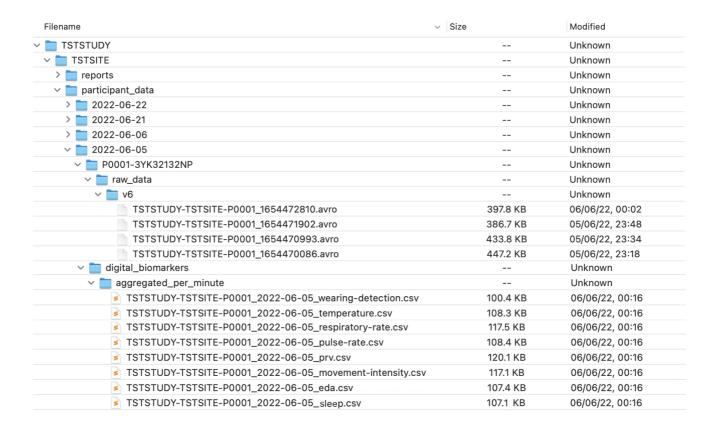
# Data hierarchy

In the Empatica Cloud, the data is organized by site, date, participant ID, and EmbracePlus serial number. Downloaded EmbracePlus data follows the below hierarchy and naming structure.

```
1 # Structure
2 s3://[s3-bucket-name]
3
    /[organization_id]
4
      /[study_id]
5
        /[site_id]
6
          /metadata
7
             /[filename_metadata].csv
          /participant_data
8
9
             /[date]
               /[participant_id]-[EmbracePlus_sn]
10
11
                 /digital_biomarkers
12
                   /aggregated_per_minute
                 /[filename_aggregated_per_minute].csv
13
14
                   /raw_data
15
                     /[schema_version]
16
                        /[filename_raw_data].avro
```



# Data hierarchy

## 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.

# File types and access

#### PROCESSING THE DATA COLLECTED

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 file generated by EmbracePlus.

#### **AVRO FILES**

Avro is an open-source project that provides data serialization and data exchange services. It provides a compact, fast, binary data format, allowing a significant reduction in the size of the stored data files. Avro files (extension avro) make use of high-level schemas, that are stored within the files themselves; this process ensures that the file can be opened and processed later by any program (Apache Avro™ 1.11.0 Documentation). Avro files are self-contained: they expose the schema defining the format type, and the sampling rate of the data the EmbracePlus has recorded.

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™ 1.11.0 Getting Started (Python)

MATLAB Avro interface: GitHub - mathworks-ref-arch/matlab-avro: MATLAB interface for Apache Avro files.

Getting started with Java: Apache Avro™ 1.11.0 Getting Started (Java)

Avro C API: Avro C

Avro C++ API: Avro C++: Main Page
Avro C# API: Avro C#: Main Page

#### **CSV FILES**

The content of the .csv files is comma-separated, which means that, for 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.

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

Python: csv — CSV File Reading and Writing — Python 3.10.2 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