COST Action CA20136
Opportunistic precipitation sensing network - OpenSense

D1.5 Benchmark dataset and documentation in the form of a report

Version 1.0

Working Group	WG1
Deliverable	D1.5
Date	28/11/2023
Version	1.0
Authors	Martin Fencl (CTU-Czech Rep.) and Roberto Nebuloni (CNR-Italy)
Reviewer	Vojtěch Bareš (CTU-Czech Rep.)

Description	This document describes the benchmark datasets that will be used by the OpenSense Community to test the algorithms for processing opportunistic sensing data and assess data applications.
Keywords	Repository, datasets, OpenSense, CML, SML, PWS

About OpenSense (COST Action CA20136). *OpenSense brings together scientists investigating different opportunistic sensors (e.g. microwave links, citizen science), experts from weather services, and end-users of rainfall products to build a worldwide reference opportunistic sensing community. The overarching goals of the COST are to overcome key barriers preventing data exchange and acceptance as hydrometeorological observations, define standards to allow for large-scale benchmarking of opportunistic sensing precipitation products and develop new methods for precipitation retrieval, coordinate integration of the opportunistic observations into traditional monitoring networks, and identify potential new sources of precipitation observations. Further details can be found here:*

Table of contents

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

Funded by
the European Union

www.cost.eu
www.opensenseaction.eu

COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation.

Glossary

CML	Commercial Microwave Links
csv	Comma-separated values (filename extension associated with text files)
DM.N	Deliverable N of working group M
GPK	Grant period K
MoU	Memorandum of Understanding
NDA	Non-Disclosure Agreement
nc	Filename extension of the NetCDF (Network Common Data Form) data format
OpenSense	Opportunistic precipitation sensing network
OS	Opportunistic Sensors
PWS	Personal Weather Stations
SML	Satellite Microwave Links
WG	Working Group

1. Introduction

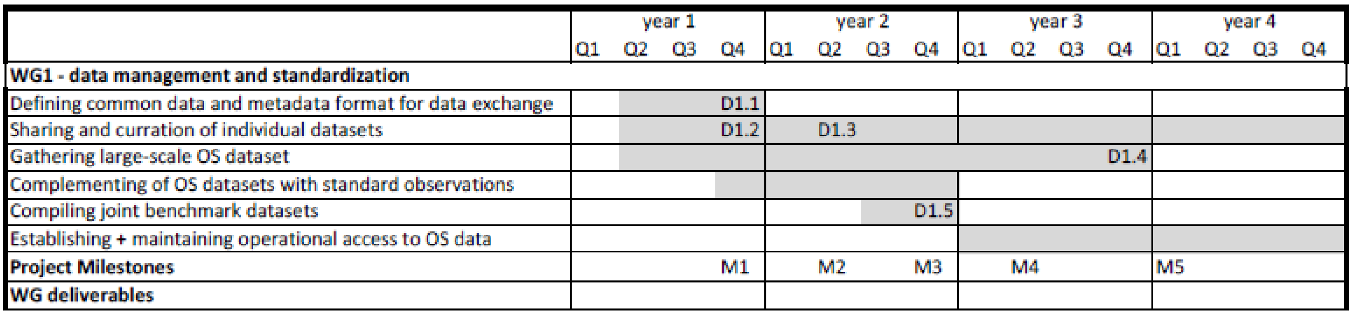
This document reports an official OpenSense deliverable D1.5 *Benchmark dataset representing different climatic regions and documentation in the form of a report*. It is an output of Activity 2 *Compiling of (benchmark) datasets*. Details about WG1 activities, milestones and deliverables due reported in the OpenSense MoU [1] are listed in Table 1.

D1.5 builds on the activities of the internal deliverable D1.1 that was delivered at the end of GP1, where sample datasets were collected. An important step towards D1.5 was also the database with metadata about datasets of individual OpenSense members, which was created jointly with WG4 in GP1. The final form of the benchmarking datasets is based on the needs of WG2, which is responsible for the actual benchmarking of OS methods.

WG1 and WG2 have jointly agreed on the approach where a collection of benchmarking datasets is created and gradually extended by new datasets. This collection now contains the OpenMRG dataset from Sweden with commercial microwave link (CML) data [2], the dataset with personal weather stations (PWS) from Amsterdam and for a period of one month the whole Netherlands [3]. A dataset with observations of satellite microwave links (SML) was just very recently obtained from HD Rain company with permission to make it available for OpenSense members and it is currently available in the Action’s internal repository.

The collection of benchmarking datasets consists now of open-access datasets with already fixed data formats stored at different repositories. Interoperability is achieved by wrappers gradually built by WG2 in a sandbox environment (D2.3). We envision that interoperability of new benchmarking datasets progressively added in future to the collection, including SML dataset currently available only to OpenSense members, will be achieved by following OpenSense data format conventions [4].

Table 1 Timetable of WG1 activities, milestones and deliverables according to OpenSense MoU.



	year 1				year 2				year 3				year 4			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
WG1 - data management and standardization																
Defining common data and metadata format for data exchange		▨	▨	D1.1												
Sharing and curration of individual datasets		▨	▨	D1.2	▨	D1.3	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
Gathering large-scale OS dataset		▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	D1.4				
Complementing of OS datasets with standard observations				▨	▨	▨	▨	▨								
Compiling joint benchmark datasets							▨	D1.5								
Establishing + maintaining operational access to OS data									▨	▨	▨	▨	▨	▨	▨	▨
Project Milestones				M1		M2		M3		M4			M5			
WG deliverables																

WG1 milestones
M2 large-scale dataset in standardized format available for benchmarking of algorithms
M5 - transboundary OS precipitation product available
M6 - operational access to subset of OS data established

WG1 deliverables
D1.1 Repository for individual OS datasets
D1.2 White paper on data standards/formats for investigated types of OS sensors
D1.3 Documentation of past datasets shared in the repository
D1.4 Large-scale OS dataset completed by standard observations + report
D1.5 Benchmark dataset and documentation in the form of a report
D1.6 Documentation for accessing operational OS data

2. Benchmarking datasets

OpenMRG datasets

The OpenMRG dataset [2] consists of attenuation data at 10-s resolution and metadata including true coordinates from 364 full-duplex CMLs in Gothenburg, Sweden (Fig. 1). As a reference, high-resolution data from 11 precipitation gauges and Swedish operational weather-radar composite of the area are provided. The dataset span over a period of three months (June-August, 2015). The dataset time series are near-complete with a minimum of outages, the availability by CMLs reaches 99.99 %, by rain gauges it is 100 %, and by weather-radar composite, it is 99.6 %.

The total rainfall in the study period according to gauge records was approximately 260 mm, with rainfall occurring in 6 % of each 15-minute interval. During the most intense event (which occurred on 28 July 2015), the Torslanda gauge recorded a peak of 1.1 mm min^{-1}.

The data are stored in netCDF format. Although the format does not follow OpenSense data format conventions, as they were designed later, it resembles them closely. In fact, the OpenMRG dataset largely inspired OpenSense data format conventions. The dataset is accessible under a CC BY-SA 4.0 license at https://doi.org/10.5281/zenodo.7107689 [5].

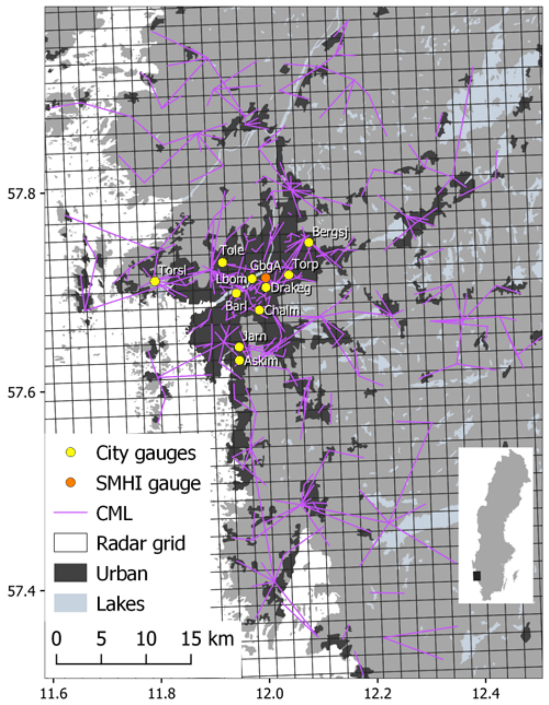


Fig. 1 Map of CMLs and rain gauges in the Gothenburg area [2].

Amsterdam and Netherlands PWS dataset

The Amsterdam PWS dataset [3] consists of 5-min data from NetAtmo PWS [5] located in the Amsterdam metropolitan area (~575 km^2), defined as the area between 4.67–5.05° longitude and 52.24–52.44° latitude (Fig. 2). There are 134 NetAtmo PWS stations [6] available in the area, which corresponds to one PWS per 4.3 km^2 on average. The average distance between PWS is 0.7 km. The observation period is between 1st May 2016 and 1st June 2018. For the period of one month (May 2018), the PWS dataset is available for the whole Netherlands.

The dataset is accessible under a CC BY 4.0 license at https://data.4tu.nl/articles/_/12703250/1.

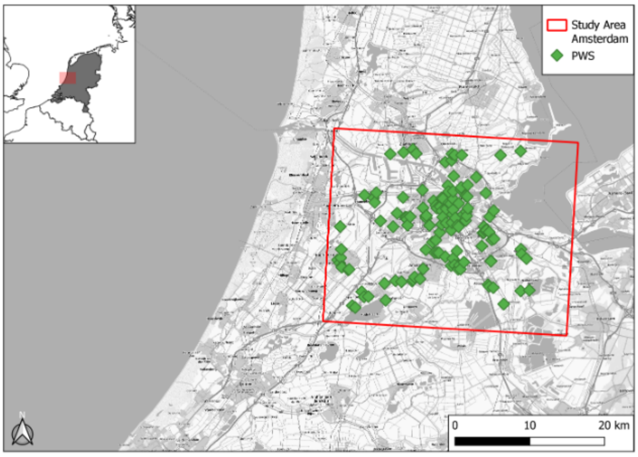


Fig. 2 Map of Amsterdam metropolitan area with all the PWS [7].

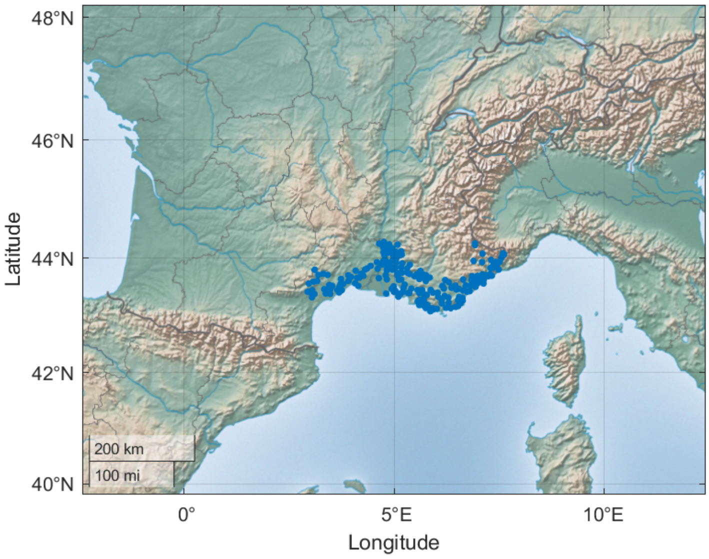


Fig. 3 Location of the 215 HD Rain ground receivers in Southern France

HD Rain SML dataset

The SML dataset has been provided to the OpenSense community by Dr. François Mercier-Tigrine of HD Rain, a French company that commercializes weather products [8]. HD rain dataset includes attenuation data of 215 ground receivers in Southern France and spans the period of five months between August and December 2022. The sensors receive the signals transmitted by the GEOstationary satellites Astra19 and HotBird13E at frequencies ranging in the upper part of the Ku band, i.e. between 11.7 and 12.75 GHz, with horizontal polarization.

Two sets of conventional meteorological data, courteously provided by Meteo France, under a research license agreement valid until the end of the OpenSense action (31 October 2025) are provided as a reference:

- “COMEPHORE” Mosaic of accumulated precipitation during 1 hour in 1 km resolution over France.
- French Mosaic of accumulated precipitation during 5 min in 1 km resolution over France.

The SML dataset is currently stored at the OpenSense internal repository in a csv format. Attenuation data are stored in a unique file per sensor whereas metadata are stored in one csv table for all the sensors. The radar and rain gauge data were provided in Netcdf4 format (.nc file extension). Reference data from Meteo France are not shared on GDrive even though they are available to the OpenSense members upon request to WG1 leaders. Details on the SML dataset including reference data are provided in the WG1 report of deliverable D1.3.

Other Open Access datasets not yet included in the collection

Several other datasets become available during the WG1 activity on compiling benchmarking datasets, however, we do not yet include them in the benchmarking collection since they have not been yet tested during WG2 benchmarking activities.

We list here only datasets covering large areas. For PWS, the WOW dataset [9] covers Great Britain and spans between 2011 and 2020. It contains rainfall, temperature, pressure and humidity data. For CMLs, 15-min min/max attenuation data from the whole Netherlands [10] are available for 21 days in the period 9th June up to and including 11th September 2011 and then all days from 30th May up to and including 1st September 2012.

These data are planned to be included in the benchmarking collection, once tested within WG2 benchmarking activities.

3. Strategy for extending the collection of OS benchmarking datasets

Our ultimate goal is to unlock large OS datasets, make them available in an Open Access mode, and thus enable substantial extension of OS benchmarking dataset collection. Up to now, the publishing of a few Open Access OS datasets was enabled thanks to strong personal relations between individual research groups and OS data owners. OpenSense builds on these relations and supports the process of OS data unlocking at different levels:

- provides an opportunity to publish datasets as a part of OpenSense collection under different access modes
- provides tools and assistance when preparing data to be compatible with OpenSense data conventions
- supports compilation of reference data
- coordinates OS data curation and unlocking with different stakeholders
- supports the process of dataset publishing with COST NETWORKING tools

To support the publishing of OS data, OPENSENSE has initially created an OpenSense community at Zenodo [11]. The OS data maintainers are, however, mostly not allowed to publish datasets under Open Access licenses, on the other hand, most public repositories including Zenodo require that datasets be publicly accessible. Most of the individual OS datasets are thus still stored locally on the hard drives of OpenSense members. WG1 thus negotiated with EUMETNET [12], a network of 31 European National Meteorological & Hydrological Services, a possibility to store individual OS datasets at the CEDA repository [13] with the access requiring authentication. OpenSense strongly recommends reformatting the datasets before publishing following OpenSense data to and metadata conventions [4]. This approach will lead to findable, interoperable and reusable datasets. Accessibility will be fully under the control of data maintainers.

To facilitate the process of data standardization, software tools for converting CML, SML, and PWS data to OpenSense format were developed in collaboration with WG2. In GP3, the Action plans to support the standardization process of individual datasets by COST networking tools such as Virtual Mobility grants and Short-term Scientific Missions. Finally, WG1 provides assistance with the complementation of OS data with standard meteorological observations from National Met Offices.

WG1 is also involved in the organization of a round table with stakeholders about the exploitation of OS data which aims at a common understanding of the needs of potential OS data end-users (e.g. Met Offices) and barriers preventing OS data owners from providing their data. We envision that these efforts will gradually enable us to unlock datasets currently accessible only to individual research groups, which is a prerequisite for extending the collection of OS benchmarking datasets.

References

[1] Memorandum of Understanding for the implementation of the COST Action "Opportunistic precipitation sensing network" (OpenSense) CA20136, downloadable at https://www.cost.eu/actions/CA20136/

[2] Andersson, J. C. M., Olsson, J., van de Beek, R. (C. Z.)., and Hansryd, J.: OpenMRG: Open data from Microwave links, Radar, and Gauges for rainfall quantification in Gothenburg, Sweden, Earth Syst. Sci. Data, 14, 5411–5426, https://doi.org/10.5194/essd-14-5411-2022, 2022.

[3] L. de Vos, H. Leijnse, A. Overeem, and R. Uijlenhoet, "Quality Control for Crowdsourced Personal Weather Stations to Enable Operational Rainfall Monitoring," Geophysical Research Letters, vol. 46, no. 15, pp. 8820–8829, 2019, doi: https://doi.org/10.1029/2019GL083731.

[4] Fencl M, Nebuloni R, C. M. Andersson J et al. Data formats and standards for opportunistic rainfall sensors [version 1; peer review: 1 approved, 1 approved with reservations]. Open Res Europe 2023, 3:169 (https://doi.org/10.12688/openreseurope.16068.1)

[5] J. C. M. Andersson, J. Olsson, C. Z. van de Beek, J. Hansryd, H. Andersson, and J. Persson, "The OpenMRG data set." Zenodo, Sep. 23, 2022. doi: 10.5281/zenodo.7107689.

[6] Netatmo company official site, https://netatmo.com

[7] A. El Hachem et al., "Technical note: Overview and comparison of three quality control algorithms for rainfall data from personal weather stations," Hydrology and Earth System Sciences Discussions, pp. 1–22, Aug. 2023, doi: 10.5194/hess-2023-195.

[8] HD Rain company official site, https://www.hd-rain.com/

[9] "WOW Observations 2011 - 2020." Newcastle University, Dec. 19, 2022. doi: 10.25405/data.ncl.21724970.v1.

[10] Commercial microwave link data for rainfall monitoring https://data.4tu.nl/articles/dataset/Commercial_microwave_link_data_for_rainfall_monitoring/12688253

[11] OpenSense community collection on Zenodo, https://zenodo.org/communities/opensense/about/

[12] EUMETNET official website, https://www.eumetnet.eu/

[13] CEDA repository, https://help.ceda.ac.uk/