**User Defined Analysis Simulation Data Preparation Guidance**

# Format and Structure of Simulation Data

* Since the analysis work flow is data-driven, in order to streamline the analysis, we define the format and structure of the simulation data.
* There are two parts of the simulation data: (1) meta data, and (2) sensor data.
* All data are stored in the directory of “simulation\_data/{technical\_route}/{building\_type\_or\_name}/{weather}/”, for example: “simulation\_data/general\_guidance/small\_commercial\_building/TN\_Knoxville/”.
* The users need to prepare the data themselves only when the “technical\_route” is “user\_defined\_analysis”. If “technical\_route” is “general\_guidance”, the module will provide an automatic download of simulation data, so the users do not need to prepare simulation data.

# Meta Data

* Meta data is a csv file that aggregates high-level simulation information of different fault simulation models. The file name of meta data should be the same as the field of “simulation\_metadata\_name” in the “config.json” file.
* An example of meta data is shown as below:

Graphical user interface

Description automatically generated with low confidence

* The mandate column names include: (1) “id”, (2) “fault\_type”, (3) “unmet\_hours\_during\_occupied\_cooling”, (4) “unmet\_hours\_during\_occupied\_heating”. (3) and (4) are used for sensor cost analysis.
* Each id requires one simulation data csv file in the same folder.
* One fault type called “baseline” has to be in the column of fault\_type, and its data file has to exist in the same folder. The fault type of “baseline” means the non-fault simulation case which is essential to the whole analysis workflow.

# Simulation Data

* Simulation data are basically EnergyPlus simulation results in csv format. One simulation data file represents for one fault simulation.
* An example of simulation data is shown below:

Table

Description automatically generated

* The file name of simulation data should be “{simulation\_id}\_sensors.csv”, for example, “2ce70b53-365f-4214-8625-2dbcdad712e9\_sensors.csv”. And the “simulation\_id” should exist in the “id” column of the meta data file.
* The first column should be timestamp. There is no requirement for the column name of timestamp and the time format. The time resolution should be 15 minutes. The ideal simulation time should cover one year of simulation data, making the total row of the ideal data to be 4 \* 8760 = 35040.
* For sensor cost analysis, the data should contain two columns named “electricity\_facility [W]”, “gas\_facility [W]”, which are common EnergyPlus output.