

SummerFAIR schema proposal (V0.4)

AUTHOR
Stefano Rapisarda (RDM)

PUBLISHED
July 24, 2024

1 Updates

- A "environment_events.csv" file has been added to the schema. This follows Egil's request on collecting events (in particular measurements) related to the environment;
- The "events.csv" file has been changed to "host_events.csv". This is to differentiate between "static" files ("host.csv" and "environment.csv", to be initialised only at the beginning of the experiment) and "dynamic" files, to be updated any time an intervention is performed on hosts or environments;
- In the "host_events.csv" file, the column "measurement_outcome" column has been added. This is supposed to contain the boolean outcome of a measurement, either positive(true) or negative(false);

2 .csv file columns

2.1 Hosts

	name	description	type	format
0	host_id	Host unique identifier	string	AA0_00000
1	host_groupNumber	Integer indicating the group of the host	integer	NaN
2	host_sex	Sex of the host (F/M)	string	A
3	host_age	Age of the host at the beginning of the experi...	integer	NaN
4	host_death	Date of death of the host from the beginning o...	integer	NaN
5	host_species	Species of the host	string	NaN
6	host_breed	Breed of the host within the species	string	NaN

2.2 Host Events

	name	description	type	format
0	host_id	Host unique identifier	string	AA0_00000
1	event_day	Date of the event in days after the beginning ...	integer	NaN
2	event_time	Time of the event_time (local time)	string	HH:MM
3	event_type	Type of event. It can be either measurement, i...	string	NaN
4	measurement_type	Measurement type	NaN	NaN
5	measurement_quantity	Measured quantity	float	NaN
6	measurement_outcome	Measurement outcome of a certain measurement t...	string	NaN
7	measurement_unit	Measured quantity unit	string	NaN

	name	description	type	format
8	inoculation_type	Inoculation type	string	NaN
9	inoculation_pathogen	Inoculation pathogen	string	NaN
10	inoculation_dose	Amount of inoculated pathogen	float	NaN
11	inoculation_unit	Inoculation dose unit	string	NaN
12	treatment_type	Treatment type	string	NaN
13	treatment_dose	Amount of a substance administrated in the tre...	float	NaN
14	treatment_unit	Treatment dose unit	string	NaN

2.3 Environment

	name	description	type	format
0	environment_level	Environment level	integer	NaN
1	environment_id	Environment unique identifier	string	A0_A0

2.4 Environment Events

	name	description	type	format
0	host_id	Host unique identifier	string	NaN
1	environment_id	Environment unique identifier	string	A0_A0
2	event_day	Date of the event in days after the beginning ...	integer	NaN
3	event_time	Time of the event_time (local time)	string	HH:MM
4	event_type	Type of event. It can be either measurement or...	string	NaN
5	measurement_type	Measurement type	string	NaN
6	measurement_quantity	Measured quantity	float	NaN
7	measurement_unit	Measured quantity unit	string	NaN

3 File description and usage:

When starting an experiment, researchers are supposed to create a new directory with the name of the experiment with the following structure:

```

experiment1
├─ metadata.toml
├─ _validation_schema.toml
├─ data
│   ├─ environment.csv
│   ├─ environment_events.csv
│   ├─ hosts.csv
│   └─ host_events.csv

```

3.1 Metadata description

- The **metadata.toml** file contains keywords describing general information about the experiment and, for each file, column name, description, and unit;
- The **_validation_schema.toml** file contains column names, formats, values, etc. This file is not supposed to be edited by researchers (unless they need to modify the data schema)

3.2 File description

- The **hosts.csv** file is supposed to be edited when host information is collected and when hosts die;
- The **host_events.csv** file is supposed to be edited every time an event occur. An event can be either a measurement, an inoculation, or a treatment. For each event, the operator has to specify host id, event day and time, and all the information related to that specific event, leaving empty the fields related to the other two;
- The **environment.csv** file is supposed to be edited when environment information is collected at the beginning of the experiment;
- The **environment_events.csv** file is supposed to be edited every time a host is relocated or a measurement about the environment parameters is performed;