectivity provided by the composite of the Italian National Radar Network. The data measured by the weather radars composing the national radar network are processed and mosaicked at DPC headquarters. The variable displayed by default is the last available image processed by the system.

Metadata and format

Metadata:

Format: 'tif'

Height: 2466

Width: 2681

BitDepth: 32

ColorType: 'grayscale'

ModelType: 'ModelTypeGeographic'

GCS: 'WGS 84'

Datum: 'World Geodetic System 1984'

Ellipsoid: 'WGS 84'

SemiMajor: 6378137

SemiMinor: 6.3568e+06

PM: 'Greenwich'

PMLongToGreenwich: 0

UOMAngle: 'degree'

UOMAngleInDegrees: 1

Format: The data are provided in geotiff.

Scale

Each file is about 300 Kb (500 m grid spacing)

Provision method

The products are provided via sftp protocol.

Available temporal and geographical coverage

The data are available since july 2018 at 5 minutes temporal resolution.

Geographical coverage:

Longitude bounds [5.6 19.0]

Latitude bounds [35.25 47.58]

WP in which the data are used and connection to the other WPs

WP4 User requirements analysis & KPIs definition

WP5 Forecast system demonstration & validation

WP7 Data assimilation system development

WP8 Data provision system development

Contingency plan for non-availability

The data are available on real-time manner, for research purposes, in the framework of the cooperation with the Italian Civil Protection Department that maintains the backup.

- **Dataset name: VIL (Vertical Integrated Liquid)**

Project level data: Level S2

Data type, general data use and description

The VIL (Vertically Integrated Liquid) contains a map that represents the total amount of liquid water (expressed in g/m^3) estimated within a column above the vertical of each pixel.

Given a volume of reflectivity, the algorithm consists in the integration for each radar cell of all the values recorded within a vertical interval above the cell. The extent of the vertical range depends on the sequence of elevations.

Metadata and format

Metadata:

Format: 'tif'

Height: 2466

Width: 2681

BitDepth: 32

ColorType: 'grayscale'

ModelType: 'ModelTypeGeographic'

GCS: 'WGS 84'

Datum: 'World Geodetic System 1984'

Ellipsoid: 'WGS 84'

SemiMajor: 6378137

Founding Members

SemiMinor: 6.3568e+06

PM: 'Greenwich'

PMLongToGreenwich: 0

UOMAngle: 'degree'

UOMAngleInDegrees: 1

Format: The data are provided in geotiff.

Scale

Each file is about 30 Kb (500 m grid spacing).

Provision method

This is an intermediate product produced by the project and it is provided to the partners via sftp protocol.

Available temporal and geographical coverage

The data are available since july 2018 at 5 minutes temporal resolution.

Geographical coverage:

Longitude bounds [5.6 19.0]

Latitude bounds [35.25 47.58]

WP in which the data are used and connection to the other WPs

WP4 User requirements analysis & KPIs definition

WP5 Forecast system demonstration & validation

WP7 Data assimilation system development

WP8 Data provision system development

Contingency plan for non-availability

The dataset is produced within the SINOPTICA project.

- **Dataset name: POH (Probability Of Hail)**

Project level data: Level S2

Data type, general data use and description

The Probability of Hail (POH) is a map containing the probability of the occurrence of hail. It is determined according to the difference between the altitude of the cloud top and the altitude of the freezing level, according to the following relationship:

POH = 0.000133 * (ETM-FLH) +0.319

Where ETM is the maximum altitude in which a signal greater than 45 dbZ is detected.

Metadata and format

Founding Members

Metadata:

Format: 'tif'

Height: 2466

Width: 2681

BitDepth: 32

ColorType: 'grayscale'

ModelType: 'ModelTypeGeographic'

GCS: 'WGS 84'

Datum: 'World Geodetic System 1984'

Ellipsoid: 'WGS 84'

SemiMajor: 6378137

SemiMinor: 6.3568e+06

PM: 'Greenwich'

PMLongToGreenwich: 0

UOMAngle: 'degree'

UOMAngleInDegrees: 1

Format: The data are provided in geotiff.

Scale

Each file is about 30 Kb (500 m grid spacing).

Provision method

This is an intermediate product produced by the project and it is provided to the partners via sftp protocol.

Available temporal and geographical coverage

The data are available since july 2018 at 5 minutes temporal resolution.

Geographical coverage:

Longitude bounds [5.6 19.0]

Latitude bounds [35.25 47.58]

WP in which the data are used and connection to the other WPs

WP4 User requirements analysis & KPIs definition

WP5 Forecast system demonstration & validation

WP7 Data assimilation system development

WP8 Data provision system development

Contingency plan for non-availability

Founding Members

The dataset is produced within the SINOPTICA project.

- **Dataset name: SRI (Surface Rainfall Intensity)**

Project level data: Level S1

Data type, general data use and description

The SRI variable is the radar-detected precipitation intensity provided by the composite of the Italian National Radar Network. The data measured by the weather radars composing the national radar network are processed and mosaicked at DPC-CFC. The variable displayed by default is the last available image processed by the system. The SRI maps are updated every 10 minutes and users can load each map within the time range.

Metadata and format

Metadata:

Format: 'tif'

Height: 2466

Width: 2681

BitDepth: 32

ColorType: 'grayscale'

ModelType: 'ModelTypeGeographic'

GCS: 'WGS 84'

Datum: 'World Geodetic System 1984'

Ellipsoid: 'WGS 84'

SemiMajor: 6378137

SemiMinor: 6.3568e+06

PM: 'Greenwich'

PMLongToGreenwich: 0

UOMAngle: 'degree'

UOMAngleInDegrees: 1

Format: The data are provided in geotiff.

Scale

Each file is about 30 Kb (500 m grid spacing)

Provision method

The products are provided via sftp protocol.

Available temporal and geographical coverage

The data are available since july 2018 at 5 minutes temporal resolution.

Geographical coverage:

Longitude bounds [5.6 19.0]
Latitude bounds [35.25 47.58]

WP in which the data are used and connection to the other WPs
WP4 User requirements analysis & KPIs definition
WP5 Forecast system demonstration & validation
WP7 Data assimilation system development
WP8 Data provision system development

Contingency plan for non-availability
The data are available on real-time manner, for research purposes, in the framework of the cooperation with the Italian Civil Protection Department that maintains the backup.

- **Dataset name: Italian authoritative weather stations**

Project level data: Level S1

Data type, general data use and description
About 5000 weather stations (2m temperature, 10 m wind, 2 m dew point temperature) available over the Italian territory

Metadata and format
Metadata: Not available.
Format: The data are provided in ascii or netcdf-cf.

Scale
It depends on the selected time window.

Provision method
The products are provided via sftp protocol.

Available temporal and geographical coverage
The data are available since july 2018 at 10 minutes temporal resolution.
Geographical coverage:
Longitude bounds [5.6 19.0]
Latitude bounds [35.25 47.58]

WP in which the data are used and connection to the other WPs
WP4 User requirements analysis & KPIs definition
WP5 Forecast system demonstration & validation
WP7 Data assimilation system development

WP8 Data provision system development

Contingency plan for non-availability

The data are available on real-time manner, for research purposes, in the framework of the cooperation with the Italian Civil Protection Department that maintains the backup.

- **Dataset name: European personal weather stations**

Project level data: Level S1

Data type, general data use and description

About 32000 weather stations (2m temperature, 10 m wind, 2 m dew point temperature) available over the Europe territory

Metadata and format

Metadata: Not available.

Format: the data are provided in ascii or netcdf-cf.

Provision method

The products are provided via sftp protocol.

Scale

It depends on the selected time window.

Provision method

The products are provided via sftp protocol.

Available temporal and geographical coverage

The data are available since july 2018 at 60 minutes temporal resolution over entire Europe.

WP in which the data are used and connection to the other WPs

WP4 User requirements analysis & KPIs definition

WP5 Forecast system demonstration & validation

WP7 Data assimilation system development

WP8 Data provision system development

Contingency plan for non-availability

The data are available on real-time manner, for research purposes, in the framework of the cooperation with Wunderground.

- **Dataset name: lightning strikes**

Project level data: Level S1

Data type, general data use and description

This layer offers the localization and the intensity of the lightnings strikes observed through the SFLOC method (acronym for 'Sferic location'). A 'spherical' interference is caused by radio waves emitted by a lightning or a thunder. From the measurement of the time of arrival of the interference it can be inferred, through suitable sensors, the location of the phenomenon. The observational network -called Lampinet- became operational in 2004 and currently consists of 15 sensors VAISALA IMPACT ESP distributed on the whole italian territory. In its current configuration, Lampinet measures both cloud-to-cloud and cloud-and-ground lightnings, with an accuracy of detection higher than 90% for amperage higher than 5 kA and a 500-meter localization precision.

Metadata and format

Metadata: Not available.

Format: The data are provided in ascii or netcdf-cf.

Scale

It depends on the selected time window.

Provision method

The products are provided via sftp protocol.

Available temporal and geographical coverage

The data are available since july 2018 at 5 minutes temporal resolution.

Longitude bounds: [5.6 19.0]

Latitude bounds: [35.25 47.58]

WP in which the data are used and connection to the other WPs

WP4 User requirements analysis & KPIs definition

WP5 Forecast system demonstration & validation

WP7 Data assimilation system development

WP8 Data provision system development

Contingency plan for non-availability

The data are available on real-time manner, for research purposes, in the framework of the cooperation with the Italian Civil Protection Department that maintains the backup.

3.4 ATM data sources

Dataset name: Georeferenced polygons of extreme weather areas measurement and nowcast

Project level data: Level S3

Data type, general data use and description

For the deviation approach trajectory calculation, position, shape, lifespan and nowcast data of extreme weather areas around airports are necessary. The data are used to generate realistic trajectories for individual flights in the actual traffic context including advisories (clearances) for pilots, target times at significant waypoints like threshold, position in the actual arrival sequence and the animated visualization on air traffic controllers' radar displays as part of controller working positions. Thus, data must be periodically updated with an ideal frequency of five or ten minutes.

For the use of weather data, a distinction is made between required and optional data: while **required data** are essential for the deviation-trajectory calculation and for displaying extreme weather areas on a controller's radar display, the **optional data** covers information needed to extend some AMAN functionalities.

Metadata and format

Polygon of GPS coordinates encircle extreme weather area. The last point of the polygon is identical with the first one.

- Aeronautical severity level of the polygon
- Valid-from time (in seconds)
- Valid-to time (in seconds)
- GPS-position of the geographical center of gravity of the polygon (extreme weather area)
- Actual moving speed of the center of gravity (meter per second or knots)
- Actual moving direction of the center of gravity (degree)
- Minimum and maximum altitude of the area covered by the extreme weather polygon (meter of feet)
- Unique identification number of every polygon

Format: The data file format is open as the interface can be adapted. The preferred formats will be comma separated values (CSV) or extended markup language (XML).

Scale

There are no requirements for a minimum or maximum size of the weather data files.

Provision method

The products are provided via sftp protocol.

Available temporal and geographical coverage

The weather data should ideally cover the area around the airport to be studied within the project. The distance of the coverage from the center (reference point) of the airport will be at least 180 kilometers (around 100 miles) uninterrupted within this area.

The time coverage of an extreme weather event will be more than one hour and less than 2 hours. This corresponds to the typical traffic scenarios as they are usually used in automatic and human-in-the-loop simulations. This scenario length allows the verification and validation of the trajectories due to extreme weather conditions diverted approaches procedures and air traffic controller support functionalities.

WP in which the data are used and connection to the other WPs

The data is used in WP6 for the verification for in the project developed displays and AMAN guidance support functionalities.

Contingency plan for non-availability

In the case that the data is not available in time or not in the desired quality or quantity, older weather data can be used. These do not necessarily have to be measured in the project target area, as they can be shifted to the pilot-airport and traffic scenarios by means of a temporal and spatial translation. In case no suitable weather data is available at all, artificial weather could be generated using Excel. DLR has already gained experience with the preparing of artificially generated weather situations, as this was needed in the past for software verifications.

4 FAIR data

Data should follow the FAIR logic, so they have to be Findable, Accessible, Interoperable and Reusable (FAIR).

4.1 Making data findable

To allow project data to be findable, data produced in the SINOPTICA project will be discoverable with metadata.

The naming conventions will be as follows:

- CAPPI (Constant Altitude Plan Position Indicator)

CAPPI_ dddddNdddddW _ YYDDDhhmm _V.DDMMYY.tif
Where dddddNdddddW represents the airport coordinates with ddddd is a 5-digit degrees *100, N is North and W is West; YYDDDhhmm represents the event starting date/time with YY is the 2-digit year, DDD is the day of the year, hh is the starting hour of the event and mm is the starting minute; V.DDMMYY represents the product Version (V) with the date in which the product was created where DD is the day, MM is the month and YY is the 2-digit year.

- VIL (Vertical Integrated Liquid)

VIL_ dddddNdddddW _ YYDDDhhmm _V.DDMMYY.tif
Where dddddNdddddW represents the airport coordinates with ddddd is a 5-digit degrees *100, N is North and W is West; YYDDDhhmm represents the event starting date/time with YY is the 2-digit year, DDD is the day of the year, hh is the starting hour of the event and mm is the starting minute; V.DDMMYY represents the product Version (V) with the date in which the product was created where DD is the day, MM is the month and YY is the 2-digit year.

- POH (Probability Of Hail)
POH_ dddddNdddddW _ YYDDDhhmm _V.DDMMYY.tif
Where dddddNdddddW represents the airport coordinates with ddddd is a 5-digit degrees *100, N is North and W is West; YYDDDhhmm represents the event starting date/time with YY is the 2-digit year, DDD is the day of the year, hh is the starting hour of the event and mm is the starting minute; V.DDMMYY represents the product Version (V) with the date in which the product was created where DD is the day, MM is the month and YY is the 2-digit year.

- Georeferenced polygons of extreme weather areas measurement and nowcast

POLYGON_ dddddNdddddW _ YYDDDhhmm _V.DDMMYY.csv
Where dddddNdddddW represents the airport coordinates with ddddd is a 5-digit degrees *100, N is North and W is West; YYDDDhhmm represents the event starting date/time with YY is the 2-digit year, DDD is the day of the year, hh is the starting hour of the event and mm is the starting minute; V.DDMMYY represents the product Version (V) with the date in which the product was created where DD is the day, MM is the month and YY is the 2-digit year.

Metadata will be created and are described in section 4 for each product.

4.2 Making data openly accessible

Some of the data Level S1 are publicly and freely accessible from the original sources (reported in section 3) and in this case SINOPTICA project will not duplicate its accessibility. Other Level S1 data are available to the project through a specific agreement between the SINOPTICA project and third parties; in this case, data will not be made accessible.

All the data (Level S1, Level S2 and Level S3) of a relevant list of case studies will be made freely available at the GFZ Data Services (https://dataservices.gfz-potsdam.de/portal/), assigning a Digital Object Identifiers (DOI) to the dataset and making the dataset citable as any other scientific article. All data will be provided with a Creative Commons licence that is indicated on the DOI Landing Page and the metadata will be in the public domain.

The dataset description, including metadata and possible data use, will be published in the Earth System Science Data (ESSD) Journal edited by Copernicus Publications, which is an open access journal with impact factor 9,197.

Data processing software will not be made available as open source code, since the SINOPTICA project has a strong market uptake aim.

4.3 Making data interoperable

Data produced within the SINOPTICA project will adhere to widely adopted standard formats, in particular facilitating the compatibility with available open applications. In this respect, all the SINOPTICA generated data will be interoperable. Standard and open formats will be chosen whenever possible. In any case, format specifications will be defined and provided.

4.4 Increase data re-use

Openly available data will be licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Data will be made available for re-use after the end of the SINOPTICA project (December 2022) and are intended to remain re-usable indefinitely.

Founding Members

5 Data sharing

A number of datasets will be used during the SINOPTICA project. The WP7 will produce an amount of research results which will be widely shared, with respect to legal and IPR constraints, by specifying how, where and to whom the data could be made available.

As reported in section 4.2, some of the Level S1 data are publicly and freely accessible from the original sources and, in this case, the SINOPTICA project will not duplicate its accessibility. Other Level S1 data are available to the project through a specific agreement between SINOPTICA and third parties which, will not be made accessible. All the data (Level S1, Level S2 and Level S3) of a relevant list of case studies will be made freely available at the GFZ Data Services (https://dataservices.gfz-potsdam.de/portal/), assigning a Digital Object Identifiers (DOI) to the dataset and making the dataset citable as any other scientific article. All data will be provided with a Creative Commons licence that is indicated on the DOI Landing Page and the metadata will be in the public domain.

Specifically regarding the IPR and legal restrictions, we refer in general to the SINOPTICA Consortium agreement, with some additional detail reported in the following table:

Dataset	Limitation and conditions for implementation	Limitation and conditions for dissemination and exploitation
European personal weather stations	Access rights to background is only granted to the extent that is needed for implementation of the project	Can be disseminated following The agreement and license rules
Italian authoritative weather stations	Access rights to background is only granted to the extent that is needed for implementation of the project	Not available for dissemination and exploitation
VMI (Maximum vertical intensity)	Access rights to background is only granted to the extent that is needed for implementation of the project	Not available for dissemination and exploitation
SRI (Surface Rainfall Intensity)	Access rights to background is only granted to the extent that is needed for implementation of the project	Not available for dissemination and exploitation
Lightning strikes	Access rights to background is only granted to the extent that is needed for implementation of the project	Not available for dissemination and exploitation

All the data to be shared within the project partners, will be stored on state of the art computing and storage system managed by the CIMA Research Foundation (project coordinator) corresponding to the following European Tender https://ted.europa.eu/udl?uri=TED:NOTICE:561171-2020:TEXT:EN:HTML&tabId=0 under the Directive 2014/24/EU.

6 Archiving and preservation

The SINOPTICA project aims to address issues concerning long term storage of datasets as well as research generated data. These issues relate to safety and long term data storage, which are complex to manage considering that data evolves and could be merged with other data coming from other sources. Within such context and within the SINOPTICA project, it is crucial to consider the volume of the data to be archived, to define what "long-term" means and what datasets can be relevant for future studies.

Preserved datasets need to be updated, under a data preservation policy and process clearly identified. A long term preservation has a cost which depends on the real use and exploitation after the end of the project itself. Such costs are mainly related to: personnel time for data management, documentation and preservation; hardware and software needed for data management, backup and security, documentation and preservation. These topics will be discussed during the project meetings and conclusions or recommendations will be reported in an updated DMP.

As already reported in the previous section, all the data (Level S1, Level S2 and Level S3) of a relevant list of case studies will be archived and preserved at the GFZ Data Services (https://dataservices.gfz-potsdam.de/portal/), under a Creative Commons licence. Regarding GNSS, GNSS-RO and Sentinel data, that could be needed in WP8 during the project duration, are automatically downloaded by tasks scheduled to run daily on GReD servers. Iterative download procedures will be implemented in the event that a dataset cannot be partially accessed at a given time (e.g. due to network outages, remote server downtime, and so on). All downloaded data are archived on GReD NAS storage system (32 TB total storage space with 12 disks in RAID-6 configuration to guarantee data availability). GNSS raw data, being retrieved from different data providers (i.e. different GNSS stations networks operators in Italy), are quality checked and their format is homogenized.

The Radar data, lightning data, weather stations data, the WRF model outputs (Level S2) including data assimilation, and the polygons (Level S3), will be stored at CIMA Research Foundation on secured servers and protected by Redundant Array of Independent Disks (RAID) technology. The access to the servers will be possible via VPN for all the partners with a personal login interface.

7 References

[1] SJU, Project Execution Guidelines for SESAR 2020 Exploratory Research, Edition 01.00.00, 07/12/2016

[2] SINOPTICA, SINOPTICA 892362 Consortium Agreement, v1.0

[3] Participant Portal H2020, Online Manual, https://ec.europa.eu/research/participants/docs/h2020-funding-guide/index_en.htm

[4] European Commission, Guidelines on Data Management in Horizon 2020, https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

[5] SINOPTICA, SINOPTICA 892362 Grant Agreement, v1.0