"ERORUN-STAFOR OBSERVATORY DATA MANAGEMENT PLAN FOR CRITICAL ZONE DATASETS (REUNION ISLAND)"

RESEARCH DATA MANAGEMENT PLAN CREATED USING DMP OPIDOR, (FRENCH ANR - DMP TEMPLATE (ENGLISH) TEMPLATE)

PLAN DETAILS

Plan title	ERORUN and STAFOR observatory DMP (Reunion island)
Fields of science and technology (from OECD classification)	1. Natural sciences / 1.5 Earth and related Environmental sciences
Language	eng
Creation date	2023-10-20
Last modification date	2024-03-04

PROJECT DETAILS

Project title	ERORUN-STAFOR observatory DMP for critical zone datasets
Acronym	ERORUN-STAFOR CZ DMP

Victor K/BIDI[1], Jean-Lambert JOIN[2], Alix TOULIER[2,3], Pierre STAMENOFF[1], Yoan BENOIT[1], Sophie FERREIRA[1], Magali DAVID[1], Emmanuel CORDIER[1], Rémy DECOUPES[1], Geneviève LEBEAU[2], Claudine AH-PENG[4], Laurent MICHON[2], Bhavani BENARD[2], Olivier MAGAND[1]

[1] Observatoire des Sciences de l'Univers de La Réunion (OSU-Reunion), UAR 3365, Université de La Réunion, CNRS, Météo France, IRD, 97744 Saint-Denis, La Réunion, France
[2] Université de La Réunion, Laboratoire GéoSciences Réunion, F-97744, Saint Denis, La Réunion, France
[3] Water Resources Research Center, University of Hawai'i at Manoa, Honolulu, HI, 96848 USA
[4] UMR PVBMT, Université de la Réunion, 97410 Saint Pierre, La Réunion, France

Victor K/bidi ORCiD : https://orcid.org/0009-0005-0522-9941
Jean-Lambert Join ORCiD : https://orcid.org/0000-0002-2321-5341
Alix Toulier ORCiD : https://orcid.org/0009-0003-1179-5013
Emmanuel Cordier ORCiD : https://orcid.org/0000-0001-5194-4026
Claudine Ah-Peng PersonId : 969161
Laurent Michon ORCiD : https://orcid.org/0000-0002-4486-5378
Bhavani Benard ORCiD : https://orcid.org/0000-0002-9512-529X
Olivier Magand ORCiD : https://orcid.org/0000-0001-8711-4155

Contributors and roles

Name	Affiliation	Roles
K/BIDI Victor	OSU-Reunion	Data steward Data contact
JOIN Jean-Lambert	LGSR	Principal Investigator (PI)
TOULIER Alix	LGSR, Univ. Hawai'i	Co-PI
STAMENOFF Pierre	OSU-Reunion	Assistant Technical coordinator
BENOIT Yoan	OSU-Reunion	Technical operator (field, laboratory and instrumentation)
FERREIRA Sophie	OSU-Reunion	Technical operator (field, laboratory and instrumentation)
DAVID Magali	OSU-Reunion	Technical operator (field, laboratory and instrumentation)
CORDIER Emmanuel	OSU-Reunion	Data steward (2019-2023)
DECOUPES Rémy	OSU-Reunion	Data steward (2014-2019)
LEBEAU Geneviève	LGSR	Technical operator (field, laboratory and instrumentation)
AH-PENG Claudine	PVBMT	Science coordinator
MICHON Laurent	LGSR	Science coordinator
BENARD Bhavani	LGSR	Science coordinator
MAGAND Olivier	OSU-Reunion	Technical and Science coordinator

ABOUT ERORUN-STAFOR OBSERVATORY DMP

The ERORUN-STAFOR Observatory Data Management Plan (DMP) is related to the French network of critical zone observatories (OZCAR) located in an insular tropical and volcanic context, integrating a “Tropical Mountain Cloud Forest " (TMCF). This collaborative observatory is located in the northern part of La Réunion island (Indian Ocean) within the watershed of Rivière des Pluies (45.0 km²) which hosts the TMCF of Plaines des Fougères, one of the best preserved natural habitats in La Réunion Island. Since 2015, the ERORUN-STAFOR monitoring in collaboration with local partners collected a multidisciplinary dataset with a constant improvement of the instrumentation over time. At the watershed scale and in its vicinity, the ERORUN-STAFOR Observatory includes 10 measurement stations covering the upstream, midstream and downstream part of the watershed. The stations record a total of 48 different variables through continuous (sensors) or periodic (sampling) monitoring. The dataset consists of continuous time series variables related to (i) meteorology, including precipitation, air temperature, relative humidity, wind speed and direction, net radiation, atmospheric pressure, cloud water flux, irradiance, leaf wetness and soil temperature, (ii) hydrology, including water level and temperature, discharge and electrical conductivity of stream, (ii) hydrogeology, including groundwater level, water temperature and electrical conductivity in two piezometers and one groundwater gallery completed by soil moisture measurements under the canopy. The database is completed by periodic time series variables related to (iv) hydrogeochemistry, including field parameters and water analysis results. The periodic sampling survey provides chemical and isotopic compositions of rainfall, groundwater, and stream water at different locations of this watershed.

The ERORUN-STAFOR monitoring database extends from November 2014 to April 2022, with varying start and end dates according to the different variables. The acquisition frequency is from 10 min to hourly for the sensor variables and from weekly to monthly frequency for the sampling. Despite the frequent maintenance of the monitoring sites, several data gaps exist due to the remote location of some sites and instrument destruction by extreme events such as cyclones. This observatory is a unique research site in an insular volcanic tropical environment offering three windows of observation for the study of critical zone processes through upstream-midstream-downstream measurements sites. This high-resolution database is valuable to assess the response of volcanic tropical watersheds and aquifers at both event and long-term scales (i.e. global change). It will also allow various progress in understanding the significant role of the TMCF in the recharge processes, the hydrogeological conceptual model of volcanic islands, the watershed hydro sedimentary responses to extreme climatic events and their respective evolution under changing climatic conditions.

Associated website

- ERORUN OZCAR website (https://www.ozcar-ri.org/fr/observatoire-erorun/)
- STAFOR presentation (https://www.osureunion.fr/les-stations-dobservation/stafor/presentation/)
- OZCAR IR (https://www.ozcar-ri.org/)
- Theia Data and Services center for continental surfaces (https://www.theia-land.fr/en/homepage-en/)
- GEOSUR, OSU-Reunion’s Geonetwork website (https://geosur.osureunion.fr)

Associated funding

- CNRS-INSU (Institut National des Sciences de l’Univers)
- University of La Réunion
- Météo-France
- IRD (Institut de Recherche pour le Développement)
- LGSR (Laboratoire GéoSciences Réunion)
- OSU-Reunion (Observatoire des Sciences de l’Univers de La Réunion)
- French network of critical zone observatories (OZCAR)
- European Regional Development Fund (ERDF)
- La Réunion Regional council Commission with LABELS-IR-INFRA program (GURDTI/20201872-0025807)
- OMNCG Research federation

Associated partners

- La Réunion water office (https://eaureunion.fr/)
- CVH-DEAL(https://www.vigicrues-reunion.re/)
- Météo-France (https://publitheque.meteo.fr/okapi/accueil/okapiWebPubli/index.jsp)
- La Réunion Regional council (https://regionreunion.com/)
- La Réunion National Park (https://www.reunion-parcnational.fr/fr)
- STRATAGEM 974 (http://www.stratagem974.com/)
- Water Resources Research Center, University of Hawai’i at Manoa, Honolulu (https://www.wrrc.hawaii.edu/)

ERORUN-STAFOR Observatory DMP version

ERORUN-STAFOR Observatory DMP version 1.0 - March 11th, 2024 version

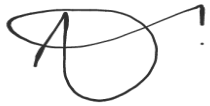
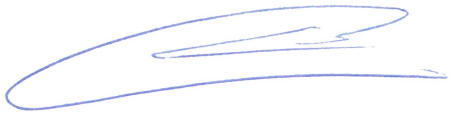
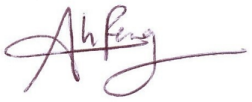
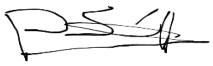
Date	Version #	Status	Author.s	Author's affiliation
11/03/2024	1.0	DMP edition and writing	K/BIDI et al. (see co-author's list)	OSU-Reunion, LGSR, PVBMT, Water resources Research Center (Univ Hawai'i)
Date	**Version #**	**Status**	**Approved by**	**Signatures**
11/03/2024	1.0	DMP approval	K/BIDI Victor, Data steward	
11/03/2024	1.0	DMP approval	MICHON Laurent, Science coordinator	
11/03/2024	1.0	DMP approval	AH-PENG Claudine, Science coordinator	
11/03/2024	1.0	DMP approval	BENARD Bhavani, Science coordinator	
11/03/2024	1.0	DMP approval	MAGAND Olivier, Technical coordinator	
11/03/2024	1.0	DMP approval	STAMENOFF Pierre, Assistant Technical operator	

Table of contents

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 ?

The data description for the ERORUN-STAFOR Observatory presented here is obtained from Toulier *et al.* (2024) (https://doi.org/10.1002/hyp.15061)[1]. These data and the way they are managed, described in the following paragraphs, are collected from November 2014 to April 2022 with varying start and end dates according to the different variables. Data collected later (after April 2022) will be managed by a different DMP following a local OSU-Reunion reorganization of the Observatory and will not be described here. Data treatment (comprising QA-QC issues) as well as storage and backup processes are described in sections 2, 3 and 5. Only final qualified data are freely available and currently downloadable for (inter)national research communities as well as public and commercial purpose uses (CC BY 4.0 license) in open access as on Zenodo platform through the link https://doi.org/10.5281/zenodo.7983138 giving access to datasets used in Toulier *et al*. (2024) related Data paper. Same final qualified data are also stored within Theia/OZCAR in-situ data and service center (https://in-situ.theia-land.fr/), now one of the active members of the Earth System Research Infrastructure Data Terra, also through CC BY 4.0 license. However, ERORUN-STAFOR Observatory data stored in Theia/OZCAR are not yet downloadable ; the dedicated portal user interface being in Beta release at the time of this DMP edition. Finally, and as described further (section 5), ERORUN-STAFOR raw data, additionally to the qualified ones previously noted, are stored on the OSU-Reunion Data Servers giving capacity to preserve data in another repositories ; raw data being reusable for research community only on request to the scientific and technical managers.

The ERORUN-STAFOR Observatory is divided into three main instrumented sites corresponding to different windows of observation of the critical zone, respectively with (A) Plaine des Fougères upstream (noted PF), (B) Rivière des Pluies midstream (RP) and (C) Campus Moufia downstream (CM), located from 50 to 1300 masl (fig. 1).

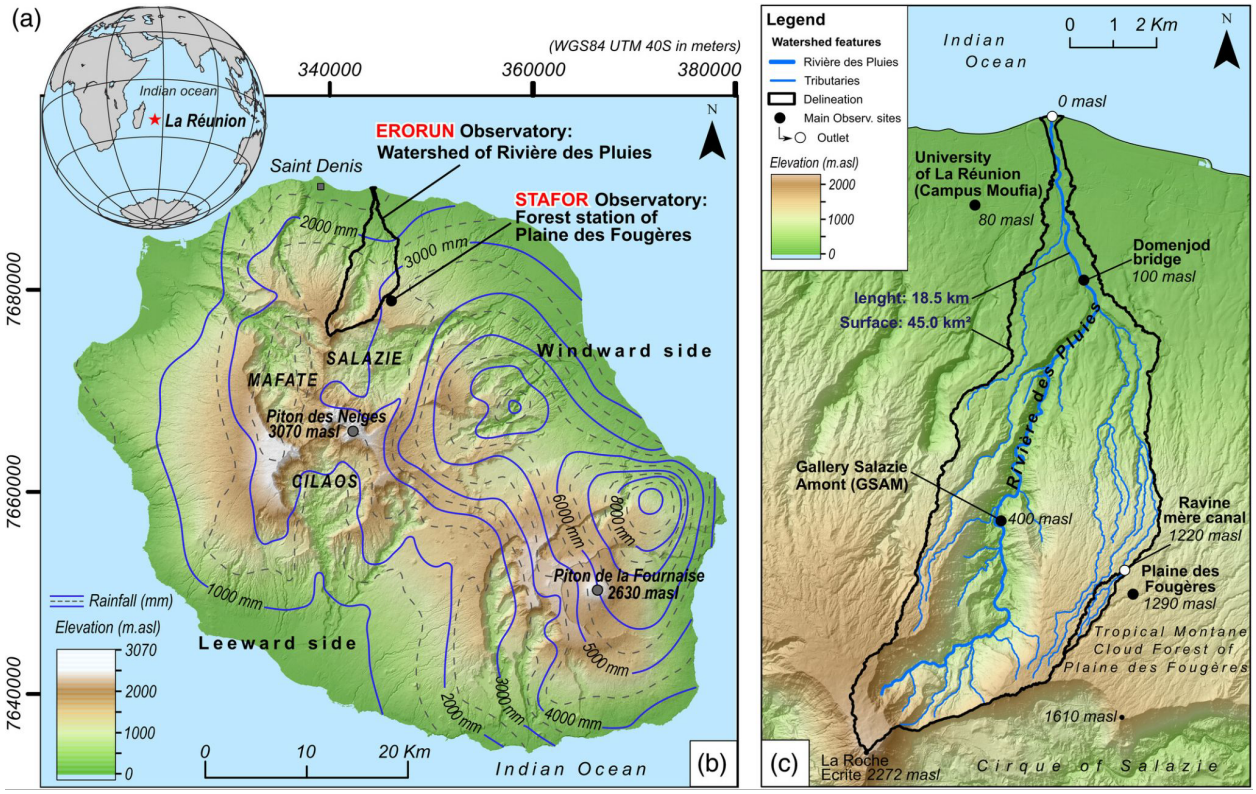


Figure 1. a Location map of La Réunion Island **b** Location and elevation map of La Réunion with isohyets (blue dotted lines and mean annual values) and showing contours of the Rivière des Pluies watershed with the Plaine des Fougères site monitored by the ERORUN-STAFOR Observatory **c.** Elevation map and watershed features of Rivière des Pluies (extracted from Toulier *et al.*, 2024).

[1] Toulier et al.. (2024). ERORUN-STAFOR: A collaborative observatory for the multidisciplinary study of the critical zone processes in a tropical volcanic watershed including a Tropical Montane Cloud Forest. Hydrological Processes, 38(2), e15061. ttps://doi.org/10.1002/hyp.15061

Each site is monitored by several types of stations including (1) climate, (2) forest, (3) river and (4) ground stations. The climate station mainly focuses on measurement of meteorological conditions while the forest station concerns plants diversity and hydric behavior within the microclimate of the forest understory. The river and ground stations focus on surface and soil-groundwater flow dynamics, respectively. The three monitored sites (fig. 2) follow different objectives :

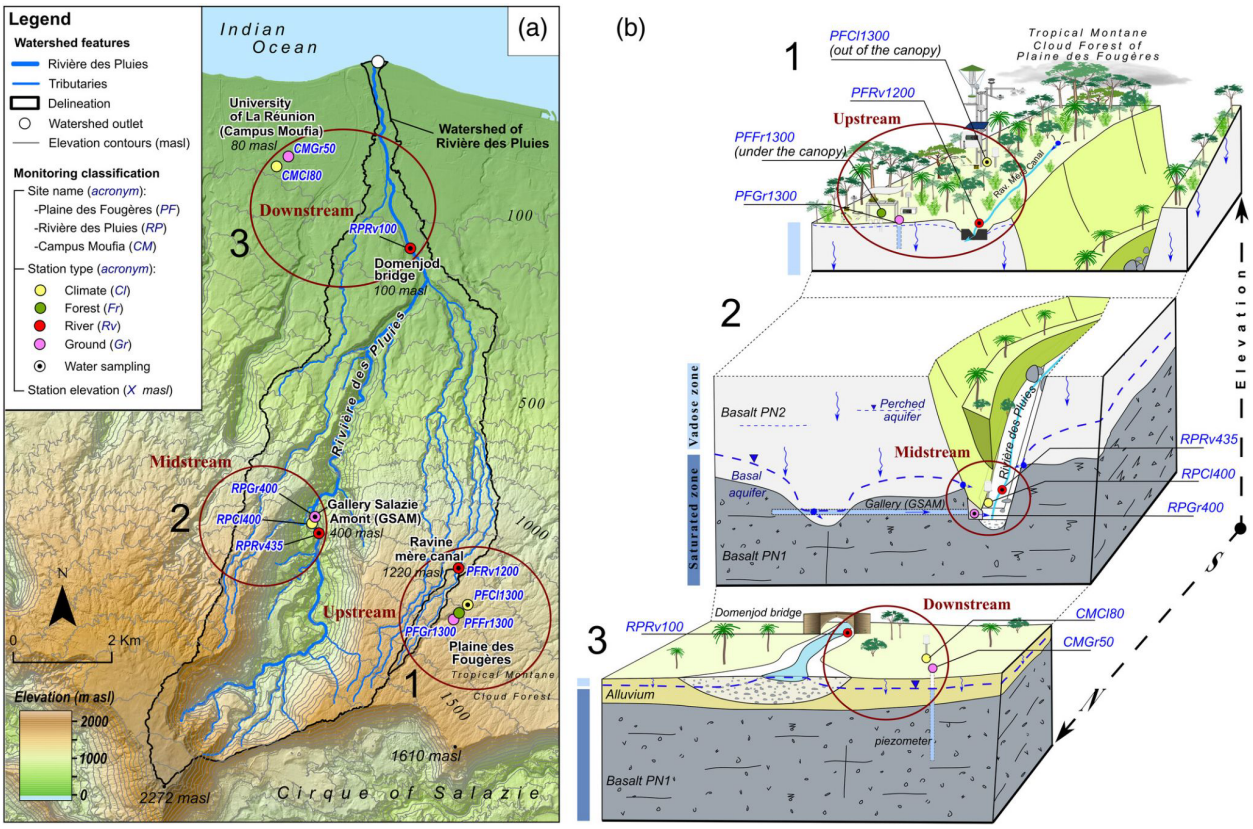


Figure 2 a. Location of instrumented sites and sampling stations, grouped into three sites (upstream, midstream and downstream sites) and four station types (climate, forest, river and ground), **b.** Simplified schema of the instrumented watershed of Rivière des Pluies (modified after Join et al. 2019). Each station name (in blue) is based on the ERORUN-STAFOR monitoring classification (see the Electronic Supplementary Material *ESM1*) corresponding to the acronym of site name (Plaine des Fougères: *PF*, Rivière des Pluies: *RP*, Campus Moufia: *CM*) followed with the acronym of the station type (Climate: *Cl*, Forest: *Fr*, River: *Rv*, Ground: *Gr*) and completed by the station's elevation (e.g. *CMGr50*) (extracted from Toulier *et al.*, 2024).

(A) Upstream site collects data in PF (>1200 masl) to better understand the recharge processes and the role of the TMCF of PF in the hydrological cycle. The PF instrumented site is divided into four complementary stations: (1) a climatic station in a forest clearing (i.e. out of the canopy), (2) a forest understory station (i.e. under the canopy), (3) a river station in Ravine Mère Canal river "co-monitored" with the Office de l'Eau Réunion (OdER) and, (4) a ground station located near the forest station. Here, the term "co-monitored" refers to co-located sites whose instrumentation is shared between local partners and ERORUN-STAFOR Observatory. At such co-monitored sites, equipment maintenance and data collection / processing are independently performed by the instrument's technical manager (e.g. OdER) ;

(B) Midstream site (≈400 masl) offers direct access to the main basal aquifer of the island through the GSAM gallery drilled in basaltic bedrock and also allows assessment of the river behavior in the middle part of the watershed. This site includes three stations: (1) a ground station at the gallery outlet co-monitored with the Département de La Réunion (CG974), (2) a climatic station outside of the gallery and, (3) another river station in the river bed located slightly upstream from the gallery (435 masl).

(C) the downstream site (<100 masl) monitors watershed and aquifer responses with three stations: (1) a river station at the Domenjod bridge co-monitored with the Cellule de Veille Hydrologique de la Direction de l'Environnement, de l'Aménagement et du Logement de La Réunion (CVH-DEAL Réunion), completed by (2) a climatic station and (3) a groundwater station, the latter two of which are located in the Moufia campus (University of La Réunion).

Stations provide high frequency data acquisition using a data logger (e.g. CR1000, Campbell Scientific) with power supplied by a 12V battery and solar panel. In addition, water samples are taken in the three instrumented sites for rainfall, groundwater and surface water analysis (major elements - Cl^-, NO_3^-, SO_4^{2-}, HCO_3^-, Na^+, K^+, Mg^{2+}, Ca^{2+} - and water stable isotopes - $^{18}O/^{16}O$ and D/H -). The sampling frequency ranges from monthly to weekly, depending on the station. Since 2015 each station has been regularly visited, 1-4 times a month in order to (1) download data, (2) clean and check sensors, (3) collect water samples and perform physico-chemical field measurements and (4) perform additional measurements for calibration control (e.g. discharge flow measurements for stream discharge rating curves).

Database described here is acquired with standard equipment (table 1) previously used at other sites of the OZCAR research infrastructure followed by standardized treatment and quality check (see section 2). The data from automatic sensors are downloaded with a laptop computer during each field visit. The data are recorded with a main frequency of 15 min, and more rarely from 10 min to 1 hour (table 1). Detailed information on each of the instruments used and the measured variables is shortly summarized below and in table 1. Measured variables are grouped into four categories : **meteorology, hydrology, hydrogeology, and hydrogeochemistry**.

Meteorology

Since 2015, the common set of measurements at both Climate and Forest stations of PF (PFCl1300; PFFr1300) are:

- Air temperature, relative air humidity (CS215, Campbell Scientific);
- Photosynthetically active radiation (i.e. irradiance; SKP215, Campbell Scientific);
- Leaf wetness rate (237-L, Campbell Scientific);
- Soil temperature (TCAV, Campbell Scientific).

Since 2020, Climate station (PFCl1300) has been improved with addition of new measured variables:

- Net radiation measured with a radiometer (NR01, Hukseflux);
- Wind speed and direction measured with a 2D ultrasonic anemometer (WindSonic 75, GILL instruments);
- Atmospheric pressure was measured with a Baro-Diver barometer (Van Essen instruments), respectively, in the Climate station of PF and at the Moufia campus (PFCl1300 and CMCl80);
- Cloud water flux measurements thanks to a fog collector composed of a Juvik-type fog gauge connected to a 0.1 mm automatic tipping-bucket rain gauge (R13029, Précis Mécanique).

Hydrology

River water level, temperature and electrical conductivity are measured with an automatic CTD sensor (CTD-Diver, Van Essen instruments) installed in PFRv1200, RPRv435 and RPRv100 from 2020. Under low flow conditions, complementary discharge measurements are performed manually with a bucket or other graduated collection container (e.g. at the weir of the PFRv1200 station).

Hydrogeology

Soil volumetric water content (i.e. soil moisture content) is measured with four moisture sensors of two types in the PFGr1300 station:

- Three TDR moisture sensors (Time-Domain Reflectometer) installed from 0 to 55 cm depths since 2019 (CS616, Campbell Scientific)
- One TDR moisture sensor (Soilvue 10, Campbell Scientific) installed near the previous TDR sensors and with nine depths of measurement ranging from 5 to 100 cm depth since 2020.

Groundwater level, temperature and electrical conductivity are measured using two piezometers of 2.20 m depth upstream (PFGr1300; since 2020) and 60.60 m depth downstream (CMGr50; since 2014). Each piezometer is equipped with automatic CTD sensors (CTD-Diver, Van Essen instruments). The piezometric data are provided in masl. Complementary CTD-Diver data are taken in the groundwater gallery GSAM of Rivière des Pluies by the CG974 (RPGr400; since 2019).

Hydrogeochemistry

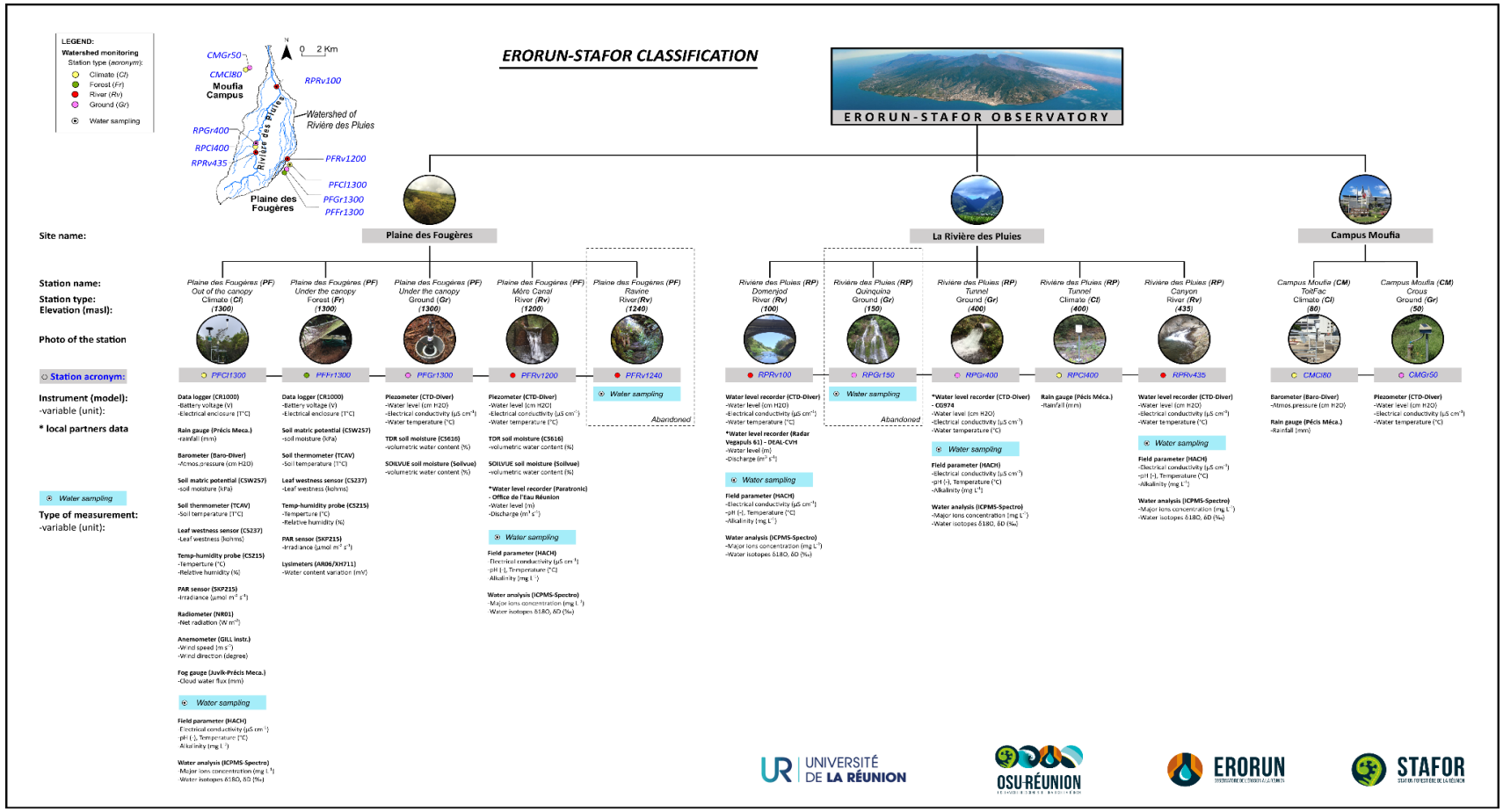
Water samples are manually collected at five stations since 2016, from surface water (PFRv1200, RPRv435, RPRv100), groundwater (RPGr400) and rain water (PFCl1300). Samples are taken monthly, except at RPRv100 which is considered as the watershed outlet and sampled every week (more on hydrogeochemistry on page 9).

Table 1. Summary of the instrumentation and monitored variables for each station of the ERORUN-STAFOR Observatory, with associated precision. The data provided by local partners are precise in brackets and italics. The water analyses are performed in the Laboratory of the Institut de Physique du Globe de Paris (IPGP-PARI laboratory). This table is extracted from Toulier *et al.* (2024).

Monitoring	Dataset	Station name	Instrument - model *(*data from local partner)*	Variables (unit)	Precision	Frequency	Period
Automatic monitoring	Meteorology	*PFCl1300*	Rain gauge - Précis Mécanique R53029	Precipitation (mm)	± 5% (0.5 mm)	15 min	2017-
			Temperature and relative humidity - Campbell CS215	Air temperature (°C)	± 0.4°C (5 to 40°C)		2015-
				Relative humidity (%)	± 4% (0 to 100%)		2015-
			Anemometer - Gill WindSonic 75	Wind speed (m s^{-1})	± 2%		2020-
				Wind direction (degree)	± 2°		2020-
			Radiometer - Hukseflux NR01	Net radiations (W m^{-2})	± 10%		2021-
			Barometer - Van Essen Baro Diver	Atmospheric pressure (cm H_2O)	± 0.5 cm H_2O		2020-
			Fog gauge - Juvik & Rain gauge Précis Mécanique R13029	Cloud water flux (mm)	± 4% (0.1 mm)		2020-
			Photosynthetically active radiation - Campbell SKP215	Irradiance (µmol m^{-2} s^{-1})	Absolute accuracy ±5%		2015-
			Leaf wetness - Campbell 237-L	Leaf wetness (kΩ)	-		2015-
			Soil temperature - Campbell TCAV	Soil temperature (°C)	± 0.1°C over 0 to 40°C		2015-
		PFFr1300	Temperature and relative humidity - Campbell CS215	Air temperature (°C)	± 0.4°C (5 to 40°C)		2015-
				Relative humidity (%)	± 4% (0 to 100%)		2015-
			Photosynthetically active radiation - Campbell SKP215	Irradiance (µmol m^{-2} s^{-1})	Absolute accuracy ±5%		2015-
			Leaf wetness - Campbell 237-L	Resistance (kΩ)	-		2015-
			Soil temperature - Campbell TCAV	Soil temperature (°C)	± 0.1°C over 0 to 40°C		2015-
		RPCl400	Rain gauge - Précis Mécanique R53029	Precipitation (mm)	± 5% (0.5 mm)		2017-
		CMCl80	Rain gauge - Précis Mécanique R53029 (data not shown)	Precipitation (mm)	± 5% (0.5 mm)		2015-
			Barometer - Van Essen Baro Diver	Atmospheric pressure (cm H_2O)	± 0.5 cm H_2O		2020-
	Hydrology	*PFRv1200*	Paratronic 0-10m (**OdER*)	Water level (m)	± 0.005 m	15 min	2006-
				Discharge (m^3 s^{-1})	± 0.001 to ±0.1 m^3 s^{-1}		2006-
			CTD - Van Essen CTD Diver 0-10m	Electrical conductivity (µS cm^{-1})	± 1% (min. 10 µS cm^{-1})		2020-
				Water temperature (°C)	± 0.1°C		2020-
		RPRv100	Radar Végapuls 61 (**CVH-DEAL*)	Water level (m)	± 0.01 m		2007-
				Discharge (m^3 s^{-1})	± 0.1 to ±1 m^3 s^{-1}		2007-
			CTD - Van Essen CTD Diver 0-10m	*Same than previous CTD*	*Same than previous CTD*		2020-
		RPRv435	CTD - Van Essen CTD Diver 0-10m	*Same than previous CTD*	*Same than previous CTD*		2020-
	Hydrogeology	*PFGr1300*	3 x TDR soil moisture - Campbell CS616; 0-31-51cm depth	Volumetric water content (% H_2O)	± 2.5%	1 h	2019-
			TDR soil moisture - Campbell Soilvue 10 (0 -100cm depth)	Volumetric water content (% H_2O)	± 1.5% (with most soils)		2020-
			CTD - Van Essen CTD Diver 0-10m	*Same than previous CTD*	*Same than previous CTD*		2020-
		RPGr400	CTD - Van Essen CTD Diver 0-10m (**CG974*)	*Same than previous CTD*	*Same than previous CTD*	10 min	2019-
		CMGr50	CTD - Van Essen CTD Diver 0-10m	*Same than previous CTD*	*Same than previous CTD*	30 min	2014-
Manual monitoring	Hydrogeochemsitry	*PFCl1300, PFRv1200, RPRv435, RPGr400, RPRv100*	Field multiparameter - HACH HQ40d	Electrical conductivity (µS cm^{-1})	± 0.5% (1 - 200 µS cm^{-1})	Monthly, except at RPRv100 is weekly	2016-
				pH (-)	± 0.1		
				Temperature (°C)	± 0.3°C		
			Manual Titrator – Digitial Titrator HACH	Alkalinity (mg L^{-1})	± 10 mg L^{-1} (estimated)		
			Mass spectrometer Neptune+, ThermoFisher (IPGP, France)	Water isotopes (δ18O, δD ‰)	± 0.03‰, ±0.3‰		
			Ionic chromatography Dionex ICS5000 plus - ThermoFisher and Dionex Integrion HPIC, ThermoFisher (IPGP, France)	Major ions concentration (mg L^{-1})	± 5%		

Next page, figure 3, extracted from Toulier *et al.* (2024)'s supplementary material (figure ESM1), complements the above text as well as table 1.

Figure 3. ERORUN-STAFOR Observatory monitoring classification.



Hydrogeochemistry (continued and end)

Prior to sampling, physico-chemical parameters (electrical conductivity, pH, temperature) are measured in the field using a multiparameter probe (HQ40d, HACH, Germany), regularly calibrated with commercial standard solutions. Then, water sampling is performed at each site according to the following protocol: Anions and cations (major ions) are collected in amber HDEP bottles of 30 ml and stored without air bubbles. Water isotopes are collected in 15 ml amber HDEP bottles and also avoid entrapped air bubbles. Alkalinity is collected in 250 ml HDEP bottles. Cation samples are preserved by acidifying to pH 2 with concentrated trace grade HCl. Each vial is rinsed three times with source water prior to sampling. Anions and cations are filtered using 0.2 µm Millipore cellulose acetate membrane (USA). Samples are kept refrigerated (4°C) prior to analysis. Alkalinity is determined at the GéoSciences Réunion Laboratory in the few hours after sampling by titration with H_2SO_4 (0.16 N) using a manual titrator (Digital Titrator, HACH, Germany). All samples are sent to the High-Resolution Analytical platform (PARI) of the Institut de Physique du Globe de Paris (IPGP) in France, and analyzed as described in Toulier *et al.* (2024) and with analytical instruments noted in table 1 (previous page).

1B. WHAT DATA (FOR EXAMPLE THE KIND, FORMATS, AND VOLUMES), WILL BE COLLECTED OR PRODUCED?

Data collected are the ones belonging to all variables precisely described in subsection 1A and distributed in four categories (meteorology, hydrology, hydrogeology and hydrogeochemistry). The type, format and volume of data collected or produced vary depending on whether we are looking at raw (i.e., data directly emitted or produced by the instruments) or qualified/validated data. In addition, as shown in Figure 4 (see page 11 - figure extracted from Toulier *et al.* (2024), and amended in the present DMP), raw data can be generated automatically (transmitted or not) in the field-sites or produced after *in-situ* manual measurements or laboratory analysis. We then would divide this subsection between data files issued from the raw data production process and the validated (or qualified) crosschecked ones.

Raw data files

A file summarizing the various paths to the raw data is available on the OSU-Reunion cloud server. Below is more detailed information for each station.

- **Meteorology data**

Four stations (PFCI1300, PFFr1300, RPCI400, CMCI80) in this data category systematically generate automatic data files corresponding to precipitation (PFCI1300, RPCI400), air temperature (PFCI1300, PFFr1300), relative air humidity (PFCI1300, PFFr1300), wind speed and direction (PFCI1300), net radiation (PFCI1300), irradiance (PFCI1300, PFFr1300), atmospheric pressure (PFCI1300, CMCI80), cloud water flux (PFCI1300), leaf wetness (PFCI1300, PFFr1300) as well as soil temperature (PFCI1300, PFFr1300). From data issued from all variables automatically collected and recorded by the 4 stations previously identified, around 84% are directly transmitted from the field to appropriate files on OSU-Reunion servers, and only 3 batches of data (precipitation for the PFCI1300 and RPCI400 stations, and atmospheric pressure for the CMCI80 station) are directly retrieved in the field by downloading them from an acquisition unit.

- **PFCI1300**

Raw precipitation data from station PFCI1300 are stored in a directory on the OSU-Reunion cloud server, enabling synchronization between field computers. The instrument produces one single monthly .csv file (< 50 Kb).

For all other parameters (air temperature, relative air humidity, wind speed, wind direction, net radiation, atmospheric pressure, cloud water flux, irradiance, leaf wetness, soil temperature), data are daily uploaded to the OSU-Reunion Zarlor server. Each single .txt file, containing all above-cited variables, is less than 100 Kb.

- **PFFr1300**

All raw data from station PFFr1300 are daily uploaded to OSU-Reunion's Zarlor server. Each single .txt file, containing the collected variables, is less than 1.5 MB in size.

- **RPCI400**

Raw precipitation data from station RPCI400 are stored in a directory on the OSU-Reunion cloud server, enabling synchronization between field computers. The instrument produces one single monthly .csv file (< 50 Kb).

- **CMCI80**

Raw atmospheric pressure data from station CMCI80 are stored in a directory on the OSU-Reunion cloud server in two files in .csv and .mon format, which are incremented on a monthly basis (file size < 5 MB).

- **Hydrology data**

Three stations (PFRv1200, RPRv100, RPRv435) in this data category systematically generate automatic data files corresponding to water level (PFRv1200, RPRv100, RPRv435), discharge[2] (PFRv1200, RPRv100), electrical conductivity (PFRv1200, RPRv100, RPRv435) and water temperature (PFRv1200, RPRv100, RPRv435). All data issued from the 3 stations previously identified, are directly retrieved in the field by downloading them from an acquisition unit.

- **PFRv1200**

All raw data[3] from station PFRv1200 are stored in .csv format (monthly frequency) in a directory on the OSU-Reunion cloud server. Recordings are made every 15 minutes. Each single file, containing above-cited station dedicated variables except discharge ones, is less than 1 MB.

- **RPRv100**

All raw data[4] from station RPRv100 are stored in .csv format (monthly frequency) in a directory on the OSU-Reunion cloud server. Recordings are made every 15 minutes. Each single .txt file, containing above-cited station dedicated variables except discharge ones, is less than 1 MB.

- **RPRv435**

All raw data from station RPRv435 are stored in .csv format (monthly frequency) in a directory on the OSU-Reunion cloud server. Recordings are made every 15 minutes. Each single .txt file, containing all above-cited station dedicated variables, is less than 1 MB.

- **Hydrogeology data**

Three stations (RPGr400, PFGr1300, CMGr50) in this dataset category systematically generate automatic data files corresponding to water level (RPGr400), water temperature (RPGr400, PFGr1300, CMGr50), electrical conductivity (RPGr400, PFGr1300, CMGr50), TDR volumetric water content (PFGr1300), soil volumetric water content (PFGr1300) and piezometry (PFGr1300). From all data issued from the 3 stations previously identified, around 82% are retrieved in the field by downloading them from an acquisition unit, and only 2 batches of data are directly transmitted to OSU-Reunion's Zarlor server (DR volumetric water content and soil volumetric water content collected in the PFGr1300 station).

- **RPGr400**

Raw water level, electrical conductivity and water temperature data are available on the OSU-Reunion cloud server, after download processes from the field, and stored in a .txt format. We have three independent files of 800 Kb each for a total of 2.3 MB.

- **PFGr1300**

Raw TDR and soil volumetric water content data (.txt format) are automatically uploaded daily to OSU-Reunion's Zarlor server. TDR (593 files) and soil volumetric water content (460 files) daily data files are less than 1.5 MB and 100 Kb in size, respectively. The raw data for the other parameters (piezometry, electrical conductivity and water temperature) are available on the OSU-Reunion cloud server and are stored in a .csv file (< 300 Kb) from the instrument, which takes a measurement every 15 minutes.

- **CMGr50**

Raw data from the CMGr50 station for piezometry, electrical conductivity and water temperature are available on the OSU-Reunion cloud server and stored in a .csv file from the instrument, which takes a measurement every 15 minutes. This file is less than 2.5 MB in size.

[2] Discharge data are owned by the Réunion Water Board (Office de l'eau Réunion ; Office de l'eau Réunion (eaureunion.fr)). They are not available for sharing and distribution at this stage of the observatory's development.

[3] Except discharge data (see comment on previous footnote)

[4] Except discharge data (see comment on previous footnote)

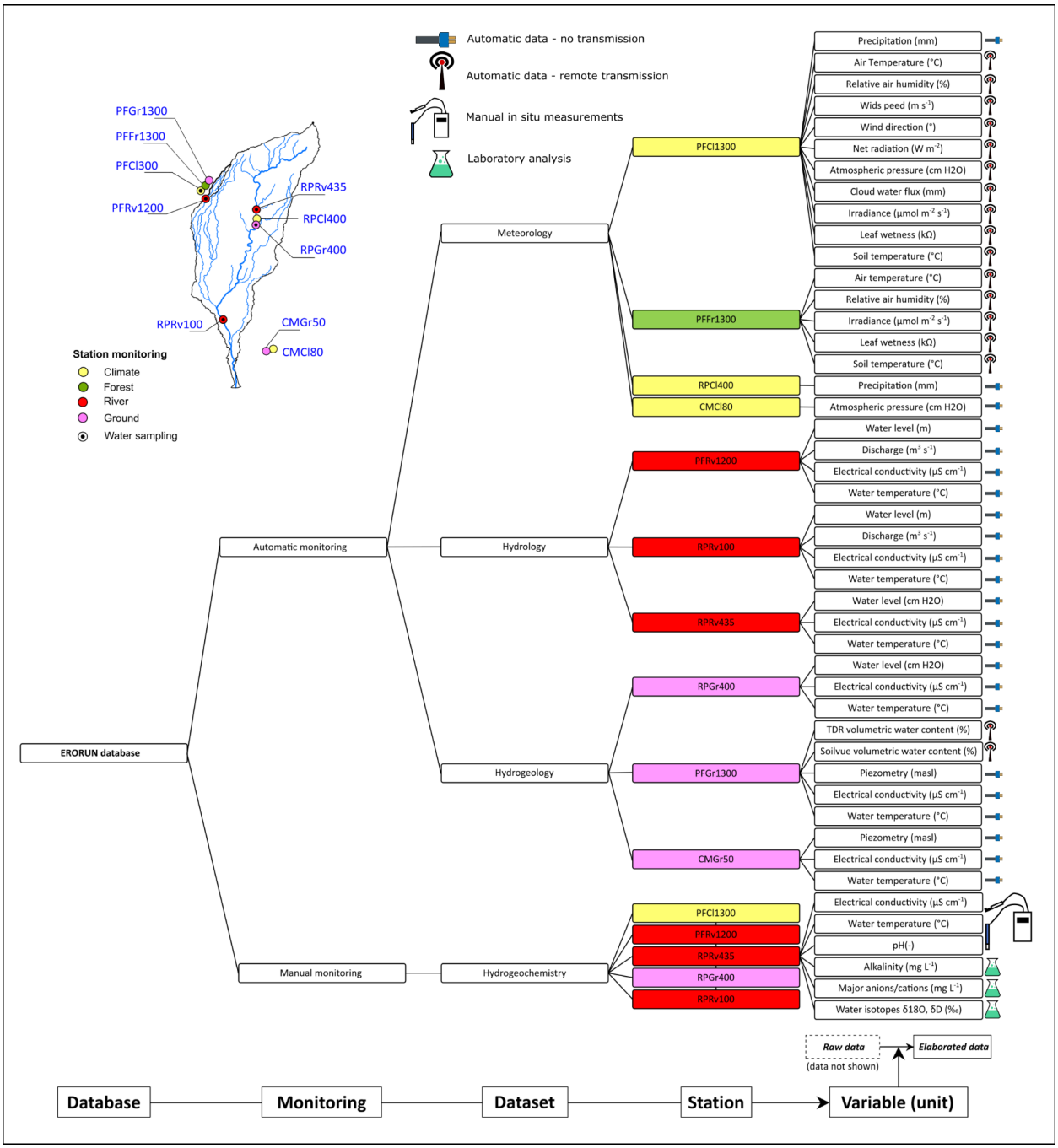


Figure 4. Organization chart of the ERORUN-STAFOR Observatory database (amended from Toulier et al., 2024).

- **Hydrogeochemistry data**

Unlike the data generated in the previous three categories, hydrogeochemistry data are generated by direct *in situ* measurements (electrical conductivity, water temperature and pH) or by analysis following field previous collection and sampling actions (alkalinity, major ions, isotopes). Five stations are covered by hydrogeochemistry data (PFCl1300, PFRv1200, RPRv435, RPGr400, RPRv100). The raw hydrogeochemistry data are stored in a directory shared by ERORUN-STAFOR group on Google drive[5].

[5] Storage on google drive will no longer be used as part of the restructuring of the Observatory and the reorganization (storage) of data sets.

Raw alkalinity data are entered into an Excel file every week after each field trip. We have one single file per year for each of the five stations mentioned above. Each file is less than 200 Kb (total size close to 700 Kb).

Chemistry data (major ions such as Cl^-, NO_3^-, SO_4^{2-}, HCO_3^-, Na^+, K^+, Mg^{2+}, Ca^{2+}) are derived from analysis carried out by IPGP (Institut de Physique du Globe de Paris) via a subcontract within the framework of a partnership around the chemistry theme. Major ion data are compiled in a single Excel format file (< 0.5 Mb) for the five stations mentioned above.

Water isotope ratio measurements ($^{18}O/^{16}O$ and D/H) are also derived from analysis carried out by IPGP subcontracting. Water isotopes data are compiled in a single .xlsm file per station (individual file size < 200 Kb); files which are incremented as and when analysis are carried out, with no defined frequency.

The remaining variables (electrical conductivity, water temperature and pH) are derived from direct measurements on field and/or in laboratory, once back from the field. The corresponding raw data are available in a single .csv file (< 200 kb) containing variables from all the stations mentioned.

Qualified data files

Qualified data files issued from ERORUN-STAFOR Observatory raw data treatment processing are distributed among three different storage infrastructures (OSU-Reunion data server named Zarlor, Zenodo platform and Theia/OZCAR DSC[6]) according to three different structural configurations specifically dedicated to each of these infrastructures. While sections 3 to 5 will precisely describe the three previously cited infrastructures, we will limit ourselves here to list the types, format, number and size of the qualified data files produced after raw (elaborated) data processing.

- **Zenodo platform**

Data on Zenodo are the qualified data on which the Data paper is based. Data are classified in directories by measurement type (hydrogeology, hydrology, meteorology and hydrogeochemistry) and are in .xlsx format (not interoperable format). Each file represents the data for one station with the measurement time step. There are 16 files with a total size of 66 MB.

- **Zarlor OSU-Reunion data server**

Data on Zarlor OSU-Reunion data server is a copy of the data from Zenodo, reclassified in order to have a folder for each dataset and with some slight modifications to the file headers to make them suitable for delivery to Theïa. The files are in the same format (not interoperable format) and have the same size as the Zenodo data.

- **Theia/OZCAR DSC**

For the Theïa delivery, data from Zarlor OSU-Reunion server are processed by a Python program that generates an interoperable .txt file for each variable. These files contain the following information:

- Observation identifier;
- Dataset title;
- Name of the measured variable;
- Measurement start and stop dates;
- Coordinates;
- Quality flag;
- Value.

These files are then zipped by dataset. Finally, the zipped datasets are assembled into a zip file named ERUN_THEIA.zip for THEIA delivery. This final zip is around 15 Mo in size.

In case of missing information, a request for supplementary documents can be made to the ERORUN-STAFOR Observatory coordinators.

[6] On the date of current DMP validation, qualified data are only downloadable from Geosur - linked to OSU-Reunion Zarlor data server - and from Zenodo platform). See sections 3 to 5.

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?

Metadata are defined at several levels.

Firstly, they can be accessed via the Geosur link to the general ERORUN-STAFOR dataset sheet, where we can find information such as:

- Identifier, title and description of the dataset;
- Contacts for the resource;
- Status (archived, creation or update required, being created, finalized, continuous update, obsolete, planned);
- Technical information such as update frequency, representation type, scale and coordinate system;
- Contacts for the metadata information;
- Spatial extent;
- Temporal scope;
- Data download link.

This metadata follows the INSPIRE (*INfrastructure for SPatial InfoRmation in the European community*) directive, which aims to promote the exchange of data within the European Community in the field of the environment in the broadest sense.The Geosur sheet for the ERORUN-STAFOR Observatory can be accessed through this link : https://geosur.osureunion.fr/geonetwork/srv/fre/catalog.search#/metadata/90be6d47-8c40-47fe-9b9a-df20c95a6a96

Secondly, we also have metadata defined in the ERORUN-STAFOR observatory dataset accessible on Zenodo. For each dataset, we have a metadata file in .txt format defining the data history such as the gaps in the variables time series and providing the processing and correction method if needed.

Finally, we have two specific metadata for Theïa delivery, in two Excel files: one concerning the producer and the other linked to the datasets. They are reviewed by the data steward and the technical operators, as well as validated by the PIs. Producer and datasets metadata for Theia delivery are as follows:

- **Producer metadata**

This file describes the metadata linked to the data producer, including the following information:

- ID, name, title and description of the producer;
- Contacts;
- Fundings;
- Links to the resource.

- **Datasets metadata**

This file describes all the metadata for each dataset, including the following information:

- Name, title, description and objective;
- Coordinates (X, Y, Z, LAT, LON) and spatial extent;
- Dataset lineage: describes the genealogy of a dataset, the history of the dataset and if known, its life cycle, from acquisition and capture to its compilation with other datasets and the variants of its current form;
- Contacts;
- Topic categories: Thematic category of the dataset from the ISO standard 19115/Inspire;
- Inspire theme: Thematic dataset required by INSPIRE;
- Keywords;
- Online resource: The various online resources related to the dataset (url, doi, web services);
- Data constraints (license used): defines the constraints on access to and use of spatial data sets and the conditions applicable to restrictions on public access;
- Variables with their measured units;
- Theïa categories: URI identifying a category of variable term in the Theia/OZCAR thesaurus and associates the producer variable name with the Theia/OZCAR taxonomy variable category;
- Sensors used: including model, manufacturer, serial number, sensor type and running activity periods;
- Lineage information: to document data post-acquisition processing;
- Result of missing values and associated quality flags;
- Data type: observation result format. This field will enable the application to perform operations on the observation, such as visualization. This field must contain one of the following notations, i.e., "Numeric", "Text", "Vector", "Raster", "Photo", "Video", "Audio", "Other";
- Information on whether the data set in question is part of a time series;

- Processing level (observation processing level): The object is one of the following notations, i.e., "Raw data", "Quality-controlled data" and/or "Derived products".

2B. WHAT DATA QUALITY CONTROL MEASURES WILL BE USED?

As a prelude, this is important to remind that ERORUN-STAFOR database is divided into two monitoring types: (1) automatic monitoring from sensors including meteorology, hydrology and hydrogeology datasets and (2) manual monitoring including the hydrogeochemistry dataset. Each dataset is divided into two domains, allowing to keep track of all the data processing steps: raw data domain (directly from sensors; data will be available on the OZCAR-Theia database) and elaborated data domain (processed data: sorted data, removed outliers, converted unit, corrected drift, compensated, time-step homogenization and gap filling). The elaborated data correspond to data that have been verified and corrected by data collectors and managers and are provided as the deliverable database. They can be noted as final validated or qualified data.

The following paragraphs provide a brief overview of the quality control procedures applied to the various datasets. More detailed information is available in Toulier *et al.* (2024):

- **Meteorology data**

All data measured by Campbell instruments are directly downloaded in the field from the data logger (CR1000). Data acquired with Diver sensors are processed with Diver Office 2021 software (Eijkelkamp), and corrected for atmospheric pressure fluctuations with a Baro-Diver and from instrumental drift or rating curve shift with manual measurements (e.g. water level measurement at wells and rivers with a manually operated contact gauge KL 010, OTT, Germany). Data from rain gauges are downloaded as a “tipping time value” (for instance 0.5 mm for each tipping with Précis Mécanique R53029). Such data are computed to obtain continuous 15 min cumulative precipitation to align with the other measured variables that are at the 15 min time-step (data processing script of the cumulative precipitation is provided in metadata). Precipitation data from the Climate station of Plaine des Fougères (PFCl1300) are also cross-checked using data from the weather station of Météo-France (Plaine_des_Fougères_SAPC-1064 masl, noted “MF1064” in this paper) located 1.8 km away. As for discharge data, precipitation data from Météo-France are not provided.

In terms of data correction, only gap filling is processed for precipitation gaps of the station RPCl400, using a linear correlation with the station PFCl1300 located 4 km away. For this station, the elaborated data both with and without gap filling correction is provided. The gaps and their corrections - whenever this was possible - are described in metadata (text files) associated with each dataset. Other information such as calibration parameters of sensors are also described in metadata.

- **Hydrology and hydrogeology data**

Hydrology data are corrected following procedures described in Toulier et al. (2024). Whenever possible, complementary measurements of the piezometric level are performed during each field visit (typically monthly) with a manually operated contact gauge (KL 010, OTT, Germany) for the sole purpose of checking any instrumental drift or failure.

- **Hydrogeochemistry data**

For the hydrogeochemistry dataset, field measurements and chemical analysis are performed using certified samples for the calibration of all sensors and analyzers (see standard operating procedures described in Toulier *et al.*, 2024). A quality criterion is provided for the major ions results : only results for samples with an Ionic Balance between -7 % and 7 % are preserved in the delivered dataset.

3. STORAGE AND BACKUP DURING THE RESEARCH PROCESS

3A. HOW WILL DATA AND METADATA BE STORED AND BACKED UP DURING THE RESEARCH?

Below, we describe how raw data, qualified data and metadata are recorded and stored during research:

- **Raw data**

Transmitted raw data are stored on OSU-Reunion's Zarlor data server (see **5. DATA SHARING AND LONG-TERM PRESERVATION**) in order to be backed up with a 45-day retention period thanks to the IT service of University of La Réunion.

Raw data without transmission are saved on the OSU-Reunion Cloud server, enabling automatic synchronization of field computers. This storage facility enables the same maintenance of University of La Réunion mentioned above.

Finally, hydrogeochemistry analysis results are obtained in the laboratory or from our partners and are stored on a Google drive shared by the team[7].

- **Qualified data**

Validated data are stored on the Zarlor archiving data server and have the same guarantees as defined above. The data paper data archive is on the Zenodo platform and can be accessed through the link: https://doi.org/10.5281/zenodo.7983138

- **Metadata**

The metadata are entered into an Excel file for submission to Theia and back up on the OSU-Reunion Cloud server for preservation. They are also available on the OSU-Reunion GEOSUR platform.

3B. HOW WILL DATA SECURITY AND PROTECTION OF SENSITIVE DATA BE TAKEN CARE DURING THE RESEARCH

At the stage of the updated DMP version, no sensitive data is collected in the framework of ERORUN-STAFOR Observatory structure works. The downloadable qualified data are only scientific measurements related to critical zone and forest studies. In case of data loss (IT storage troubleshootings), the intent is to make use of the redundant copies of the data at the various cited locations and/or the sustainable archiving data provided by dedicated services thanks to the duplication processes. Finally, only contact information (first names, last names, ORCID identifiers and e-mail addresses of contributors) may be considered as personal sensitive data and are treated as such by dedicated IT services.

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?

ERORUN-STAFOR Observatory global policy is compliant with GDPR ("*General Data Protection Regulation"*) one and action plan that is being implemented within OSU-Reunion and then, consequently, meets the requirements for properly handling personal data as defined in the EU dedicated law. To remind, the main objectives of the GDPR are to increase both the protection of people affected by processing of their personal data and the accountability of those involved in this processing. Consequently, all created datasets do not contain personal data.

Downloading ERORUN-STAFOR Observatory final qualified data on Zenodo platform through the link https://doi.org/10.5281/zenodo.7983138, giving access to dataset used in Toulier *et al*. (2024) and that do not contain personal data, do not require also the use of personal data from potential re-users who connect to the above cited website. The same validated data, which are also stored, but currently not yet downloadable, on dedicated Theia data portal (https://www.theia-land.fr/en/homepage-en/), do not contain personal data. Raw and elaborated data stored on the third previously cited repository (OSU-Reunion Data Server named Zarlor) do not store personal data. Data concerned by the current DMP are derived from scientific measurements and other related quantities that do not require personal data storage. At present, the current above cited Data centers and servers where potential data re-users can access the processed data can do so without using or archiving any personal data.

[7] Storage on google drive will no longer be used as part of the restructuring of the Observatory and the reorganization (storage) of data sets.

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

All data produced by the ERORUN-STAFOR Observatory are funded by French public funding. As such, in agreement with national policy including CNRS Open Science National Plan, the data have to be freely available and downloadable. Data are under Creative "Commons Attribution" License 4.0 (CC BY 4.0).This license enables reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. Reusers should also inform the PIs, science coordinators and/or the data steward of their published work featuring the data. Reusers should cite the data with the following sentence:

"*ERORUN-STAFOR observatory dataset has been collected thanks to the financial support of Institut National des Sciences de l'Univers (INSU-CNRS), Université de La Réunion (UR), Météo-France, Institut de Recherche pour le Développement (IRD), Laboratoire Géosciences Réunion, Observatoire des Sciences de l'Univers de La Réunion (OSU-Reunion), French network of critical zone observatories (OZCAR) and also from European Regional Development Fund (ERDF, La Réunion Regional Council, European Commission, French State) under program LABELS-IR-INFRA (GURDTI/20201872-0025807). Qualified datasets can be accessed from zenodo repository (https://doi.org/10.5281/zenodo.7983138) as well as from OZCAR-Theia repository (https://catalogue.theia-land.fr/)".*

It is important to note that, ERORUN-STAFOR Observatory data stored in Theia/OZCAR are not yet currently downloadable; the dedicated portal user interface being in Beta testing version at the time of this DMP edition. However, data download capacity from Theia/OZCAR will be available once the beta version has been consolidated. Consequently, in the long term, once the internal structure of the OSU-R would have been changed and a new observatory would have been created following on all works initiated by the ERORUN-STAFOR Observatory, the storage and downloading processes of newly generated data will be limited to 2 Data Centers and Servers, which will be explained in a new dedicated DMP.

Data produced by partners and collaborators of ERORUN-STAFOR Observatory are only shared within the members of the Observatory and their availability to the public is managed by each organization according to their own policy.

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

Data produced by the ERORUN-STAFOR observatory do not raise any ethical issues.

The data providers (ERORUN-STAFOR observatory science and technical coordinators) are well aware of their responsibility for the correctness of data and metadata. In general, everyone should work in a socially ethical way keeping the integrity and FAIRness, and maintaining a high level of trust and respect among the people working in the dedicated observatory and with the users and other stakeholders. One should always consider that the mission of all above-mentioned research studies is to provide effective access for a wide user community to its resources and services, in order to facilitate high-quality dedicated critical zone research, to increase the excellence in Earth system research, and to provide information and knowledge on developing sustainable solutions to societal challenges.

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?

Data from the ERORUN-STAFOR observatory, collected from 2014 to 2022, have been published in a journal article: *ERORUN-STAFOR: A collaborative observatory for the multidisciplinary study of the critical zone processes in a tropical volcanic watershed including a Tropical Montane Cloud Forest* (https://doi.org/10.1002/hyp.15061). Additionally, they will be made available on Theia, the dedicated public platform for the critical zone observatories. Data collected later on will be managed by a different DMP.

Data produced by partners and collaborators of ERORUN-STAFOR observatory will not be displayed.

ERORUN-STAFOR observatory data are not subject to embargo periods.

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

All ERORUN-STAFOR Observatory data collected and qualified are considered as data with long-term value and consequently, to be long-time archiving, considering that they cannot be recreated or produced. As previously noted, ERORUN-STAFOR qualified data (as well as raw data) are currently stored in three different repositories, as explained as follows:

As mentioned above, final validated data cited in the data paper are archived on the Zenodo platform which has assigned them an identifier (https://doi.org/10.5281/zenodo.7983138). Via Zenodo infrastructure, data are safely stored in the CERN Data Center corresponding to the same cloud infrastructure as research data from CERN's Large Hadron Collider; infrastructure that uses CERN's battle-tested repository Invenio software and dedicated digital library (open source one). Data files and related metadata are kept in multiple online and offline copies, guaranteeing backup processes and avoiding loss capacities. More information about Zenodo infrastructure is available at http://about.zenodo.org/infrastructure.

The final validated dataset, previously cited and stored in the Zenodo repository, is also stored on the OSU-Reunion data server named Zarlor. Zarlor is OSU-Reunion's main storage facility. It was set up to centralize and store all data from the four Observatories and stations managed by the institution, as well as from the research programs carried out by the various affiliated laboratories belonging to the OSU-R Research Federation. The Zarlor server is hosted in a highly available storage area (SAN) managed by the University of La Réunion. It currently operates 4 TB on the datastore, but is set to grow rapidly in the coming years. Zarlor data is shared with the various servers running the data via the NFS protocol. This enables access in read or read-write mode. Going through the university's system means we can have a backup with a 45-day retention period. To further protect the data, OSU-Reunion makes a backup of this data on another university site. Additionally to final validated data, raw data are also stored in the Zarlor data server enabling data to be cataloged from acquisition to final validation and qualification.

Lastly, and as previously pointed, final validated ERORUN-STAFOR Observatory datasets are also stored on dedicated Theia/OZCAR critical zone in situ measurements service and data infrastructure (see https://www.theia-land.fr/en/homepage-en/). The current Theia/OZCR Information System under development is based on data exchange standards (INSPIRE, OGC) and is committed to implementing the FAIR principles. Being part of Data Terra research infrastructure, Theia/Ozcar is the data and services hub specialized in land surfaces. Data Terra's IT infrastructure is built around eight Tier2 HPC data and processing centers combining computing and storage capacity dedicated to hosting and exploiting large volumes of data. The infrastructure is built in such a way as to guarantee the long-term archiving of validated data deposited in dedicated DSC, via the transversal services it includes.

Consequently, ERORUN-STAFOR Observatory final validated data are stored and backed up both in Zenodo and Theia/OZCAR repositories, as well as in Zarlor's /data/observatory/erorun directory and are preserved long-term, i.e., as long as the repositories exist. The same for ERORUN-STAFOR Observatory raw data, but only in Zarlor's data server. We have to keep in mind that, at this stage, data collection long-term preservation is not always guaranteed whatever the data center. The expectation is a data preservation process for at least one decade.

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

As previously noted and from the current validation date of this DMP, ERORUN-STAFOR data are stored (i) on Zenodo platform, (ii) on dedicated Theia data portal as well as on (iii) OSU-Reunion's Data Server called Zarlor (see section 4A). However, ERORUN-STAFOR data concerned by the present DMP are only accessible on Zenodo platform and on OSU-Reunion's Zarlor data server; the dedicated Theia data portal data do not yet permit download processes (portal in beta testing at the production date of this document).

On the Zenodo platform - the CERN Data Centre-backed research data repository for the long-tail of science,an innovative and easy to apply web-platform using Invenio free open source software -, ERORUN-STAFOR data can be easily and freely downloaded through the link https://doi.org/10.5281/zenodo.7983138, without access permission. Validated data are also easily accessible via OSU-Reunion's Zarlor data server. This last service, maintained by the OSU-R IT team, provides a stable and open HTTP access link (without identification) to freely download the data. Metadata can also be accessed via OSU-Reunion's GEOSUR Catalog data service via the CSW protocol. This service is based on the geographic data sharing catalog GEONETWORK which is an open source application. Download information (methodology, software used) enabling data and metadata to be retrieved via the THEIA data portal (https://in-situ.theia-land.fr/) will be provided in a future DMP, once the THEIA data platform would be ready for, and when the new Observatory structure for collecting, processing and making available data from the critical zone studied in Réunion will be implemented.

Validated (qualified) data downloadable on Zenodo platform (.xlsx format) and on OSU-Reunion's Zarlor Data Server (.txt interoperable format), can be easily reused by various processing programs.

5D. HOW WILL THE APPLICATION OF A UNIQUE AND PERSISTENT IDENTIFIER (SUCH AS A DIGITAL OBJECT IDENTIFIER (DOI)) TO EACH DATA SET BE ENSURED?

Data acquired from 2014 to 2022 have been published as a dataset accompanying Toulier et al. (2024)'s paper (https://doi.org/10.1002/hyp.15061). The dataset was uploaded on the open data repository ZENODO and can be accessed through the link https://doi.org/10.5281/zenodo.7983138. ZENODO is hosted by CERN (European Organization for Nuclear Research) and follows FAIR principles. Ultimately, OSU-Reunion would have the agreement to directly assign DOIs to further coming datasets.

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 data steward manager and scientific observatory coordinators, clearly identified to the partners (see associated partners list), are in charge of coordinating data management for the whole project from data capture to archiving and sharing steps.

Each partner has to define a "data correspondent" responsible for the management of the data produced in his team for the identified sites. The data correspondent has to transmit all the necessary information to the ERORUN-STAFOR data steward and scientific observatory coordinators to establish and update the current data management plan.

The ERORUN-STAFOR data steward and scientific as well as technical observatory coordinators update and submit the corresponding DMP when necessary (versioning step) as well as send it to the data correspondent each time a new version is released.

The ERORUN-STAFOR data steward and scientific as well as technical observatory coordinators are responsible for:

- Verifying the data quality: structure, format and naming according to the DMP recommendations themselves taking into account standardization requests expressed by Data Centers and Servers to which the data will be uploaded;
- Verifying that metadata are correctly filled in;
- Transferring the data to the Theia/OZCAR *in-situ* data and service center.

The ERORUN-STAFOR Observatory data steward and correspondents may change during the project and any changes will be recorded in the data management plan.

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)?

ERORUN-STAFOR Observatory group of the OSU-R ensures that ERORUN-STAFOR data are FAIR.

Data management is under the responsibility of ERORUN-STAFOR scientific, technical Observatory and data steward managers, who meet the objectives and FAIR principles recommended for the management and sharing of monitoring data within Theia/Ozcar in-situ data and service center (*https://in-situ.theia-land.fr/*), which is one of the active data center member of Earth System Research Infrastructure (Data Terra - https://www.data-terra.org/).

Financial effort is provided through recurring annual funding as well as recruitment of technical staff, trainees / PhD / post-doctoral students, who will be in charge or helpful in the data management processes. This will be completed by time used by the science observatory managers to coordinate the whole data management.