

---

# "Sensor Augmented weather Prediction at HIgh Resolution" [ANR-21-CE04-0014-03] project DMP

*Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "ANR - DMP template (english)" fourni par Agence nationale de la recherche (ANR).*

## Plan Details

<b>Plan title</b>	"Sensor Augmented weather Prediction at HIgh Resolution" [ANR-21-CE04-0014-03] project DMP				
<b>Deliverable</b>	Deliverable 1.1.1				
<b>Version</b>	First version				
<b>Plan purpose/scope</b>	<p>This initial version of the SAPHIR Data Management Plan (DMP) is the first deliverable of the work package 1.1 "Data Base design and Data Management Plan". This document has been elaborated by the Data Managers together with all partners, in the first six months of the SAPHIR project. We consider it is an ongoing process and a useful tool for scientific discussions on all research Data aspects.</p> <p>This DMP will be continuously adapted and updated during the project duration with an update at T24 and a definitive version at the end of the project (T48), considering especially the sharing and preservation of SAPHIR research data after the end of the project. We will pay close attention to the integration of FAIR (Findable, Accessible, Interoperable and Reusable) principles to the DMP redaction.</p> <p>The structure of the document follows the ANR DMP template.</p>				
<b>Fields of science and technology (from OECD classification)</b>	Earth and related environmental sciences, Physical sciences, Computer and information sciences				
<b>Language</b>	eng				
<b>Creation date</b>	2022-02-04				
<b>Last modification date</b>	2022-07-06				
<b>License</b>	<table><tr><td><b>Name</b></td><td>Creative Commons Attribution Non Commercial Share Alike 4.0 International</td></tr><tr><td><b>URL</b></td><td><a href="http://spdx.org/licenses/CC-BY-NC-SA-4.0.json">http://spdx.org/licenses/CC-BY-NC-SA-4.0.json</a></td></tr></table>	<b>Name</b>	Creative Commons Attribution Non Commercial Share Alike 4.0 International	<b>URL</b>	<a href="http://spdx.org/licenses/CC-BY-NC-SA-4.0.json">http://spdx.org/licenses/CC-BY-NC-SA-4.0.json</a>
<b>Name</b>	Creative Commons Attribution Non Commercial Share Alike 4.0 International				
<b>URL</b>	<a href="http://spdx.org/licenses/CC-BY-NC-SA-4.0.json">http://spdx.org/licenses/CC-BY-NC-SA-4.0.json</a>				
<b>Associated documents (publications, reports, patents, experimental plan...), website</b>	<ul style="list-style-type: none"><li>Project website : <a href="#">SAPHIR website</a></li><li>SAPHIR Meta Data repository : <a href="#">SAPHIR Dataverse</a></li></ul>				

## Project Details

<b>Project title</b>	Sensor Augmented weather Prediction at HIgh Resolution" [ANR-21-CE04-0014-03]
<b>Acronym</b>	SAPHIR
<b>Abstract</b>	<p>The project "Sensor Augmented weather Prediction at HIgh Resolution" (SAPHIR) takes place within the fast growing and very promising context of data-driven weather forecasting methods. SAPHIR project consists in exploiting the outputs of a high-resolution NWP system and a variety of (real-time or post-processed) data gathered from meteorological safety sensor networks, satellite observations and also from ad hoc local sensor networks through a "Deep Learning" architecture. We will focus more precisely on two different outcomes at some specific locations in Corsica, namely the prediction of episodes of intense thunderstorms with significant rainfalls and lightning activity and wind/solar power forecasting (see Fig. 1 below). Indeed, severe weather conditions and notably thunderstorms and flooding are among one of the most dangerous weather risks in the Mediterranean region. Even if warning systems based on operational NWP forecasts provide reliable recommendations, accurate predictions at high resolution of the occurrence, location, timing and intensity of extreme events remain very challenging especially in a complex mountainous and maritime region such as the island of Corsica. This issue is especially critical since the global warming notably increases the occurrence of extreme weather events. In the field of renewable energy (RE), the power delivery of wind farms and solar power plants is strongly dependent on the volatile energy resource and its intermittency. As a consequence, it is characterized by a large amount of uncertainty that may be costly and difficult to manage for energy suppliers. However, the share of electricity produced by photovoltaics and wind energy systems increases and according to France Multi Annual Energy Plan, the installed capacity for renewable electricity in 2028 will be doubled as compared to 2017. Then, accurate wind speed and solar irradiance predictions and thereby forecast of power delivery are more and more important for the energy managers in order to make these intermittent energy sources amenable to integration into a power grid. Our ambition in that respect will be (i) to study the possibility of improving current predictions at horizons from few tens of minutes to 1 day, (ii) to consider microgrid flow control issues associated with such predictions and (iii) to study the possibility to predict extreme wind events in order to mitigate their impact on energy production.</p>
<b>Funding</b>	<ul style="list-style-type: none"> <li>• Agence Nationale de la Recherche : ANR-21-CE04-0014-03</li> </ul>
<b>Start date</b>	2022-01-06
<b>End date</b>	2026-02-05
<b>Partners</b>	<ul style="list-style-type: none"> <li>• University of Corsica Pasquale Paoli, UMR CNRS 6134 "Sciences pour l'Environnement" (SPE) <a href="#">168670399</a></li> <li>• Inria Paris Research Center <a href="#">86790</a></li> <li>• Laboratory of Aerology, LAERO UMR 5560 <a href="#">150195117</a></li> </ul>

#### **Research outputs :**

1. SAPHiR data (Dataset)

## Contributors

Name	Affiliation	Roles
Bosseur Frédéric - <a href="https://orcid.org/0000-0002-8108-0887">https://orcid.org/0000-0002-8108-0887</a>	SPE UMR 6134 CNRS, University Of Corsica Pasquale Paoli - 200012203U	
Filippi Jean-Baptiste - <a href="https://orcid.org/0000-0002-6244-0648">https://orcid.org/0000-0002-6244-0648</a>	SPE UMR 6134 CNRS, University Of Corsica Pasquale Paoli - 200012203U	
Muzy Jean-François - <a href="https://orcid.org/0000-0003-0956-7874">https://orcid.org/0000-0003-0956-7874</a>	SPE UMR 6134 CNRS, University Of Corsica Pasquale Paoli - 200012203U	• Project coordinator
Nivet Marie-Laure - <a href="https://orcid.org/0000-0003-1425-4450">https://orcid.org/0000-0003-1425-4450</a>	SPE UMR 6134 CNRS, University Of Corsica Pasquale Paoli - 200012203U	• DMP manager
Paoli Christophe - <a href="https://orcid.org/0000-0002-3093-1119">https://orcid.org/0000-0002-3093-1119</a>	SPE UMR 6134 CNRS, University Of Corsica Pasquale Paoli - 200012203U	

### Droits d'auteur :

Le(s) créateur(s) de ce plan accepte(nt) que tout ou partie de texte de ce plan soit réutilisé et personnalisé si nécessaire pour un autre plan. Vous n'avez pas besoin de citer le(s) créateur(s) en tant que source. L'utilisation de toute partie de texte de ce plan n'implique pas que le(s) créateur(s) soutien(nen)t ou aient une quelconque relation avec votre projet ou votre soumission.

# "Sensor Augmented weather Prediction at High Resolution" [ANR-21-CE04-0014-03]" project DMP

---

## 1. Data description and collection or re-use of existing data

### 1a. How will new data be collected or produced and/or how will existing data be re-used?

A part of the data will be produced ex nihilo and collected in various ways:

- Ground measurement campaigns based on sensors (cup or sonic anemometers, thermocouples, solar measuring PV cells, micro rain radar, rain gauges, disdrometers,...) installed over a set of chosen sites. Eventually, a detailed interactive map will be produced and available on the web.
- Computed outputs of the high-resolution NWP meteorological research code [Meso-NH](#), including the ground fields of temperature (°C), pressure (Pa), water vapour content (kg/kg), cumulative rain precipitation (mm), turbulent kinetic energy (m<sup>2</sup>s<sup>-2</sup>) and horizontal wind components (km/h).
- Computed outputs of the SAPHiR forecasting models under development.

Another part are existing data with different uses:

- Ground meteorological observation data made available under strict conditions by the French national meteorological service or others national meteorological services;
- Ground meteorological observation data recorded in Ajaccio, France, on the SPE UMR CNRS 6134 lab main building;
- European Centre for Medium-Range Weather Forecasts (ECMWF) operational forecasts;
- Time and space locations of VHF sources that are radiated by lightning flashes in the 60-66 MHz VHF band;
- Data provided by [Vigicrues](#) available on many Corsican rivers and historical record of recent rain events.

This list of data is not exhaustive and may be further expanded with other data types provided by our consortium partners. The metadata of new or existing SAPHiR data will be referenced in a repository for research data as the Harvard Dataverse [SAPHiR](#), making our data more discoverable to the research community.

---

### 1b. What data (for example the kind, formats, and volumes), will be collected or produced?

Most of the data manipulated in the SAPHiR project will be in digital format.

They will be stored in agreed weather common formats and standards: csv, NetCDF4 (Network Common Data Form) which contains a straightforward description of the contained data (matrices tensors dimensions, variables names and stored measure units), and GRIB (GRIdded Binar).

We estimate the need for data delivered by various data sources throughout the project duration to be at least 20 Tb of data storage (sensor data, specific data collected from existing sources and data generated within SAPHiR WPs) and 100 Tb of archiving capacity (NWP model outputs).

---

## 2. Documentation and data quality

### 2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

Our medium-term objective concerning SAPHiR data is to be hosted on the [AERIS web platform](#). AERIS is an atmosphere Data and Services Centre that federates national data management activities and atmospheric science expertise. Subsequently, we choose the AERIS metadata standard to describe our data sets.

Pending the acceptance of the SAPHiR project by AERIS, we first create a [SAPHiR Dataverse collection](#) to allow partners to rigorously describe their research data sets. Data set creation and deposit in this research data repository will enhance the web visibility and academic credit of the data, both inside and outside the SAPHiR project.

---

## **2b. What data quality control measures will be used?**

For data issued from Ground measurement campaigns, data consistency and quality will be defined and controlled depending on the type of data and the physical signal measured.

Indeed, the concept of data quality depends on many factors, relating to the research objectives and methodologies used. Ad hoc quality control procedures will be used and described throughout the project duration.

In all the cases, the raw data will be stored and qualified with metrics collectively determined, according to community standards (amount of missing data, duration of the lack of data,...). Raw data without processing will also be saved.

---

## **3. Storage and backup during the research process**

### **3a. How will data and metadata be stored and backed up during the research?**

Pending the acceptance of the SAPHIR project by AERIS we use a Dataverse instance to publish [SAPHIR metadata](#).

Concerning the data themselves, they will be stored on the SPE laboratory server with an incremental automatic daily backup provided by the University of Corsica IT support service.

---

### **3b. How will data security and protection of sensitive data be taken care during the research**

The University of Corsica IT support service provides us a data storage and recover service throughout SAPHIR project duration.

Different cases must be considered concerning the access to the data:

- Ground measurement campaigns and SAPHIR forecasting will be upload to a website dedicated to SAPHIR data: data collected or produced over the course of the research will be freely accessible to everyone on condition that due acknowledgement is made.
  - Reused data set will be accessible after authentication (login/psw) to the only partners.
- 

## **4. Legal and ethical requirements, code of conduct**

### **4a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?**

No personal data are processed.

---

### **4b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?**

Data produced during the SAPHIR project will be openly accessible according to the principles "as open as possible and as closed as necessary" as promoted by ANR in its Open Research Data orientation initiative.

---

#### **4c. What ethical issues and codes of conduct are there, and how will they be taken into account?**

Scientists involved in SAPHiR project pay particular attention to research integrity promoted by [The European Code of Conduct for Research Integrity](#) and the [french national charter for research integrity promoted by CNRS](#).

---

## **5. Data sharing and long-term preservation**

#### **5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?**

First, we created a [SAPHIR Dataverse collection](#) to allow partners to rigorously describe their research data sets. All the data sets will be accessible and visible both inside and outside the SAPHiR project.

At the end of the SAPHiR project, AERIS website will host our data sets, ensuring and maximising FAIR principles respect. Currently, no embargo period is planned for data produced throughout the project duration.

---

#### **5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?**

No decision on data preservation selection has been made for the moment. This point will be discussed among partners. Despite this, we have already identified [AERIS](#), the Data and Service for the Atmosphere Center as the best solution for long-term data preservation. We will submit a data hosting application to AERIS scientific comity.

---

#### **5c. What methods or software tools are needed to access and use data?**

[Dataverse](#) repository will be used for metadata exposition and storage.

Ground measurement data will be stored in an open-source Time Series Database [Influx DB](#), and PostgreSQL database systems.

Numerical Weather Prediction input and output will be directly stored on the high-performance computer data directories and be accessible by classic FTP.

[Grafana](#) an open-source analytics and monitoring solution will be used to query and visualize times series.

Data will be stored according to standard format (csv, NetCDF,...). The re-use of data will be possible using the standard software or library (Python xarray, Pandas, MATLAB, spreadsheet software, ParaView, etc.).

Ad hoc Python code will be provided to facilitate the re-use of complex re-aggregated data (input and output of Original Forecasting SAPHiR models)

---

#### **5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?**

Every set of data will be declared and published on the [SAPHIR Dataverse](#) and - at the end of the project - on [AERIS](#). Both these repositories automatically assign an unique Digital Object Identifier (DOI) which is a persistent identifier.

---

## **6. Data management responsibilities and resources**

**6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?**

The ground measurement campaigns will be under the operational responsibility of Mr. Xavier SILVANI ([silvani\\_x@univ-corse.fr](mailto:silvani_x@univ-corse.fr)) research engineer at SPE Laboratory.

Each partner will have a designated data manager in charge of data inventory and responsible for guaranteeing the availability of his data. The quality procedures are discussed between all the data expert partners and are specific to the different data.

Mr. Frédéric Bosseur BOSSEUR ([bosseur\\_f@univ-corse.fr](mailto:bosseur_f@univ-corse.fr)) research engineer in charge at SPE Laboratory of hight performance computing facilities will be in charge of overseeing the Data Research Infrastructure support. The two heads of the DMP are Mr. Christophe PAOLI ([paoli\\_c@univ-corse.fr](mailto:paoli_c@univ-corse.fr)) and Ms. Marie-Laure NIVET ([nivet\\_m@univ-corse.fr](mailto:nivet_m@univ-corse.fr)) both senior lecturers at SPE Laboratory. They will jointly assume the role of data stewards and ensure the regular updating of the DMP.

Archiving will be managed jointly by the IT facilities of the University of Corsica and AERIS.

---

**6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?**

Question sans réponse.