



HAL
open science

Dataset of an operating education modular building for simulation and artificial intelligence

Pierre Antoine Cormier, Quentin Laporte-Chabasse, Maël Guiraud, Julien Berton, Jean Daniel Penot, Dominique Barth

► To cite this version:

Pierre Antoine Cormier, Quentin Laporte-Chabasse, Maël Guiraud, Julien Berton, Jean Daniel Penot, et al.. Dataset of an operating education modular building for simulation and artificial intelligence. Data in Brief, 2024, 57, 10.1016/j.dib.2024.110889 . hal-04724611

HAL Id: hal-04724611

<https://hal.science/hal-04724611v1>

Submitted on 12 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License



Data Article

Dataset of an operating education modular building for simulation and artificial intelligence



Pierre-Antoine Cormier^{a,*}, Quentin Laporte-Chabasse^a,
Maël Guiraud^b, Julien Berton^b, Dominique Barth^c,
Jean-Daniel Penot^b

^a CESI LINEACT, Campus CESI Orléans, 1 allée du Titane, 45100 Orléans, France

^b CESI LINEACT, Campus CESI Nanterre, 93 boulevard de la Seine, BP602, 92006 Nanterre Cedex, France

^c DAVID Laboratory, UVSQ, Versailles, France

ARTICLE INFO

Article history:

Received 24 July 2024

Revised 20 August 2024

Accepted 26 August 2024

Available online 2 September 2024

Dataset link: [One year dataset of an education modular building \(Original data\)](#)

Keywords:

Building occupant comfort

Smart building

Indoor physical parameter

Thermal comfort

Energy consumption

ABSTRACT

Improving energy efficiency in the building sector is a subject of significant interest, considering the environmental impact of buildings. Energy efficiency involves many aspects, such as occupant comfort, system monitoring and maintenance, data treatment, instrumentation... Physical modeling and calibration, or artificial intelligence, are often employed to explore these different subjects and, thus, to limit energy consumption in buildings. Even though these techniques are well-suited, they have one thing in common, i.e., the need for user cases. This is why we propose to share a part of the large volume of data collected on our modular education building. The building is located on Nanterre's CESI Engineering school campus and welcomes approximately 80 students daily. A network of more than 150 sensors and actuators allows monitoring of the physical behavior of the entire building, preserving optimal comfort and energy consumption. The dataset includes the indoor physical parameters and the operating conditions of each system to describe the physical behavior of the building during a year.

* Corresponding author.

E-mail address: pacormier@cesi.fr (P.-A. Cormier).

Specifications Table

Subject	Control and Systems Engineering
Specific subject area	The dataset comes from a modular education building and includes two intimately linked data types: the indoor physical parameters and the operating conditions of each system.
Type of data	Table in .csv files
Data collection	Data were collected using three types of sensors: specific ones for measuring physicochemical parameters (temperature, humidity, particle concentrations..., etc.), others installed on systems (HVAC operating condition, shutter openings), or ones to observe the state of the openings (windows and doors). A Cisco network links all sensors and stores data in a relational database.
Data source location	The dataset concerns an education building (named NR3) of the CESI campus of Nanterre located at 93 Boulevard de la Seine (France). All data were then collected at the NR3 building location (Approximate GPS coordinates: 48.9033, 2.1925). The dataset is part of a database in the NR3, where all daily collected data are stored.
Data accessibility	Repository name: Zenodo Data identification number: 10.5281/zenodo.12188578 Direct URL to data: https://doi.org/10.5281/zenodo.12188578

1. Value of the Data

- The building of interest is equipped with a variety of sensors, monitoring systems, and a weather station to record operating conditions, as well as indoor and outdoor physico-chemical parameters in real time. The data are collected under real-life usage conditions over the course of a year, providing a new and comprehensive case study, to be used by the building research community.
- Building modeling is frequently employed to improve energy consumption in buildings. Modelers can use physico-chemical parameters and occupancy as a comparison to improve their building models.
- Predictive maintenance is an effective approach for minimizing maintenance costs and service disruptions. Artificial intelligence (AI) methods are commonly utilized for predicting system failures, but they require a large amount of data to train efficient AI algorithms. Our dataset could be valuable for these types of applications.
- This dataset can help to train supervised and unsupervised learning models aimed at classifying the different uses of the building.

2. Background

Our primary motivation for building this dataset is to reconcile user comfort and energy efficiency in buildings. These notions are more or less antagonistic since most energy expenditures on buildings are spent on maintaining comfortable conditions for their users. This is particularly true for thermal and air quality, depending on HVAC systems [1,2].

To reach this objective, we mainly employ indoor/outdoor physical parameters (temperature, humidity, weather..., etc.) and system operating data (CMV, heat pump, lighters..., etc.) of an experimental building called "Smart building Nanterre 3 (NR3)." This building, located on the CESI Engineering school campus of Nanterre (France), has more than 150 sensors and welcomes

Table 1

List of data stored in the vmc.csv file.

Column in the CSV file	Symbol used in the CSV file	Possible value	Description
0	Timestamp	YYYY-MM-DD hh:mm:ss	Every minute
1	T_ex (°C)	float	Extract temperature
2	T_su (°C)	float	Supply temperature
3	T_out (°C)	float	Outdoor temperature
4	P_ex (Pa)	float	Extract pressure
5	P_su (Pa)	float	Supply pressure
6	Exc_state	OFF=0 / Manual=1 / Auto=2	The operating condition of the Exchanger
7	AHU	Manual off = 0 / Manual reduced speed = 1 / Manual normal speed = 2 / Auto = 3	State of the AHU
8	F_damper	OFF=0 / Manual=1 / Auto=2	State of the fresh air damper
9	Rec_damper	OFF=0 / Manual=1 / Auto=2	State of the recirculation damper
10	Q_ex (m ³ /h)	float	Extract air flow rate.
11	Q_su (m ³ /h)	float	Supply air flow rate
12	C_mode	OFF=0 / Manual=1 / Auto=2	State of the cooling mode
13	H_mode	OFF=0 / Manual=1 / Auto=2	State of the heating mode

students to follow their courses. This work aims to provide a part of the large volume of data from our experimental building.

This dataset is a valuable resource for the academic building modeling community, data scientists, and the building sector as a whole. It offers a new real-world case study that can help test and enhance existing modeling and machine-learning approaches. Collecting various physical parameters allows for a comprehensive evaluation of the building, encompassing more than just its energy performance.

3. Data Description

Our dataset consists of 13 CSV (Comma separated Values - UTF-8 encoding) files and 2 simple Python scripts. CSV is a very convenient file format for data recording and data analysis. It does not require any query language to be manipulated, and the data is easily readable. CSV files contains four types of data: (i) operating conditions and orders sent to the building systems, (ii) indoor and outdoor physico-chemical parameters measured by specific sensors, (iii) state of the openings (windows, doors and shutters) and (iv) occupancy. Each value is collected periodically and identified by a timestamp parameter as follows: YYYY-MM-DD hh:mm:ss, DD is the day, MM the month, YYYY the year, hh the hour, mm the minute, and ss the second. The Python scripts (.zip file for the code and .log file for the results) allow identifying possible missing data in each table. In the following, we describe each file.

1. Building systems

We provide the operating conditions and orders sent to 3 main systems:

- The Air Handling Unit (AHU) controlling the heat recovery ventilation systems stored in the vmc.csv file (see on [Table 1](#)).
- The heating/cooling system consisting on a central heat pump and indoor units placed in each room stored in ac_values.csv file (see on [Table 2](#)).
- The LED lighters stored in three different files
 - o lighting_level.csv (see on [Table 3](#)) containing the lighting levels of each lighter installed in all spaces of the building.

Table 2

List of data stored in the ac_value.csv file.

Column in the CSV file	Symbol used in the CSV file	Possible value	Description
0	Timestamp	YYYY-MM-DD hh:mm:ss	Every minute
1	OnOfforder	ON=1 / OFF=0	Order sent to modify the state of the heat pump
2	OnOfvalue	ON=1 / OFF=0	State of the heat pump (read value)
3	ErrorCode	Normal=1 / Other errors=2 / Refrigeration system fault=3 / Water system error=4 / Air system error=5 / Electronic system error=6 / Sensor fault=7 / Communication error=8 / System error=9	Alert coming from the heat pump (read value)
4	Mode_order	Order: Cool=1 / Heat=2 / Fan=3 / Auto=4 / Dry=5 / Setback=6	The order was sent to the heat pump system to choose the operating mode.
5	Mode_state	Read: Cool=1 / Heat=2 / Fan=3 / Auto=4 / Dry=5 / Setback=6	Operating mode read from the heat pump system.
6	Controller_T (°C)	Read value (float)	The temperature of the room, given by the controller placed on the wall at the entry of the room
7	SetPoint_T (°C)	Read value (float)	Temperature set point of the room
8	ControllerMode_order	Order: ON=1 / OFF=0	Order sent to activate or not the room controller
9	ControllerMode_state	Read: ON=1 / OFF=0	The value of the control mode of the room controller is read from the system.
10	ProhibitionTemp	Read: ON=1 / OFF=0	Activation of the temperature control
11	Fan_state	Order: Low=1 / High=2 / Mid2=3 / Mid1=4 / Auto=5	The state of the fan of each indoor unit is sent to the heat pump
12	Fan_order	Read: Low=1 / High=2 / Mid2=3 / Mid1=4 / Auto=5	The state of the fan of each indoor unit read from the heat pump
13	UI_ROOM	TESLA / LUMIERE / NOBEL / NOBEL2 / BELL / LOCAL_TECH	Room location of the indoor unit

Table 3

List of data stored in the lighting_level.csv file.

Column in the CSV file	Symbol used in the CSV file	Possible value	Description
0	Timestamp	YYYY-MM-DD hh:mm:ss	Every minute
1	Lighting_level	An integer between 1 and 100	Light level supply by the lighter, as a rate of the maximum light (value 1 and 100)
2	Lighting_model	PowerBalance #XXXX	XXXX denotes the serial number of the lighter
3	ROOM	TESLA / LUMIERE / TURING / NOBEL / LOCAL_TECH / BELL / CIRCULATION_RDC / CIRCULATION_ETAGE1	Name of the room in which the lighter is installed

- lighting.csv (see on [Table 4](#)) containing information about the operating conditions of each lighter.
- lightingEnergy.csv (see on [Table 5](#)) containing the lighters' total energy consumption for each space (the sum of the lighters for a given space).

2. Indoor and outdoor physico-chemical parameters

Table 4

List of data stored in the lighting.csv file.

Column in the CSV file	Symbol used in the CSV file	Possible value	Description
0	Timestamp	YYYY-MM-DD hh:mm:ss	For each modification of the light level
1	Lighter_temperature (°)	Integer	Temperature measured at the lighter
2	occupancy	1 = nobody / 2 = people under the lighter	The value is given by the IR sensor located on each lighter
3	Burn_hours	Integer in hour	Accumulated time of operation of each lighter in hours from the first activation
4	Lighting_model	PowerBalance #XXXX	XXXX denotes the serial number of the lighter
5	ROOM	TESLA / LUMIERE / TURING / NOBEL / NOBEL2 / LOCAL_TECH / BELL / CIRCULATION_RDC / CIRCULATION_ETAGE1	Name of the room in which the lighter is installed

Table 5

List of data stored in the lightingEnergy.csv file.

Column in the CSV file	Symbol used in the CSV file	Possible value	Description
0	period	quarter	Period of accumulation for the energy consumption evaluation (15 min)
1	from_date	MM/DD/YYYY hh:mm:ss	Start of the acquisition period
2	to_date	MM/DD/YYYY hh:mm:ss	End of the acquisition period
3	avg_power (kW)	Float	Average power calculated with start and end energy values during the quarter of acquisition
4	min_power (kW)	Float	Minimum power during the quarter of acquisition
5	max_power (kW)	Float	Maximum power during the quarter of acquisition
6	start_acc_energy (kWh)	Float	First value of energy during the quarter of acquisition
7	end_acc_energy (kWh)	Float	The last value of energy during the quarter of acquisition
8	ROOM	TESLA / LUMIERE / TURING / NOBEL / NOBEL2 / LOCAL_TECH / BELL / CIRCULATION_RDC / CIRUYCLATION_ETAGE1	Name of the considered room

Outdoor physico-chemical parameters consists of meteorological information measured by the weather station located on the building roof. All values are collected in float type every minute except the timestamp and stored in meteo.csv file (see on [Table 6](#)).

Indoor physico-chemical parameters consist of:

- The temperature and humidity values for rooms TESLA, LUMIERE, TURING, BELL, and NOBEL2 (no sensor in NOBEL room) measured by HW-Group HTemp-1Wire sensors (see on [Table 7](#)).
- Particle concentrations, noiselevel temperature and humidity measured by Cisco Meraki MT14 sensors in each building room (see on [Table 8](#)).
- Information about the energy consumption (voltage, current, power...) of the entire building measured by a DotVision Spooky 1.5 sensor stored in energy.csv file (see on [Table 9](#)). All

Table 6

List of data stored in the meteo.csv file.

Column in the CSV file	Symbol used in the CSV file	Description	Accuracy	Resolution	Range
0	Timestamp	YYYY-MM-DD hh:mm:ss	/	/	/
1	V_wind (m s ⁻¹)	Wind speed	0.5	1/16	0–70
2	Dir_wind (°)	Wind direction	5	1/16	0–360
3	P (hPa)	Atmospheric pressure	1	3	300–1100
4	RH (%)	Relative humidity	4	3	0–100
5	T_dew (°C)	Calculated dew point	/	/	/
6	Pyra_N (W/m ²)	Solar radiation incident to the north wall	10	1/16	0–1400
7	Pyra_W (W/m ²)	Solar radiation incident to the west wall			
8	Pyra_E (W/m ²)	Solar radiation incident to the east wall			
9	Pyra_S (W/m ²)	Solar radiation incident to the south wall			
10	I (lux)	External light intensity	10	1/16	5–150,000
11	T_air (°C)	Ambient air temperature	1	1/16	–40–60

Table 7

List of data stored in the temp.csv file.

Column in the CSV file	Symbol used in the CSV file	Possible value	Description	Accuracy
0	Timestamp	YYYY-MM-DD hh:mm:ss	Every minute	/
1	T_room (°C)	float	Temperature	0.8 °C
2	RH_room (%)	float	Relative humidity	5 %
3	ROOM	TESLA / NOBEL2 / TURING / LUMIERE / BELL	Room location	/

Table 8

List of data stored in the sensor_meraki.csv file.

Column in the CSV file	Symbol used in the CSV file	Possible value	Description
0	Timestamp	YYYY-MM-DD hh:mm:ss	Every minute
1	pm25 (µg/m3)	Float	The concentration of microparticles with a diameter less than 2.5 µm (PM2.5)
2	tvoc (µg/m3)	Float	Total concentration of volatile organic compounds (TVOC)
3	noise (dBA)	Float	Ambient noise
4	T_air (°C)	Float	Air temperature
5	RH (%)	Float	Relative humidity
6	ROOM	TESLA / LUMIERE / NOBEL / NOBEL2 / TURING	Name of the sensor associated to the room location name.

values are in float type (except timestamp) and collected every minute with an accuracy of 1 %.

3. Openings

We provide in the dataset the status (open/close) of the different openings such as doors, windows and shutters, and for some of them information about the sensor. Each measured data is associated to the position of the opening and stored in door.csv (see on Table 10), window.csv (see on Table 1 and Table 11) and shutters .csv (see on Table 12) files, respectively.

Table 9

List of stored in the energy.csv file.

Column in the CSV file	Symbol used in the CSV file	Description	Units
0	Timestamp	YYYY-MM-DD hh:mm:ss	/
1	WATTHRb	The active energy of phase B	Wh
2	WATTB	Active power of phase B	W
3	VARB	Reactive power of phase B	VA
4	VAHRb	Apparent energy of phase B	VAh
5	VAB	The apparent power of Phase B	VA
6	VRMSB	RMS voltage of phase B	V
7	VARHRb	Reactive energy of phase B	VAh
8	PFB	Power factor of phase B	/
9	IRMSB	RMS current of phase B	A
10	WATTHRc	The active energy of phase C	Wh
11	WATTc	Active power of phase C	W
12	VARc	Reactive power of phase C	VA
13	VAHRc	Apparent energy of phase C	VAh
14	VAc	The apparent power of Phase C	VA
15	VRMSc	RMS voltage of phase C	V
16	VARHRc	Reactive energy of phase C	VAh
17	PFC	Power factor of phase C	/
18	IRMSc	RMS current of phase C	A
19	WATTTh	Total Active energy	Wh
20	WATT	Total Active power	W
21	VAR	Total Reactive power	VA
22	VAHR	Total Apparent energy	VAh
23	VA	Total Apparent power	VA
24	TEMP	Sensor temperature	°C
25	VARHR	Total Reactive energy	VAh
26	FREQ	Line frequency	Hz
27	IRMS	Total RMS current	A
28	WATTThRA	The active energy of phase A	Wh
29	WATTa	Active power of phase A	W
30	VARa	Reactive power of phase A	VA
31	VAHRa	Apparent energy of phase A	VAh
32	VAA	The apparent power of Phase A	VA
33	VRMSa	RMS voltage of phase A	V
34	VARHRa	Reactive energy of phase A	VAh
35	PFA	Power factor of phase A	/
36	IRMSa	RMS current of phase A	A

Table 10

List of data stored in the door.csv file.

Column in the CSV file	Symbol used in the CSV file	Possible value	Description
0	Timestamp	YYYY-MM-DD hh:mm:ss	For each event (open or closed)
1	State	Close = 0 / 1 = open	State of the door
2	Battery	Integer in%	Remaining capacity of the sensor battery in%
3	Device_T	Integer in °C	Temperature of the sensor
4	Position	CIRCULATION_RDC / TESLA / LUMIERE / LOCAL_TECH / BELL / NOBEL / NOBEL2 / TURING	The location of the door is linked to the room name.

4. Occupancy

The occupancy.csv file contains information about the theoretical occupation of each room. The theoretical occupancy was adjusted to the room's maximum capacity (18 people) when it exceeded this maximum capacity. All those treatments were recorded in the Post_processing column with the value 1 (0, meaning no post-treatment was performed). This situation is fre-

Table 11

List of data stored in the window.csv file.

Column in the CSV file	Symbol used in the CSV file	Possible value	Description
0	Timestamp	YYYY-MM-DD hh:mm:ss	Every minute
1	State	Close = 0 / 1 = open	State of the door
2	Position	NOBEL_F1 / NOBEL_F2 / NOBEL_F3 / HALL_PALIER_F1 / HALL_PALIER_F2 / HALL_PALIER_F3 / TURING_F1 / TURING_F2 / TURING_F2 / LUMIERE_F1 / LUMIERE_F2 / LUMIERE_F3 / TESLA_F1 / TESLA_F2 / TESLA_F3 / TESLA_F4 / HALL_RDC_F1	Location of the window in the room
3	ROOM	CIRCULATION_RDC / CIRCULATION_ETAGE1 / TESLA / NOBEL / TURING / LUMIERE	Room location of the window

Table 12

List of data stored in the shutter_orders.csv file.

Column in the CSV file	Symbol used in the CSV file	Possible value	Description
0	Timestamp	YYYY-MM-DD hh:mm:ss	For each modification of the opening
1	Position	NOBEL_GAUCHE / NOBEL_CENTRE / NOBEL_DROITE / 1ER_ETAGE_GAUCHE / 1ER_ETAGE_CENTRE / 1ER_ETAGE_DROITE / TURING_GAUCHE / TURING_CENTRE / TURING_DROITE / LUMIERE_DROITE / LUMIERE_GAUCHE / LUMIERE_CENTRE / TESLA_DROITE / TESLA_GAUCHE / TESLA_CENTRE1 / TESLA_CENTRE2 / RDC_CENTRE	Location of the window in the room
2	Shutter_position	Integer between 0 and 100	The opening rate of the shutter

quently met when a large class is spread over several rooms. We also inferred the “type of use” from the reservation topics in our planning and room management system. In the dataset, all topics mentioning a meeting in their name were labeled MEETING. Similarly, all topics referring to defense were labeled DEFENSE. Finally, the CLASS label gathers all other topics relating to class activities, such as courses, tutorials, practical work, and exams in various subjects. [Table 13](#)

5. Additional Python code

Example_code_midding_data.zip contains two Python code examples to analyze the missing data points depending on the recording time interval. In “all_missing_values.py” code, the tables for door, ac_values, temp, and sensor_meraki are processed as follows:

- I. The number of minutes without data entries is counted and recorded for each room.
- II. The data is then grouped by month.
- III. Information for the year and by month is stored in the “print.log” file, and a file named “XXX_year.png” (with XXX being the table name) is created to display the number of missing data points by month and by room.

In the file “Month_missing_values.ipynb”, the data is first filtered for a specific month to display daily details of the missing data points on a graph.

Table 13

List of data stored in the occupancy.csv table.

Column in the CSV file	Symbol used in the CSV file	Possible value	Description
0	ROOM	TESLA / LUMIERE / NOBEL / TURING	Name of the room
1	Type_of_use	CLASS / DEFENSE / MEETING	Purpose of the session: <ul style="list-style-type: none"> - CLASS: planned class activities including courses, tutorials, exams ... - DEFENSE: planned defenses - MEETING: planned meetings
2	Start_session	YYYY-MM-DD hh:mm:ss	Start of the session
3	End_session	YYYY-MM-DD hh:mm:ss	End of the session
4	People_number	integer	Theoretical number of people exposed in the room
5	Post_processed	0 = no post-processing / 1 = post-processing was performed on column People_number	Indicates whether a post-treatment ensuring that $people_number \leq max$ Capacity has been performed on People_number.

4. Experimental Design, Materials and Methods

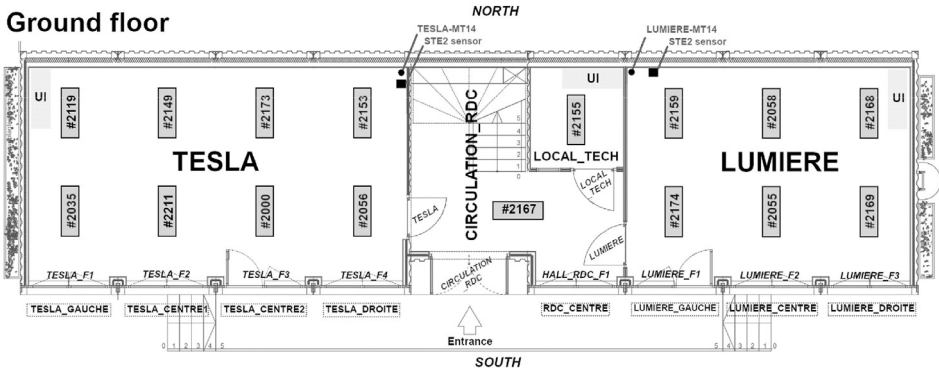
Data composing the dataset were acquired from the 1st of June 2023 to the 31 of May 2024 and concerns an education building (NR3) of CESI Engineering school located on the campus of Nanterre (France - Approximate GPS coordinates: 48.9033, 2.1925). This building comprises 16 recycled shipping containers offering an available surface of ≈ 220 m² on two levels. It integrates four classrooms named TURING (38 m²), TESLA (51 m²), NOBEL (38 m²), and LUMIERE (38 m²); a server room named BELL (8.7 m²); an office named NOBEL2 (8.7 m²); a technical room named LOCAL_TECH and a hall distributed on the two levels (CIRCULATION_RDC and CIRCULATION_ETAGE for ground and first floors, respectively). Fig. 1 permits to visualize the organization of different spaces, as well as the position of the different systems and sensors. Nevertheless, the reader can refer to Doukhari et al. [3] for a detailed description of the building.

NR3 is equipped with ≈ 150 sensors and actuators, allowing the measurements of physical environmental parameters, the operating conditions of each system, and the monitoring of the latter. In the following, we describe each system and sensor by starting with the computer architecture to collect and store related data.

1. Computer architecture and database

The building relies on a hypervisor cluster based on VMware vSphere 8 technologies for monitoring, data collection, and storage. The production cluster consists of 2 servers on which HA (High Availability ensuring an automatic restart in case of failure), DRS (Distributed Resource Scheduler – automatic distribution of the load between each virtual machine), and EVC (Enhanced vMotion Compatibility ensuring the CPU compatibility for virtual machine migration) are enabled. A Synology FS2500 storage array centralizes the virtual machines on a LUN (Logical Unit Number) exposed via the iSCSI (Internet Small Computer Systems Interface) protocol. The hypervisors are directly connected to this array through 10 Gbit/s interfaces to maximize operational performance. The sensors communicate with the primary virtual machine collecting the building data through further protocols (Modbus, BACnet, HTTP, HTTPS/REST, ZigBee / MQTT, and Bluetooth). Despite the wide variety of protocols, the building collects and stores data in a unified manner within a single MariaDB database. This consistency allows quick and efficient data access for statistical or control purposes.

Ground floor



First floor

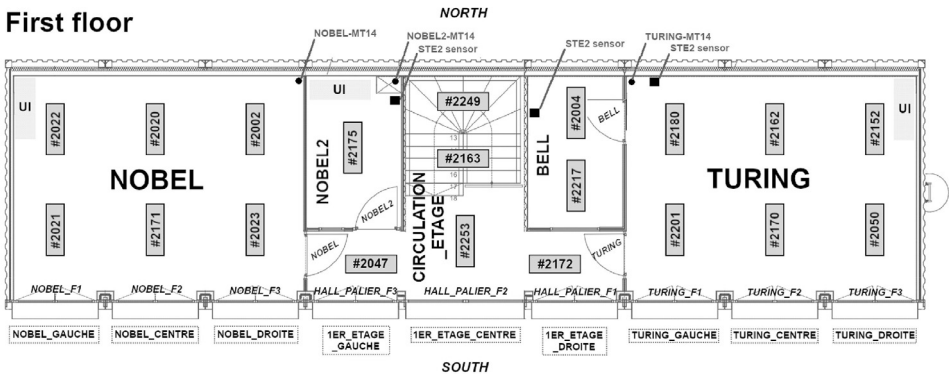


Fig. 1. Plan of the two levels of the NR3 building indicating the position of each sensor and system related to the dataset. #XXX indicates the LED panels. The names of the windows, shutters, and doors are written in italics, dotted boxes, and italics and tilted. UI means Indoor units used for heating and cooling each room.

2. Building systems

As mentioned in the DATA DESCRIPTION, we provide data coming from three main systems, i.e. the AHU, the heating/cooling and the lighting systems.

The AHU consists of a heat recovery ventilation system with a rotary thermal wheel (VIM Integral E 25 VLEXD CORRIGO)—the maximum blowing and recovery flow rate equal to 2350 and 2200 m³/h at 2500 Pa, respectively. The thermal efficiency of the exchanger is $\approx 80\%$. Fresh air is filtered by coupling a G4 (particles higher than 10 μm in diameter) and an F7 (ePM₇ according to ISO16890) filter. A M5 filter (ePM₁₀ according to ISO16890) is placed at the output of the AHU. A CORRIGO controller allows the control of the AHU and the collection of all operating data (vmc.csv file) using the Modbus protocol.

The heating/and cooling system consist of a central heat pump (Mitsubishi PUMY-P140YKM2) located on the building roof and linked to individual indoor units (IU – Mitsubishi PFFY-P20VLRMM-E) located in each room. A controller in each room offers the possibility of controlling the room temperature. The entire system can be monitored by BACnet protocol, and several data can be collected (ac_value.csv file).

The lighting of the building is ensured by 37 lighters (LED panels Philips PowerBalance with sensor – RC461B Gen1) distributed as follows: (i) 8 in Tesla room, (ii) 6 in Bell, Turing and Nobel rooms, (iii) 1 in Local_tech and Nobel2 rooms, and (iv) 6 in the hall. These lighters allow you to vary the lighting level and are equipped with temperature and IR motion sensors. They are supplied in PoE and use IPv6 and DyNet protocol to communicate with the database

(lighting_level.csv, lighting.csv and lghtingEnergy.csv files). The position and the associated serial number of each panel are shown in Fig. 1.

3. Sensors dedicated to the measurement of indoor/outdoor physico-chemical parameters

A weather station (IRDAM WST6000GTB), located on the roof of the NR3 building, measures the meteorological parameters (meteo.csv file). In addition to the information in the table, four pyranometers directly measure the solar radiation incident to each building wall. Each pyranometer has a spectral and an angular response ranging from 300 to 1100 nm and of $\pm 90^\circ$, respectively. Moreover, the lux meter has a spectral response similar to the human eye and an angular response of $\pm 90^\circ$. The station's communication protocol is ModBus.

TESLA, TURING, LUMIERE, and NOBEL2 rooms are equipped with STE2 (HW-Group) 1-wire sensors measuring the room's temperature ($^\circ\text{C}$) and relative humidity (%) – (temp.csv file). The sensors use IPv4 and HTTP&XML protocols and are fixed on the wall at approximately 1.7 m in height.

An air quality sensor, the Cisco Meraki MT14, measures different physical comfort parameters in TESLA, TURING, NOBEL, and LUMIERE rooms (sensor_meraki.csv). The sensors are positioned on the wall in a corner of the room at a height of approximately 2.2 m. All parameters are listed in Table 12. These sensors are connected to the system by Bluetooth.

An energy sensor (DotVision Energy Spoony 1.5) measures the electric consumption of the entire building (energy.csv file). The sensor's core is an ADE7758 chipset that measures the instantaneous voltage and current to calculate the active energy. This means that the electric consumption in a given period is calculated by dividing the values measured at the end and the beginning. Moreover, the sensor has three individual (A, B, and C) channels to measure each phase of a three-phase power supply or three single-phase supplied systems. Data are collected by HTTP/JSON protocol (see Table 9) and each value is given with an accuracy of 1 %.

4. Openings

All doors are equipped with Hall Effect sensors (Aqara MCGQ11LM) measuring the state of each of them, i.e., open (=1) or close (=0). The sensor also returns the level of its internal battery in% and its temperature in $^\circ\text{C}$ using a ZigBee/MQTT protocol (door.csv file). Windows are also equipped with an open/close sensor that returns 1 or 0 when the window is open or closed, respectively, using a Modbus protocol (window.csv file). The location of each window is shown in Fig. 1.

Each window is equipped with shutters (see Fig. 1) that can be controlled by the system (ZigBee/MQTT protocol) or by users with a Schneider Electric (S520567W) actuator. The sensor returns the opening rate between 0 and 100 % (shutter_orders.csv file).

5. Occupancy

The occupation of the building is not measured in real-time, but we provide theoretical value based on the school's booking calendar (occupancy.csv file). This occupation calendar does not consider some events, such as late or absent students, course cancellations, or student exits during a course...

Limitations

- The room Nobel does not dispose of an STE2 sensor.
- The lighting level data of the Tesla room may be completely uncorrelated with the data from other rooms due to an experiment that has been in progress since March 2023.
- For each table containing minute-by-minute data (vmc, ac_value, temp, window, lighting_level, sensor_meraki), there are very few missing data points when the building was operating normally (only a few packet losses). However, we observe notable failures in the sensors and data collection scripts on the following tables/dates:

- Table ac_values: data gap between November 6, 2023, and November 16, 2023.
- Table temp: A few data were missing in September 2023. Also, the Nobel room has no data in this table.
- Table sensor_meraki: There are a few missing data in September and July 2023 (less than 10 h) and in February and May 2024 (less than 12 h).

Ethics Statement

All authors have read and followed the ethical requirements for publication in Data in Brief and confirm that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

CRediT Author Statement

Pierre-Antoine Cormier: Writing - Original draft, Writing - Review & Editing, Supervision; **Maël Guiraud:** Software, Formal analysis; **Quentin Laporte-Chabasse:** Software, Formal analysis; **Julien Berton:** Resources; **Dominique Barth:** Project administration; **Jean-Daniel Penot:** Project administration.

Data Availability

[One year dataset of an education modular building \(Original data\)](#) (Zenodo).

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] D. Mariano-Hernández, L. Hernández-Callejo, A. Zorita-Lamadrid, O. Duque-Pérez, F.S. García, A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis, *J. Build. Eng.* 33 (2021) 101692.
- [2] P. Jafarpur, U. Berardi, Effects of climate changes on building energy demand and thermal comfort in Canadian office buildings adopting different temperature setpoints, *J. Build. Eng.* 42 (2021) 102725.
- [3] O. Doukari, B. Seck, D. Greenwood, H. Feng, M. Kassem, Towards an interoperable approach for modelling and managing smart building data: The case of the CESI smart building demonstrator, *Buildings* 12 (3) (2022) 362.