empatica

Data Access

PRS-0000205 REV19.0 July 2024

Empatica Health Monitoring Platform

The Empatica Health Monitoring Platform is a modular remote health monitoring platform that encompasses Empatica's product suite for professionals, and consists of Empatica's hardware, software, digital biomarkers, and Cloud. The platform enables the continuous collection, streaming and storage of physiological signals and digital biomarkers. Professionals can use the Empatica Health Monitoring Platform for different purposes, ranging from remote patient monitoring to small and large-scale scientific and clinical research studies.

Document scope and intended audience

In this document, you can discover which types of data are available, how the data is structured, which data formats are used, and how to open them. This document does not provide detailed information about how the data is generated by the Platform or about how Empatica's algorithms function. Please refer to the Empatica Health Monitoring Platform Spec Sheet for more information on data and algorithms.

This document is intended for data scientists and developers, with appropriate access rights to the platform, who want to understand how the platform data is structured in the Empatica Cloud, and to access it for further processing and analysis. A basic understanding of data structure and programming is helpful when reading this document.

Property of Empatica. **CONFIDENTIAL.** Unauthorized distribution or copying prohibited.

Some of the features mentioned in this document are only available with the Enterprise plan. For more information, see the plans table.

Index

DATA FLOW

DATA HIERARCHY

FILE TYPES

DATA FORMAT

DATA ACCESS AND RETRIEVAL

APPENDIX

01

Data Flow

Data flow

The following diagram shows the complete data flow from the EmbracePlus to your infrastructure.

Within 24 hours after data has been successfully uploaded, the data will become accessible from the Empatica Cloud.



The Empatica Health Monitoring Platform provides three types of data:

- **Raw data** is the high-frequency data recorded by the EmbracePlus sensors;
- **Digital Biomarkers** are pre-processed data by Empatica's proprietary and open source algorithms;
- **Reports** are data summaries for a group of participants that provide an overview of different physiological and behavioral metrics.

02

Data hierarchy

Data hierarchy

In the Empatica Cloud, the data is organized by Site, date, Participant, and EmbracePlus serial number. Below you can find the generic hierarchy and an example.

⚠ The availability of the different raw data, biomarkers, and reports depends on the study configuration. For further information on study configurations, please refer to your account manager.

```
# Structure

s3://[s3-bucket-name]
  /[organization_id]
    /[study_id]
      /[site_id]
        /metadata
          /[filename_demographics].csv***
          /[filename_metadata].csv
        /participant_data
          /[date]
            /[participant_id]-[EmbracePlus_sn]
              /digital_biomarkers
                /aggregated_per_minute
                  /[filename_aggregated_per_minute].csv
                /raw_data
                  /[schema_version]
                    /[filename_raw_data].avro
        /reports
          /6mwt*
            /daily
              /[filename_report].csv
          /actigraphy
            /weekly
              /[filename_report].csv
          /autonomic_arousal*
            /weekly
              /[filename_report].csv
          /extended_actigraphy*
            /weekly
              /[filename_report].csv
          /gait_features*
            /weekly
              /[filename_report].csv
          /sleep
            /weekly
              /[filename_report].csv
          /seizure*
            /weekly
              /[filename_report].csv
          /wearing_detection
            /weekly
              /[filename_report].csv
```

Data hierarchy

```
# Example

s3://empatica-us-east-1-prod-data/v2
  /ORGANIZATION
    /STUDY
      /SITE
        /metadata
          /ORGANIZATION-STUDY-SITE_demographics.csv***
          /STUDY-SITE-AB001_metadata.csv
        /participant_data
          /2021-01-14
            /AB001-3ABCDEFGHI
              /digital_biomarkers
                /aggregated_per_minute
                  /STUDY-SITE-AB001_2021-01-14_accelerometers-std.csv
                  /STUDY-SITE-AB001_2021-01-14_actigraphy-counts.csv
                  /STUDY-SITE-AB001_2021-01-14_activity-classification.csv
                  /STUDY-SITE-AB001_2021-01-14_activity-counts.csv
                  /STUDY-SITE-AB001_2021-01-14_activity-intensity.csv
                  /STUDY-SITE-AB001_2021-01-14_body-position.csv
                  /STUDY-SITE-AB001_2021-01-14_double-support.csv*
                  /STUDY-SITE-AB001_2021-01-14_eda.csv
                  /STUDY-SITE-AB001_2021-01-14_gait-speed.csv*
                  /STUDY-SITE-AB001_2021-01-14_met.csv
                  /STUDY-SITE-AB001_2021-01-14_prv.csv
                  /STUDY-SITE-AB001_2021-01-14_pulse-rate.csv
                  /STUDY-SITE-AB001_2021-01-14_respiratory-rate.csv
                  /STUDY-SITE-AB001_2021-01-14_sleep-detection.csv
                  /STUDY-SITE-AB001_2021-01-14_spo2.csv
                  /STUDY-SITE-AB001_2021-01-14_stance-time.csv*
                  /STUDY-SITE-AB001_2021-01-14_stance-time-asymmetry.csv*
                  /STUDY-SITE-AB001_2021-01-14_step-counts.csv**
                  /STUDY-SITE-AB001_2021-01-14_step-length.csv*
                  /STUDY-SITE-AB001_2021-01-14_step-length-asymmetry.csv*
                  /STUDY-SITE-AB001_2021-01-14_step-time.csv*
                  /STUDY-SITE-AB001_2021-01-14_step-time-asymmetry.csv*
                  /STUDY-SITE-AB001_2021-01-14_step-time-std.csv*
                  /STUDY-SITE-AB001_2021-01-14_step-time-variance.csv*
                  /STUDY-SITE-AB001_2021-01-14_stride-length.csv*
                  /STUDY-SITE-AB001_2021-01-14_stride-time.csv*
                  /STUDY-SITE-AB001_2021-01-14_swing-time.csv*
                  /STUDY-SITE-AB001_2021-01-14_swing-time-asymmetry.csv*
                  /STUDY-SITE-AB001_2021-01-14_temperature.csv
                  /STUDY-SITE-AB001_2021-01-14_wearing-detection.csv
              /raw_data
                /v6
                  /STUDY-SITE-AB001_1610609653.avro
                  /STUDY-SITE-AB001_1610610571.avro
        /reports
          /6mwt
            /daily
              /ORGANIZATION-STUDY-SITE_2021-01-14_6mwt-report.csv*
          /actigraphy
            /weekly
              /ORGANIZATION-STUDY-SITE_2021W02_actigraphy-report.csv
          /autonomic_arousal*
            /weekly
              /ORGANIZATION-STUDY-SITE_2021W02_autonomic-arousal-report.csv*
          /extended_actigraphy*
            /weekly
              /ORGANIZATION-STUDY-SITE_2021W02_extended-actigraphy-report.csv*
          /gait_features*
            /weekly
              /ORGANIZATION-STUDY-SITE_2021W02_gait-features-report.csv*
```

Confidential to Empatica, Inc.

```
          /health_metrics
            /weekly
              /ORGANIZATION-STUDY-SITE_2021W02_health-metrics-report.csv
          /seizure*
            /weekly
              /ORGANIZATION-STUDY-SITE_2021W02_seizure-report.csv*
          /sleep
            /weekly
              /ORGANIZATION-STUDY-SITE_2021W02_sleep-report.csv
          /wearing_detection
            /weekly
              /ORGANIZATION-STUDY-SITE_2021W02_wearing-detection-report.csv
```

* Report or biomarker are generated at the end of the study.
** Step count is generated at the end of study only in case of usage with belt clip in lumbar placement
*** Demographics file is generated only if demographics collection is enabled for the study and data is available.

organization_id

The Organization ID is an alphanumeric ID identifying the organization, consisting of a minimum of 3 characters and a maximum of 20 characters.

study_id

The Study ID is an alphanumeric ID identifying the study, consisting of a maximum of 10 characters.

site_id

The Site ID is an alphanumeric ID identifying each site, consisting of a maximum of 10 characters.

date

The date is represented in YYYY-MM-DD format (e.g., 2021-01-14) and it refers to the day in Coordinated Universal Time (UTC) time.

participant_id

The Participant ID is an alphanumeric ID uniquely identifying a participant within each site, consisting of a maximum of 10 characters.

EmbracePlus_sn

The Serial Number of the EmbracePlus.

filename_demographics

```
1 # Structure
2 [organization_id]-[study_id]-[site_id]_demographics.csv
3
4 # Example
5 ORGANIZATION-STUDY-SITE_demographics.csv
```

filename_metadata

```
# Structure
[study_id]-[site_id]-[participant_id]_metadata.csv

# Example
STUDY-SITE-AB001_metadata.csv
```

filename_aggregated_per_minute

```
# Structure
[study_id]-[site_id]-[participant_id]_[year]-[month]-[week]_[biomarker_name].csv

# Example
STUDY-SITE-AB001_2021-01-14_eda.csv
```

biomarker_name can be:

- accelerometers-std
- actigraphy-counts
- activity-classification
- activity-counts
- activity-intensity
- body-position
- double-support
- eda
- gait-speed
- mets
- prv
- pulse-rate
- respiratory-rate
- sleep-detection
- spo2
- stance-time
- stance-time-asymmetry
- step-counts
- step-length
- step-length-asymmetry
- step-time
- step-time-asymmetry
- step-time-std
- step-time-variance
- stride-length
- stride-time
- swing-time
- swing-time-asymmetry
- temperature
- wearing-detection

filename_raw_data

```
# Structure
[study_id]-[site_id]-[participant_id]_[timestamp].avro

# Example
STUDY-SITE-AB001_1610609653.avro
```

The Avro filename contains the start timestamp, in Unix timestamp (s), of the file content.

filename_report

```
# Structure (weekly report)
[organization_id]-[study_id]-[site_id]_[year]W[week]_[report_name].csv
# Structure (daily report)
[organization_id]-[study_id]-[site_id]_[year]-[month]-[day]_[report_name].csv

# Example
ORGANIZATION-STUDY-SITE_2021W02_wearing-detection-report.csv
```

report_name can be:

- 6mwt-report
- actigraphy-report
- autonomic-arousal-report
- extended-actigraphy-report
- gait-features-report
- health-metrics-report
- seizure-report
- sleep-report
- wearing-detection-report

03

File types

File types

The Empatica Health Monitoring Platform uses two formats of files: Avro and CSV. There are multiple tools, libraries, and software that can be used to access and read the files generated by the platform.

AVRO FILES

Apache Avro™ is an open-source project that provides data serialization and data exchange services. It provides a compact, fast, binary data format, which allows for a significant reduction in the size of the stored data files. Avro files (extension .avro) use of high-level schemas which are stored within the files themselves; this process ensures that the file can be opened and processed later by any program (Apache Avro™ 1.11.1 Documentation). Avro files are self-contained: they expose the schema defining the format type, and the sampling rate of the data recorded by the EmbracePlus.

Avro data format is platform-independent, and it offers an official API for Python, Java, C, C#, and C++. It is supported for MATLAB interface as well:

- Getting started with Python: Apache Avro™ 1.11.1 Getting Started (Python)
- MATLAB Avro interface: GitHub - mathworks-ref-arch/matlab-avro: MATLAB interface for Apache Avro files.
- Getting started with Java: Apache Avro™ 1.11.1 Getting Started (Java)
- Avro C API: Avro C
- Avro C++ API: Avro C++: Main Page
- Avro C# API: Avro C#: Main Page

An example python script to convert a single EmbracePlus raw data Avro file to a CSV file for each sensor data is provided in the Appendix of this document.

CSV FILES

The content of the CSV files is comma-separated. Within each row, each value is separated with a comma (,) character. Each CSV file includes a header, the first row of the file, which describes the content of the file itself and gives a name to each column.

CSV format is human-readable and represents a widespread standard for data exchange. This format can be easily opened with a variety of different tools, including spreadsheet software (e.g. Microsoft Excel, LibreOffice Calc, Apple Numbers, Google Sheets) or programming languages. Programming languages libraries for opening and processing CSV files include:

- Python: csv — CSV File Reading and Writing — Python 3.11.2 documentation
- MATLAB: Read matrix from file - MATLAB readmatrix
- Go: csv package - encoding/csv - pkg.go.dev
- .NET: How to: read from comma-delimited text files - Visual Basic

04

Data format

Data format

> ⚠ The availability of the different raw data, biomarkers, and reports depends on the study configuration.
> For further information on study configurations, please refer to your account manager.

RAW DATA - AVRO SCHEMA

Raw data is stored in Avro files. EmbracePlus creates a new Avro file about every 30 minutes of data collected. A single Avro file contains a continuous portion of data. If a gap occurs (e.g. because of a charging period), then a new file is created. The content of the Avro file is detailed in the table below.

General information

Schema version	Version of the Avro schema used
Firmware version	Version of the EmbracePlus Firmware
Hardware version	Version of the EmbracePlus Hardware
Algorithm version	Version of the algorithm used for signal processing
Timezone	Delta in seconds from UTC time at the location of the participant
Enrollment	Organization, study, site, and participant identifiers in Empatica's systems
Device Serial Number	EmbracePlus Serial Number
Device model	Device model

Raw Data/Accelerometer

Timestamp start	(long) timestamp in UTC (µs) representing the start of the accelerometer data in the file.
Sampling frequency	(float) sampling frequency (Hz) of the sensor.
IMU parameters	physical/digital Inertial Measurement Unit (IMU) parameters: ◦ Physical min: (int) the minimum physical value that can be recorded. ◦ Physical max: (int) the maximum physical value that can be recorded. ◦ Digital min: (int) the minimum value that the Analog to Digital converter (ADC) can produce. ◦ Digital max: (int) the maximum value that the ADC can produce.
x	(array/int) acceleration in the x-axis (ADC counts).
y	(array/int) acceleration in the y-axis (ADC counts).
z	(array/int) acceleration in the z-axis (ADC counts).

Raw Data/Gyroscope

Timestamp start	(long) timestamp in UTC (µs) representing the start of the gyroscope data in the file.
Sampling frequency	(float) sampling frequency (Hz) of the sensor.
IMU parameters	physical/digital IMU parameters: ◦ Physical min: (int) the minimum physical value that can be recorded. ◦ Physical max: (int) the maximum physical value that can be recorded. ◦ Digital min: (int) the minimum value that ADC can produce. ◦ Digital max: (int) the maximum value that ADC can produce.
x	(array/int) angular velocity in the x-axis (ADC counts).
y	(array/int) angular velocity in the y-axis (ADC counts).
z	(array/int) angular velocity in the z-axis (ADC counts).

Raw Data/Temperature

Timestamp start	(long) timestamp in UTC (µs) representing the start of the temperature data in the file.
Sampling frequency	(float) sampling frequency (Hz) of the sensor.
Values	(array/float) temperature (°C).

Raw Data/ElectroDermal Activity (EDA)

Timestamp start	(long) timestamp in UTC (µs) representing the start of the EDA's data in the file.
Sampling frequency	(float) sampling frequency (Hz) of the sensor.
Values	(array/float) electrodermal activity (µS).

Raw Data/Steps

Timestamp start	(long) timestamp in UTC (µs) representing the start of the steps data in the file.
Sampling frequency	(float) sampling frequency (Hz).
Values	(array/int) number of steps taken in the time interval defined by the sampling frequency (counts).

Raw Data/Blood Volume Pulse (BVP)

Timestamp start	(long) timestamp in UTC (µs) representing the start of the BVP data in the file.
Sampling frequency	(float) sampling frequency (Hz) of the BVP.
Values	(array/float) light absorption (nW).

Raw Data/Tags

Tags Time [us]	(array/long) timestamps in UTC (µs) of the event marks.

Raw Data/Systolic Peaks

Peaks Times [ns]	(array/long) timestamps in UTC (ns) of systolic peaks' positions.

DIGITAL BIOMARKERS - CSV

The aggregated per minute digital biomarker files contain all the 1-minute physiological parameters calculated by Empatica's proprietary algorithms. Every digital biomarker is stored in a separate CSV file. The tables below describe the content of each digital biomarker provided.

Accelerometer Magnitude Standard Deviation

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
accelerometers_std_g	Standard deviation of the accelerometer magnitude	g
missing_value_reason	Reason for missing value	N/A

Actigraphy Counts

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
count_x_axis	Actigraphy counts on x axis	counts
count_y_axis	Actigraphy counts on y axis	counts
count_z_axis	Actigraphy counts on z axis	counts
vector_magnitude	Actigraphy counts vector magnitude	counts
missing_value_reason	Reason for missing value	N/A

Activity Classification

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
activity_class	Detected activity class	N/A
missing_value_reason	Reason for missing value	N/A

Activity Counts

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
activity_counts	Activity counts	counts
missing_value_reason	Reason for missing value	N/A

Activity Intensity

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
activity_intensity	Detected physical activity intensity levels	N/A
missing_value_reason	Reason for missing value	N/A

Body Position

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
body_position_left	Detected body position if device worn on left wrist	N/A
body_position_right	Detected body position if device worn on right wrist	N/A
missing_value_reason	Reason for missing value	N/A

Please refer to either body_position_left or body_position_right according to which wrist the device was worn on during data collection. E.g., if EmbracePlus user wore the device on the right wrist, please consider body_position_right and vice versa.

Double Support

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
double_support_in_gait	Proportion of the stride time during which both feet are simultaneously in contact with the ground	%
initial_double_support_in_gait	Proportion of the stride time during initial phase of double support	%
terminal_double_support_in_gait	Proportion of the stride time during terminal phase of double support	%
missing_value_reason	Reason for missing value	N/A

Electrodermal Activity (EDA)

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
eda_scl_usiemens	EDA skin conductance level	uS
missing_value_reason	Reason for missing value	N/A

Gait Speed

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
gait_speed_mps	Gait speed	m/s
missing_value_reason	Reason for missing value	N/A

METs

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
met	Metabolic Equivalent of Task	MET
missing_value_reason	Reason for missing value	N/A

Pulse Rate

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the union of organization, study, site, and participant IDs	N/A
pulse_rate_bpm	Pulse Rate	beats per minute, BPM
missing_value_reason	Reason for missing value	N/A

Pulse Rate Variability (PRV)

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
prv_rmssd_ms	Root mean square of successive differences between normal heartbeats (RMSSD) from PRV	ms
missing_value_reason	Reason for missing value	N/A

Confidential to Empatica, Inc.

Respiratory Rate

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
respiratory_rate_brpm	Respiratory Rate	breaths per minute, BRPM
missing_value_reason	Reason for missing value	N/A

Sleep Detection

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the union of organization, study, site, and participant IDs	N/A
sleep_detection_stage	Sleep stage	N/A
missing_value_reason	Reason for missing value	N/A

`sleep_detection_stage` can take one of the following values:

- `0`: Awake;
- `101`: Sleep;
- `102`: Wake;
- `300`: Interruption, i.e. long wake during a sleep period.

SpO2

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
spo2_percentage	SpO2 level	percentage
missing_value_reason	Reason for missing value	N/A

Stance Time

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
stance_time_s	Stance Time	s
stance_time_right_s	Stance Time for right leg	s
stance_time_left_s	Stance Time for left leg	s
missing_value_reason	Reason for missing value	N/A

Stance Time Asymmetry

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the union of organization, study, site, and participant IDs	N/A
stance_time_asymmetry_s	Stance Time Asymmetry	s
missing_value_reason	Reason for missing value	N/A

Step Counts

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
step_counts	Number of steps	counts
missing_value_reason	Reason for missing value	N/A

Step Length

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
step_length_m	Step Length	m
step_length_right_m	Step Length for right leg	m
step_length_left_m	Step Length for right leg	m
missing_value_reason	Reason for missing value	N/A

Step Length Asymmetry

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the union of organization, study, site, and participant IDs	N/A
step_length_asymmetry_m	Step Length Asymmetry	m
missing_value_reason	Reason for missing value	N/A

Step Time

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
step_time_s	Step Time	s
step_time_right_m	Step Time for right leg	s
step_time_left_m	Step Time for right leg	s
missing_value_reason	Reason for missing value	N/A

Step Time Asymmetry

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the union of organization, study, site, and participant IDs	N/A
step_time_asymmetry_s	Step Time Asymmetry	s
missing_value_reason	Reason for missing value	N/A

Step Time Standard Deviation

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
step_time_std_s	Step Time Standard Deviation	s
missing_value_reason	Reason for missing value	N/A

Step Time Variance

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
step_time_variance_s	Step Time Variance	s^2
missing_value_reason	Reason for missing value	N/A

Stride Length

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the union of organization, study, site, and participant IDs	N/A
stride_length_m	Stride Length	m
missing_value_reason	Reason for missing value	N/A

Stride Time

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
stride_time_s	Stride Time	s
missing_value_reason	Reason for missing value	N/A

Swing Time

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
swing_time_s	Swing Time	s
swing_time_right_s	Swing Time for right leg	s
swing_time_left_s	Swing Time for left leg	s
missing_value_reason	Reason for missing value	N/A

Swing Time Asymmetry

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the union of organization, study, site, and participant IDs	N/A
swing_time_asymmetry_s	Swing Time Asymmetry	s
missing_value_reason	Reason for missing value	N/A

Temperature

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
temperature_celsius	Skin Temperature	°C
missing_value_reason	Reason for missing value	N/A

Wearing Detection

Column	Description	Unit
timestamp_unix	Unix timestamp, in ms	ms
timestamp_iso	UTC timestamp, ISO 8601 format	N/A
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	N/A
wearing_detection_percentage	Wearing Detection percentage	percentage
missing_value_reason	Reason for missing value	N/A

The aggregated data is provided each minute, for the entire duration of the data collection. Missing data can be due to different reasons; the possible values for the `missing_value_reason` field are:

- `device_not_recording`, when no sensor data is available due to the EmbracePlus battery being depleted or the EmbracePlus being in charging mode.
- `device_not_worn_correctly`, when the EmbracePlus is detected as off-wrist. Off-wrist segments are identified as periods of time with the wearing detection lower than 75%. This missing value reason applies only to the following sensor modes: Pulse Rate Pro, SpO2 Pro and Actigraphy Pro.
- `worn_during_motion`, when the EmbracePlus is worn during motion conditions. This condition may affect the PRV, Respiratory Rate, and SpO2 biomarkers.
- `worn_with_low_signal_quality`, when the EmbracePlus is worn but signal quality does not allow the biomarker computation. This condition may affect PR, PRV, Respiratory Rate, SpO2 and advanced gait biomarkers.
- `walking_not_detected`, when the EmbracePlus is not detecting walking conditions. This condition may affect the gait biomarkers.

REPORTS - CSV

The reports provide a summary overview of different physiological and behavioral metrics for each participant in the study. Each report is stored in a separate CSV file. The tables below describe the content of the different reports.

Actigraphy report

Column	Description	Example
participant_full_id	Identifier composed of the union of organization, study, site, and participant IDs	ORGANIZATION-STUDY-SITE-AB001
device_sn	Serial number of the EmbracePlus	3ABCDEFGHI
date	The string representation of the date in the format: YYYY-MM-DD	2021-11-07
timezone_offset_s	Offset in seconds relative to GMT (e.g. -18000 is equivalent to GMT-5)	-18000
weekday	UTC day of the week.	SAT
daily_steps	Total daily steps **	2435
locomotion_time_min	Total daily time classified as WALK and RUN, in minutes **	306
movement_time_min	Total daily time classified as GENERIC, WALK and RUN, in minutes **	840
pa_sedentary_min	Total daily time spent in sedentary activities, in minutes	1005
pa_light_min	Total daily time spent in light-intensity activities (LPA), in minutes	135
pa_moderate_min	Total daily time spent in moderate-intensity activities (MPA), in minutes	99
pa_vigorous_min	Total daily time spent in vigorous-intensity activities (VPA), in minutes	201
non-sedentary_time_min	Total daily time classified as pa_light_min, pa_moderate_min and pa_vigorous_min, in minutes	435
mvpa_min	Total daily time classified as pa_moderate_min and pa_vigorous_min, in minutes	300
left_arm_angle_90-180_degrees_time_min	Total daily time spent with the angle between left arm and gravitational vector greater than 90° (if device worn on left wrist), in minutes. * **	21.30
right_arm_angle_90-180_degrees_time_min	Total daily time spent with the angle between right arm and gravitational vector greater than 90° (if device worn on right wrist), in minutes. * **	21.30
m6min_met	Sum of METs in the six minutes of maximum daily activity [MET]	22.65

* Reference posture for measuring the arm elevation angle: upper arm parallel to the vertical line and pointing the ground is considered 0°

** Column availability depends on device placement.

Autonomic Arousal report*

Column	Description	Example
participant_full_id	Identifier composed of the union of organization, study, site, and participant IDs	ORGANIZATION-STUDY-SITE-AB001
device_sn	Serial number of the EmbracePlus	3ABCDEFGHI
autonomic_arousal_peak_time_unix	Unix timestamp of detected peak of arousal, in seconds	1633278713
autonomic_arousal_peak_time_iso	Time of detected peak of arousal, in ISO 8601 format	2021-10-03T18:31:53+02:00

* Autonomic arousal reports are computed at the end of the study.

Extended Actigraphy report*

Column	Description	Example
participant_full_id	Identifier composed of the union of organization, study, site, and participant IDs	ORGANIZATION-STUDY-SITE-AB001
device_sn	Serial number of the EmbracePlus	3ABCDEFGHI
date	The string representation of the date in the format: YYYY-MM-DD	2021-11-07
timezone_offset_s	Offset in seconds relative to GMT (e.g. -18000 is equivalent to GMT-5)	-18000
weekday	UTC day of the week.	SAT
distance_km	Total distance travelled in one day, in km **	6.70
calories_kcal	Total amount of calories burned in one day, in Kcal	1250

* Extended Actigraphy reports are computed at the end of the study.
** Column availability depends on device placement.

Gait Features report*

Column	Description	Example
participant_full_id	Identifier composed of the union of organization, study, site, and participant IDs	ORGANIZATION-STUDY-SITE-AB001
device_sn	Serial number of the EmbracePlus	3ABCDEFGHI
date	The string representation of the date in the format: YYYY-MM-DD	2021-11-07
timezone_offset_s	Offset in seconds relative to GMT (e.g. -18000 is equivalent to GMT-5)	-18000
weekday	UTC day of the week.	SAT
walked_distance_km	Total distance travelled in one day recorded during walking only, in km	3.20
walked_daily_steps	Total daily steps recorded during walking only	2114
gait_speed_95th_centile_mps	Gait speed 95th centile (SV95C), in mps	1.22
step_time_mean_s	Daily mean step time, in seconds	0.59
step_time_right_mean_s	Daily mean step time for right leg, in seconds	0.59
step_time_left_mean_s	Daily mean step time for left leg, in seconds	0.59
step_time_asymmetry_mean_s	Daily mean step time asymmetry, in seconds	0.1
step_length_mean_m	Daily mean step length, in meters	0.76
step_length_right_mean_m	Daily mean step length for right leg, in meters	0.76
step_length_left_mean_m	Daily mean step length for left leg, in meters	0.76
step_length_asymmetry_mean_m	Daily mean step length asymmetry, in meters	0.1
stance_time_mean_s	Daily mean stance time, in seconds	0.61
stance_time_right_mean_s	Daily mean stance time for right leg, in seconds	0.61
stance_time_left_mean_s	Daily mean stance time for left leg, in seconds	0.61
stance_time_asymmetry_mean_s	Daily mean stance time asymmetry, in seconds	0.1
swing_time_mean_s	Daily mean swing time, in seconds	0.16
swing_time_right_mean_s	Daily mean swing time for right leg, in seconds	0.16
swing_time_left_mean_s	Daily mean swing time for left leg, in seconds	0.16
swing_time_asymmetry_mean_s	Daily mean swing time asymmetry, in seconds	0.1
double_support_in_gait_mean	Mean proportion of the stride time during which both feet are simultaneously in contact with the ground	50.03
initial_double_support_in_gait_mean	Daily mean proportion of the stride time during initial phase of double support	20.12
terminal_double_support_in_gait_mean	Daily mean proportion of the stride time during terminal phase of double support	30.4
daily_walking_bouts	Count of daily walking bouts	35

* Gait Features report is computed at the end of the study.

Health Metrics Report*

Column	Description	Example
participant_full_id	Identifier composed of the union of organization, study, site, and participant IDs	ORGANIZATION-STUDY-SITE-AB001
device_sn	Serial number of the EmbracePlus	3ABCDEFGHI
date	The string representation of the date in the format: YYYY-MM-DD	2021-11-07
timezone_offset_s	Offset in seconds relative to GMT (e.g. -18000 is equivalent to GMT-5)	-18000
weekday	UTC day of the week.	SAT
temperature_mean_celsius	Mean of all valid temperature values for reference date, in degrees Celsius	30.79
temperature_std_celsius	Standard deviation of all valid temperature values for reference date, in degrees Celsius	0.95
temperature_p01_celsius	1st percentile of all valid temperature values for reference date, in degrees Celsius	28.18
temperature_p25_celsius	25th percentile of all valid temperature values for reference date, in degrees Celsius	30.10
temperature_p50_celsius	50th percentile of all valid temperature values for reference date, in degrees Celsius	30.98
temperature_p75_celsius	75th percentile of all valid temperature values for reference date, in degrees Celsius	31.52
temperature_p99_celsius	99th percentile of all valid temperature values for reference date, in degrees Celsius	32.41
temperature_coverage_min	Number of minutes of valid temperature values for reference date*	1060
pulse_rate_mean_bpm	Mean of all valid pulse rate values for reference date, in bpm	75.17
pulse_rate_std_bpm	Standard deviation of all valid pulse rate values for reference date, in bpm	17.98
pulse_rate_p01_bpm	1st percentile of all valid pulse rate values for reference date, in bpm	55.60
pulse_rate_p25_bpm	25th percentile of all valid pulse rate values for reference date, in bpm	64.74
pulse_rate_p50_bpm	50th percentile of all valid pulse rate values for reference date, in bpm	69.72
pulse_rate_p75_bpm	75th percentile of all valid pulse rate values for reference date, in bpm	79.08
pulse_rate_p99_bpm	99th percentile of all valid pulse rate values for reference date, in bpm	147.42
pulse_rate_coverage_min	Number of minutes of valid pulse rate values for reference date*	1060
prv_rmssd_mean_ms	Mean of all valid pulse rate variability values for reference date, in ms	325.29
prv_rmssd_std_ms	Standard deviation of all valid pulse rate variability values for reference date, in ms	88.50
prv_rmssd_p01_ms	1st percentile of all valid pulse rate variability values for reference date, in ms	71.33
prv_rmssd_p25_ms	25th percentile of all valid pulse rate variability values for reference date, in ms	278.57
prv_rmssd_p50_ms	50th percentile of all valid pulse rate variability values for reference date, in ms	338.83

prv_rmssd_p75_ms	75th percentile of all valid pulse rate variability values for reference date, in ms	385.31
prv_rmssd_p99_ms	99th percentile of all valid pulse rate variability values for reference date, in ms	490.81
prv_rmssd_coverage_min	Number of minutes of valid pulse rate variability values for reference date*	1060
eda_scl_mean_usiemens	Mean of all valid EDA scl values for reference date, in uSiemens	0.11
eda_scl_std_usiemens	Standard deviation of all valid EDA scl values for reference date, in uSiemens	0.14
eda_scl_p01_usiemens	1st percentile of all valid EDA scl values for reference date, in uSiemens	0.00
eda_scl_p25_usiemens	25th percentile of all valid EDA scl values for reference date, in uSiemens	0.02
eda_scl_p50_usiemens	50th percentile of all valid EDA scl values for reference date, in uSiemens	0.08
eda_scl_p75_usiemens	75th percentile of all valid EDA scl values for reference date, in uSiemens	0.14
eda_scl_p99_usiemens	99th percentile of all valid EDA scl values for reference date, in uSiemens	0.70
eda_scl_coverage_min	Number of minutes of valid EDA scl values for reference date*	1060
respiratory_rate_mean_brpm	Mean of all valid respiratory rate values for reference date, in brpm	0.70
respiratory_rate_std_brpm	Standard deviation of all valid respiratory rate values for reference date, in brpm	3.71
respiratory_rate_p01_brpm	1st percentile of all valid respiratory rate values for reference date, in brpm	0.00
respiratory_rate_p25_brpm	25th percentile of all valid respiratory rate values for reference date, in brpm	0.00
respiratory_rate_p50_brpm	50th percentile of all valid respiratory rate values for reference date, in brpm	0.00
respiratory_rate_p75_brpm	75th percentile of all valid respiratory rate values for reference date, in brpm	0.00
respiratory_rate_p99_brpm	99th percentile of all valid respiratory rate values for reference date, in brpm	20.39
respiratory_rate_coverage_min	Number of minutes of valid respiratory rate values for reference date*	950
spo2_mean_percentage	Mean of all valid SPO2 percentage values for reference date	97.57
spo2_std_percentage	Standard deviation of all valid SPO2 percentage values for reference date	1.43
spo2_p01_percentage	1st percentile of all valid SPO2 percentage values for reference date	95.6
spo2_p25_percentage	25th percentile of all valid SPO2 percentage values for reference date	97
spo2_p50_percentage	50th percentile of all valid SPO2 percentage values for reference date	98
spo2_p75_percentage	75th percentile of all valid SPO2 percentage values for reference date	98.33
spo2_p99_percentage	99th percentile of all valid SPO2 percentage values for reference date	99.64
spo2_coverage_min	Number of minutes of valid SPO2 percentage values for reference date*	950

* Coverage depends on each biomarker filter. You can access missing value reasons of each biomarker on EHMP spec sheet.

Confidential to Empatica, Inc.

Seizure report*

Column	Description	Example
participant_full_id	Identifier composed of the union of organization, study, site, and participant IDs	ORGANIZATION-STUDY-SITE-AB001
device_sn	Serial number of the EmbracePlus	3ABCDEFGHI
seizure_time_unix	Unix timestamp of the onset of the detected convulsive seizure, in seconds	1633278713
seizure_time_iso	Time of the onset of the detected convulsive seizure, in ISO 8601 format	2021-10-03T18:31:53+02:00
type	Type of the detected convulsive seizure	convulsive

* Seizure reports are computed at the end of the study.

Six Minute Walking Test (6MWT) report *

Column	Description	Example
participant_full_id	Identifier composed of the union of organization, study, site, and participant IDs	ORGANIZATION-STUDY-SITE-AB001
device_sn	Serial number of the EmbracePlus	3ABCDEFGHI
timezone_offset_s	Offset in seconds relative to GMT (e.g. -18000 is equivalent to GMT-5)	-18000
6mwt_start_time_unix	6mwt start time, Unix timestamp in seconds	1604697328
6mwt_start_time_iso	6mwt start time, in ISO 8601 format	2021-11-06T23:15:28+02:00
6mwt_stop_time_unix	6mwt stop time, Unix timestamp in seconds	1604697688
6mwt_stop_time_iso	6mwt stop time, in ISO 8601 format	2021-11-06T23:21:28+02:00
6mwt_steps	Number of steps during the test	642
6mwt_gait_speed_mps	Average gait speed during test, in mps **	1.23
6mwt_distance_m	Total distance travelled during the test, in m	450
6mwt_PR_bpm	Average pulse rate during test, in bpm. **	59

* Six Minute Walking Test reports are computed at the end of the study.
** Column availability depends on sensor mode

Sleep report

Column	Description	Example
participant_full_id	Identifier composed of the union of organization, study, site, and participant IDs	ORGANIZATION-STUDY-SITE-AB001
device_sn	Serial number of the EmbracePlus	3ABCDEFGHI
date	The string representation of the date in the format: YYYY-MM-DD	2021-11-07
weekday	UTC day of the week.	SAT
time_in_bed_start_time_unix	Time in bed start time, Unix timestamp in seconds	1604697328
time_in_bed_start_time_iso	Time in bed start time, in ISO 8601 format	2021-11-06T23:15:28+02:00
time_in_bed_stop_time_unix	Time in bed stop time, Unix timestamp in seconds	1604733448
time_in_bed_stop_time_iso	Time in bed stop time, in ISO 8601 format	2021-11-07T09:17:28+02:00
sleep_onset_latency_min	Sleep onset latency (SOL) is the amount of time from the time in bed start time to the actual sleep onset. It is computed from the actigraphy data, in minutes	40
total_sleep_time_min	Total sleep time (TST), the estimated time spent actually sleeping after the sleep onset, in minutes	536
wake_after_sleep_onset_min	Wake after sleep onset (WASO), the amount of time spent awake after the SOL, in minutes.	26
time_in_bed_duration_min	Time in bed (TIB) duration, the total length of the time in bed period, from the time in bed start time to the time in bed stop time, in minutes. If there are no gaps in the data, TIB = SOL + TST + WASO.	602
interruptions_count	Number of interruptions detected after the sleep onset. An interruption is defined as a break in sleep when movements appear to be more intense than typical nocturnal movements (e.g., when going to the bathroom).	2
wake_bouts	Number of contiguous sections categorized as wake after the sleep onset.	15
sleep_efficiency_percentage	Sleep efficiency (SE), expressed as a percentage. SE is defined as TST / TIB.	0.89
sleep_fragmentation_percentage	Sleep fragmentation (SF), an index of restlessness during the sleep period expressed as a percentage. The SF is calculated as the sum of two percentages: the proportion of minutes after the sleep onset that are mobile (i.e., where the activity count is ≥ 2) and the proportion of all the sleep bouts after the sleep onset that are ≤ 1 min in duration.	0.23

Wearing Detection report

Column	Description	Example
participant_full_id	Identifier composed of the organization ID, study ID, site ID, and participant ID	ORGANIZATION-STUDY-SITE-AB001
date_utc	String representation of the UTC date in the format: YYYY-MM-DD	2021-11-07
weekday	UTC day of the week	SAT
device_worn_correctly_min	Time spent with the device worn correctly, in minutes (wearing detection ≥ 75%)	1250
device_not_worn_correctly_min	Time spent with the device not worn correctly, in minutes	200
device_not_recording_min	Time spent with the device off, in minutes (null wearing detection)	35
status*	Status of the data collection	monitoring
monitoring_start_date_utc	UTC start date of monitoring period in the format: YYYY-MM-DD	2021-10-05
monitoring_end_date_utc	UTC end date of monitoring period in the format: YYYY-MM-DD	2021-11-11

*status can be one of the following:

- waiting: Participant has not yet been enrolled;
- monitoring: Participant is actively enrolled in the data collection;
- early_term: Participant did not complete the study duration (e.g., in case of screen failures or withdrawals);
- full_term: Participant has completed the full duration of the study.

monitoring_start_date_utc and monitoring_end_date_utc can be empty in case the study is not started/ended yet.

DEMOGRAPHICS - CSV

The demographics file contains the demographics information for the site participants that were provided through the Care portal. The content of the file is detailed in the table below.

Column	Description	Example
`participant_full_id`	Identifier composed of the organization ID, study ID, site ID, and participant ID	`ORGANIZATION-STUDY-SITE-AB001`
`year_of_birth`	Year of birth of the participant	`2000`
`height_in_cm`	Participant's height in centimeters	`165`
`weight_in_kg`	Participant's weight in kilograms	`50`
`sex_M/F`	Participant's sex	`F`

METADATA - CSV

The metadata file contains the list of the changes of metadata for each participant. The content of the file is detailed in the table below.

Column	Description	Example
`timestamp_unix`	Unix timestamp for the record, in ms	`1626697035505`
`timestamp_iso`	UTC Timestamp for the record, in ISO 8601 format	`2021-03-19T12:18:25Z`
`time_offset`	Offset in seconds relative to GMT (e.g. -18000 is equivalent to GMT-5)	`-18000`
`timezone_location`	Name of timezone location. The timezone is set at the first pairing, and it is automatically updated with participant's paired smartphone.	`America/Chicago`
`mobile_platform`	Mobile phone platform (one of ios, android)	`ios`
`mobile_platform_version`	Mobile phone platform version	`14.6.1`
`mobile_app_version`	Version of the app	`3.0.3`
`algo_version`	Version of the algorithm used to calculate the digital biomarkers	`4.4.1`
`embrace_fw_version`	Version of the EmbracePlus Firmware used to collect the data	`2.3.9`
`cdp_version`	Version of the Cloud Data Pipeline used to compute sleep biomarker, reports, and biomarkers computed at the end of a study	`3.6.0`

05

Data Access and Retrieval

Data Access and Retrieval

Your data is securely stored and encrypted within AWS Cloud Services. Access to your data is granted by a set of keys that are unique and should be kept secret. This guide covers three options that allow you to access your data:

- Cyberduck - *Suggested option for users with no programming experience*
- AWS CLI v2
- AWS SDK

Before proceeding with learning more about the access options and how to configure them, take a moment to learn the configuration and access keys required to securely access your data. These will be included in the access keys CSV file shared with you when you generate a new key through the Care Portal. The file includes the following three fields in a single row:

AWS ACCESS KEY ID

This is the AWS Access Key ID to be used when connecting to the AWS S3 bucket.

AWS SECRET ACCESS KEY

This is the AWS Secret Access Key to be used when connecting to the AWS S3 bucket. It should be treated as a password. The AWS Secret Access Key can be requested via Care Portal. Please contact your Empatica Account Manager to have access to Care Portal and Data Access Key generation.

S3 ACCESS URL

This is your organization-specific location where to find your data (it will look similar to the following example: *s3://bucket-name/version/dir-name/*).

⚠ **Important Reminder**:

Access Keys are Confidential: Always treat access keys as secret. Never publish or expose them in any public medium.
Versioning Systems: If you use a versioning system (like Git), ensure that access keys are never saved in plain text or committed. Consider using environment variables or secrets management tools.
Exposing keys can lead to data breaches and unauthorized access.

DATA ACCESS VIA CYBERDUCK

Empatica is not affiliated with nor supporting Cyberduck products.

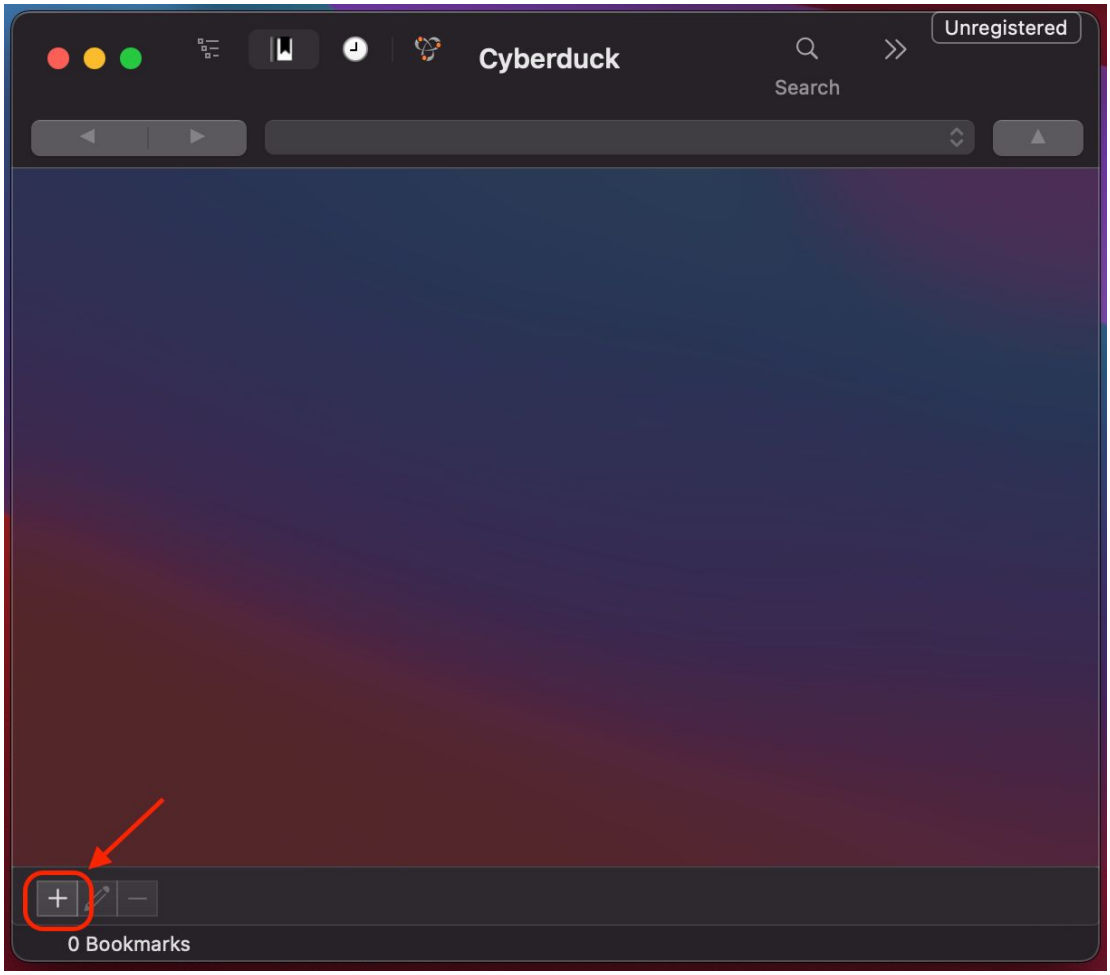
We recommend that users with no programming experience choose this option to access their data.

Cyberduck is a desktop application available for Windows and macOS. It can be downloaded here: https://cyberduck.io/download/

It can be configured for AWS S3 access, enabling FTP-like access to your data, including folder browsing and download.

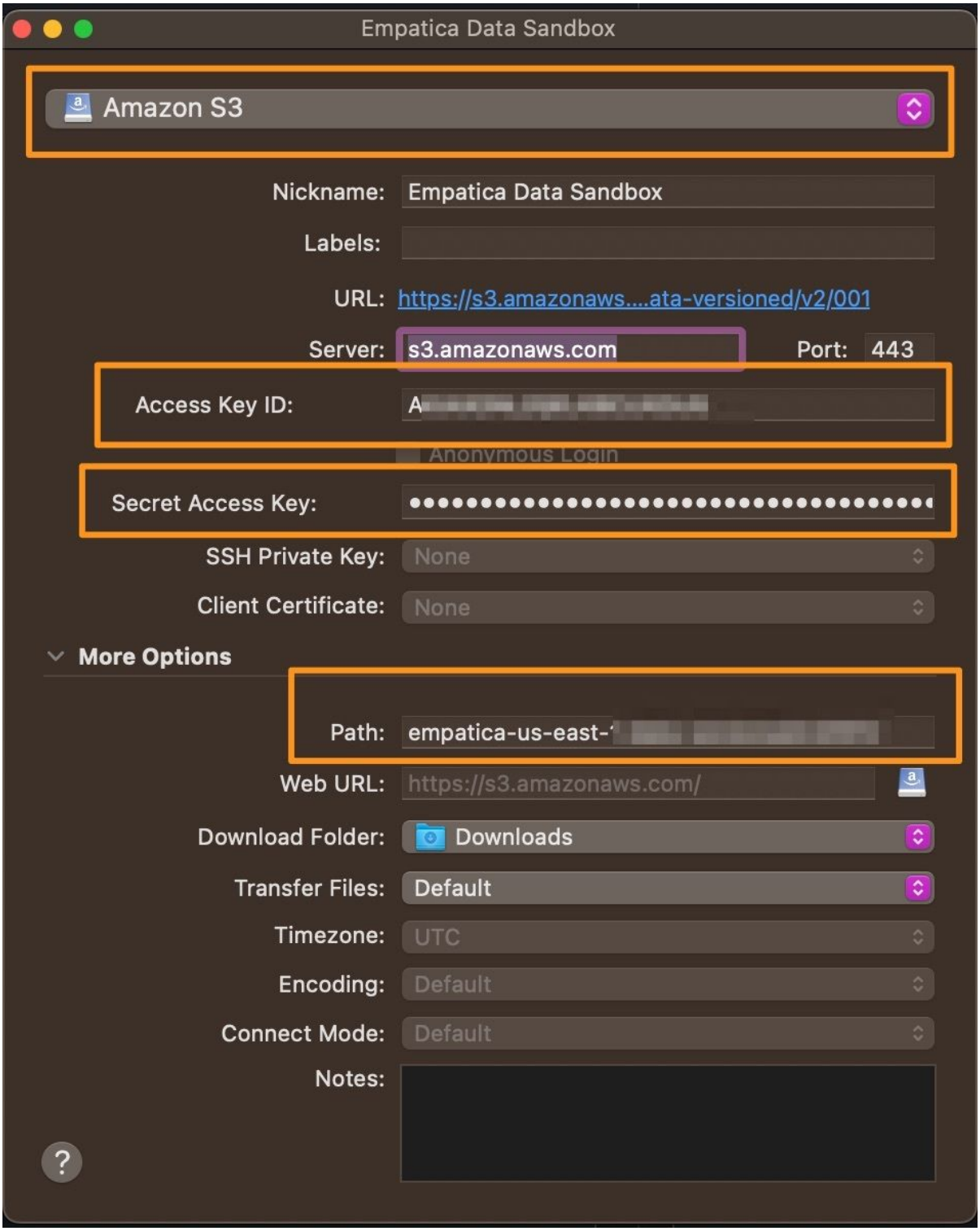
Configuration

Once Cyberduck is installed, you will need to configure the connection to access your data. Start by clicking on the + icon on the bottom left of the window.



A popup window will appear, giving you the possibility to configure the connection with your data.

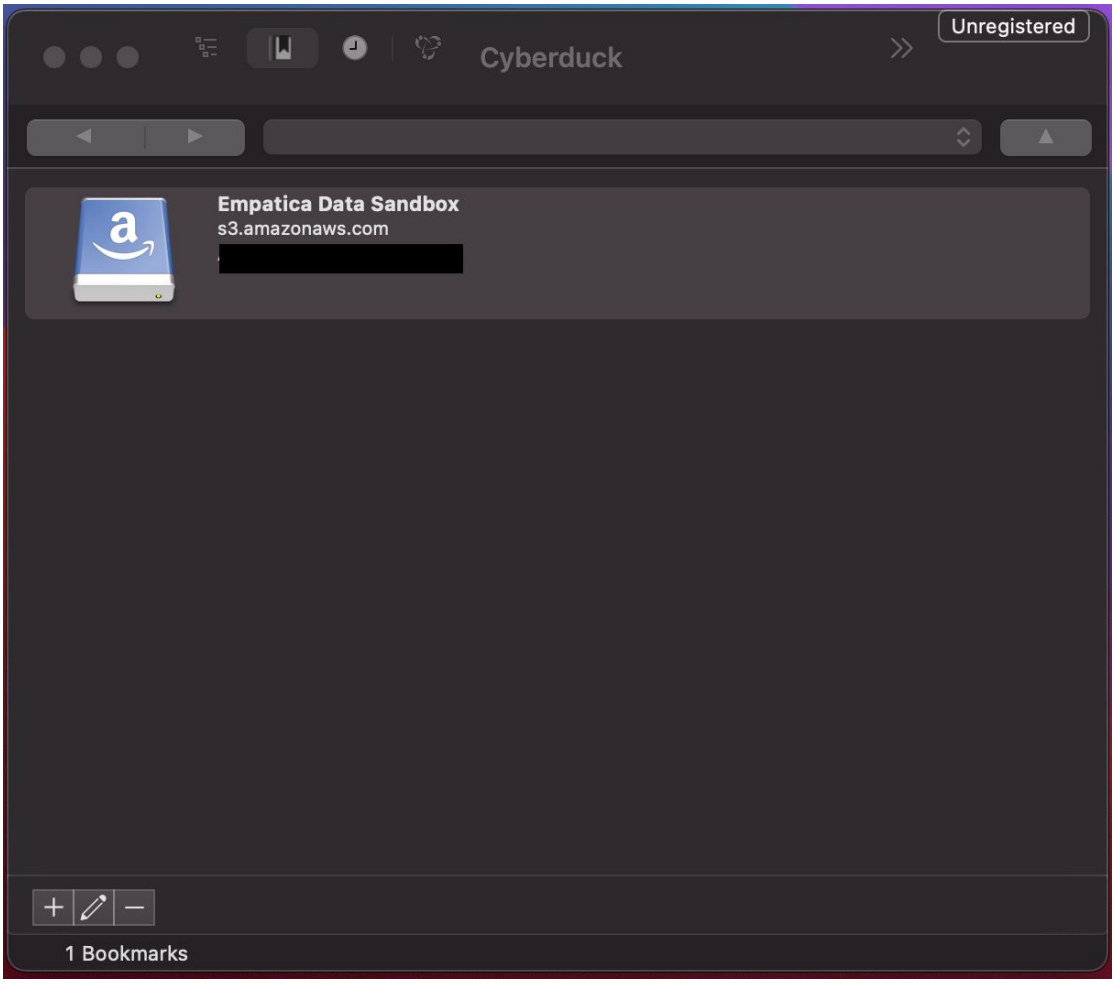
- **Select Amazon S3** as the connection type;
- **Give it a Nickname** that makes it easy to identify the connection, e.g. Empatica Data Sandbox;
- **Add the value** for Access Key ID and Secret Access Key
- **Add the value** for Path. The S3 PATH is obtained from the S3 ACCESS URL removing the leading **s3://** (the format should look like: ***bucket-name/version/dir-name/***)



By closing the popup window, the configuration is automatically saved. You can now connect to your data by double-clicking on the created resource.

If the connection is successful, a new window will open, and from there you will be able to browse through your data.

⚠ Please note that Cyberduck will return an error message if no data have been uploaded yet.



DATA ACCESS VIA AWS CLI

⚠ Empatica is not affiliated with nor supporting AWS products.

By leveraging AWS CLI v2, it's possible to read, copy, and sync data from your data path. The AWS CLI v2 is available to be installed and used from Windows, Linux, and macOS command line terminals.

Install

Windows: Installing, updating, and uninstalling the AWS CLI version 2 on Windows - AWS Command Line Interface
Linux: Installing, updating, and uninstalling the AWS CLI version 2 on Linux - AWS Command Line Interface
macOS: Installing, updating, and uninstalling the AWS CLI version 2 on macOS - AWS Command Line Interface

Usage

ℹ The following procedure shows an example of how to use AWS CLI v2 on macOS. More details on how to extensively use it with S3 is detailed in the official documentation provided by Amazon AWS: Using Amazon S3 with the AWS CLI - AWS Command Line Interface

Setup AWS Credentials

First, permanently set your access credentials using the AWS CLI configure command, prompting the Access Key ID and the Secret Access ID when required. You can leave the Default region name and the Default output format empty:

```
# AWS ACCESS KEY ID and AWS SECRET ACCESS KEY - Provided above
aws configure
AWS Access Key ID [None]: <YOUR_ACCESS_KEY_ID>
AWS Secret Access Key [None]: <YOU_SECRET_ACCESS_KEY_ID>
Default region name [None]: <LEAVE_EMPTY>
Default output format [None]: <LEAVE_EMPTY>
```

For additional information on AWS credentials, please refer to the AWS official documentation.

Copy files locally

To copy the files from the Empatica S3 bucket, first set ACCESS_URL and the output LOCAL_PATH as environment variables:

```
# ACCESS URL - Provided above
export ACCESS_URL=<insert your S3 ACCESS URL here>

# LOCAL PATH - The local path where to store files retrieved from S3
export LOCAL_PATH=<insert a local path here>
```

Then, run the following command to list the files available for your organization:

```
# The list command, recursively listing files available
aws s3 ls ${ACCESS_URL} --recursive
```

Run the following command to locally copy all the files within the folder. This command will override existing files, but will not delete files no longer available on the S3 bucket.

```
# The copy command, recursively copying files available for your organization
aws s3 cp ${ACCESS_URL} ${LOCAL_PATH} --recursive
```

Sync files locally

Conversely, this command will make sure that the files within your local directory are synced with the ones contained in the S3 bucket. Existing unchanged files will not be overwritten, while files that are no longer available on S3, will be deleted.

First, setup environment variables as described in the previous section.
Then, run the following command to sync the files locally:

```
# The sync command, recursively syncing files available for your organization
aws s3 sync ${ACCESS_URL} ${LOCAL_PATH}
```

DATA ACCESS VIA AWS SDK

⚠ Empatica is not affiliated with nor supporting AWS SDK products.

AWS comes with an extensive set of SDKs that can be used to access the S3 bucket. The actual usage of AWS SDK depends on the programming language of your choice:

- JavaScript: https://aws.amazon.com/sdk-for-javascript/
- Python: https://aws.amazon.com/sdk-for-python/
- PHP: https://aws.amazon.com/sdk-for-php/
- .NET: https://aws.amazon.com/sdk-for-net/
- Ruby: https://aws.amazon.com/sdk-for-ruby/
- Java: https://aws.amazon.com/sdk-for-java/
- Go: https://aws.amazon.com/sdk-for-go/
- C++: https://aws.amazon.com/sdk-for-cpp/

Here is an example using the python SDK (boto3) to list all the objects in the S3 bucket. As shown in the DATA ACCESS VIA AWS CLI section, remember to first setup the AWS credentials.

The value for the `BUCKET_NAME` and the `PREFIX` are taken from the S3 Access URL available in your access keys CSV.

For the example below, the S3 Access URL is: `s3://empatica-us-east-1-prod-data/v2/001/`

⚠ Depending on your plan, the S3 Access URL might differ from the one in the example.

- For `BUCKET_NAME` add only the "root" of the URL. Please also remove `s3://` from the beginning of the URL. In this case the value is: `empatica-us-east-1-prod-data`
- For `PREFIX` add the remaining part of the URL which defines the exact path of your data. In this case the value is: `v2/001/`

```
1  import boto3
2
3  BUCKET_NAME = "empatica-us-east-1-prod-data"
4  PREFIX = "v2/001/"
5
6  s3_resource = boto3.resource('s3')
7  bucket = s3_resource.Bucket(BUCKET_NAME)
8  bucket.objects.all()
9  for my_bucket_object in bucket.objects.filter(Prefix = PREFIX):
10    print(my_bucket_object.key)
```

06

Appendix

Appendix

OPEN AN AVRO FILE - PYTHON EXAMPLE

The code below shows how to load and access the fields of an Avro file (NB: this code is tested for the v6 of the Avro files).
The script saves raw data on a csv file for each sensor. Note that the availability of the different raw data depends on the study configuration: some output files could be empty.
The `avro` package needs to be installed in your python environment prior running this script (e.g. prompt `pip install avro` in your terminal).

```
from avro.datafile import DataFileReader
from avro.io import DatumReader
import json
import csv
import os

avro_file_path = "<insert the avro filename with the full path>"
output_dir = "<insert the path to a directory where to save csv files >"

## Read Avro file
reader = DataFileReader(open(avro_file_path, "rb"), DatumReader())
schema = json.loads(reader.meta.get('avro.schema').decode('utf-8'))
data= next(reader)

## Print the Avro schema
print(schema)

## Export sensors data to csv files
# Accelerometer
acc = data["rawData"]["accelerometer"]
timestamp = [round(acc["timestampStart"] + i * (1e6 / acc["samplingFrequency"]))
            for i in range(len(acc["x"]))]
# Convert ADC counts in g
delta_physical = acc["imuParams"]["physicalMax"] - acc["imuParams"]["physicalMin"]
delta_digital  = acc["imuParams"]["digitalMax"]  - acc["imuParams"]["digitalMin"]
x_g = [val * delta_physical / delta_digital for val in acc["x"]]
y_g = [val * delta_physical / delta_digital for val in acc["y"]]
z_g = [val * delta_physical / delta_digital for val in acc["z"]]
with open(os.path.join(output_dir, 'accelerometer.csv'), 'w', newline='') as f:
   writer = csv.writer(f)
   writer.writerow(["unix_timestamp", "x", "y", "z"])
   writer.writerows([[ts, x, y, z] for ts, x, y, z in zip(timestamp, x_g, y_g, z_g)])

# Gyroscope
gyro = data["rawData"]["gyroscope"]
timestamp = [round(gyro["timestampStart"] + i * (1e6 / gyro["samplingFrequency"]))
            for i in range(len(gyro["x"]))]
with open(os.path.join(output_dir, 'gyroscope.csv'), 'w', newline='') as f:
   writer = csv.writer(f)
   writer.writerow(["unix_timestamp", "x", "y", "z"])
   writer.writerows([[ts, x, y, z] for ts, x, y, z in zip(timestamp, gyro["x"], gyro["y"], gyro["z"])])
```

```
# Eda
eda = data["rawData"]["eda"]
timestamp = [round(eda["timestampStart"] + i * (1e6 / eda["samplingFrequency"]))
            for i in range(len(eda["values"]))]
with open(os.path.join(output_dir, 'eda.csv'), 'w', newline='') as f:
   writer = csv.writer(f)
   writer.writerow(["unix_timestamp", "eda"])
   writer.writerows([[ts, eda] for ts, eda in zip(timestamp, eda["values"])])

# Temperature
tmp = data["rawData"]["temperature"]
timestamp = [round(tmp["timestampStart"] + i * (1e6 / tmp["samplingFrequency"]))
            for i in range(len(tmp["values"]))]
with open(os.path.join(output_dir, 'temperature.csv'), 'w', newline='') as f:
   writer = csv.writer(f)
   writer.writerow(["unix_timestamp", "temperature"])
   writer.writerows([[ts, tmp] for ts, tmp in zip(timestamp, tmp["values"])])

# Tags
tags = data["rawData"]["tags"]
with open(os.path.join(output_dir, 'tags.csv'), 'w', newline='') as f:
   writer = csv.writer(f)
   writer.writerow(["tags_timestamp"])
   writer.writerows([[tag] for tag in tags["tagsTimeMicros"]])

# BVP
bvp = data["rawData"]["bvp"]
timestamp = [round(bvp["timestampStart"] + i * (1e6 / bvp["samplingFrequency"]))
            for i in range(len(bvp["values"]))]
with open(os.path.join(output_dir, 'bvp.csv'), 'w', newline='') as f:
   writer = csv.writer(f)
   writer.writerow(["unix_timestamp", "bvp"])
   writer.writerows([[ts, bvp] for ts, bvp in zip(timestamp, bvp["values"])])

# Systolic peaks
sps = data["rawData"]["systolicPeaks"]
with open(os.path.join(output_dir, 'systolic_peaks.csv'), 'w', newline='') as f:
   writer = csv.writer(f)
   writer.writerow(["systolic_peak_timestamp"])
   writer.writerows([[sp] for sp in sps["peaksTimeNanos"]])

# Steps
steps = data["rawData"]["steps"]
timestamp = [round(steps["timestampStart"] + i * (1e6 / steps["samplingFrequency"]))
            for i in range(len(steps["values"]))]
with open(os.path.join(output_dir, 'steps.csv'), 'w', newline='') as f:
   writer = csv.writer(f)
   writer.writerow(["unix_timestamp", "steps"])
   writer.writerows([[ts, step] for ts, step in zip(timestamp, steps["values"])])
```

AVAILABLE PLANS

The Empatica Health Monitoring Platform can be purchased with two different plans: Professional or Enterprise. In the table below, you can check the differences between the two plans. For more information on plans, talk to your account manager.

⊕ Portal features

	Professional	Enterprise
Registration	Self-registration	Supported by dedicated team
Number of sites supported	1	Unlimited
Number of participants supported	1000	Unlimited
Number of study managers supported	1	Multiple
Team Roles Management	✓	✓
Wearing detection report sent by mail	✗	✓
Data visualization graphs	Accelerometer Standard Deviation Electrodermal Activity (EDA): Skin Conductance Level (SCL) Participant Wearing Time Pulse Rate Respiratory Rate Skin Temperature	*Same as Professional plan* + Sleep SpO2 Step Counts
Data access	Data access keys through Amazon S3	*Same as Professional plan*

Data services

Raw data	Accelerometer Gyroscope Blood Volume Pulse Electrodermal Activity Skin Temperature Steps Systolic peaks (IBI) User Tags	*Same as Professional plan*
Digital biomarkers	Accelerometer Magnitude Standard Deviation Actigraphy Counts Activity Classification Activity Counts Activity Intensity Body Position Electrodermal activity (EDA): Skin Conductance Level (SCL) METs Pulse Rate Pulse Rate Variability Respiratory Rate Skin Temperature Sleep Detection Step Counts (wrist) Wearing Detection	*All from professional plan* + Gait Speed (available end of study) SpO2 Spatiotemporal Gait Features (available end of study)

Reports	✕	Actigraphy Autonomic Arousal Extended Actigraphy Gait Features Health Metrics Seizure Six Minute Walking Test (6MWT) Sleep Wearing Detection
Event tagging	✓	✓
Sensor modes	Pulse Rate Pro	*All from Professional plan* + SpO2 Pro Actigraphy Pro Actigraphy Optimised Gyroscope Max Gyroscope Ultra

⊕ Regulatory

Certifications	EmbracePlus is CE marked Care Lab is for experimental / research purpose only	EmbracePlus, Care App and Care Portal are CE marked and FDA cleared

Document History

Version	Description
1.0	• first release
2.0	• removed outdated field in the AVRO schema
3.0	• corrected typographic errors
4.0	• updated document version on the first page • corrected **filename_metadata** description • updated **Access and retrieve data** section
5.0	• added **Document scope and intended audience** section • updated **Data flow** diagram • updated **Data hierarchy** section • updated **Data format** section
6.0	• updated Avro Apache hyperlinks
7.0	• corrected typographic errors • improved description of **Access and retrieve data** section
8.0	• improved description of Empatica Health Monitoring Platform • added plans table • terminology improvement from sensor profile to sensor mode
9.0	• Added Sleep Detection biomarker to Professional plan in the Plans table
10.0	• Added Sleep Detection values • Substituted “sales representative” with “account manager” • Edited Intro section in favor of a more technical oriented description • Moved Cyberduck method up in the list • Added Cyberduck as recommended option for users with no programming experience • Added warning that Cyberduck will return error in case of empty bucket
11.0	• Updated disclaimers on biomarkers and reports that are computed at the end of the study • Update disclaimer on study configuration • corrected the `prv_rmssd_ms` biomarker name • Added the Activity Classification biomarker • Added a script to list bucket objects using python AWS SDK • Updated Plans table
12.0	• Updated **Data hierarchy** section • Added Body Position, METs, Gait Speed and Activity Intensity biomarker • Added Actigraphy and Extended Actigraphy reports • Updated Plans table • Updated **Access and retrieve data** section title to **Data Access and Retrieval**

Version	Description
14.0	• Updated **Data hierarchy** section • Added actigraphy counts biomarker • Added 6mwt report • Biomarkers and reports ordered alphabetically • Added note on data access keys confidentiality • Updated Plans table
15.0	• Removed disclaimer on Gyroscope availability
16.0	• Minor typographical correction
17.0	• Updated biomarkers and reports with gait features • Updated Plans table
18.0	• Added demographics file • Updated metadata field: cdp_version
19.0	• Added health metrics report • Added Gyroscope Ultra

empatica