PHOENIX

WP1 - Project Management and Coordination

Document Version:

D1.4 Data Management Plan

4.0

Project Number:	**Project Acronym:**	**Project Title:**
893079	PHOENIX	Adapt-&-**P**lay **H**olistic c**O**st Effective and user-frie**N**dly **I**nnovations with high replicability to upgrade smartness of e**X**isting buildings with legacy equipment

Contractual Delivery Date:	**Actual Delivery Date:**	**Deliverable Type* - Security**:**
28/02/2020	28/02/2020	R - PU

* Type: P - Prototype, R - Report, D - Demonstrator, O - Other

** Security Class: PU- Public, PP - Restricted to other programme participants (including the Commission), RE - Restricted to a group defined by the consortium (including the Commission), CO - Confidential, only for members of the consortium (including the Commission)

Responsible and Editor/Author:	Organization:	Contributing WP:
Antonio Skarmeta	UMU	WP1

Authors (organizations):

Antonio Skarmeta (UMU)

Abstract:

This document describes the strategy of project PHONIX with respect to the Data Management Plan. The document gives a description of the main data that will be collected on the pilots, and it describes the strategies to follow a Fair data philosophy. More details about how the data will be handled in terms of equipment and security aspects are also given in this document.

Keywords:

Data, Open Access, Management, Ethics, Security

Revision History

The following table describes the main changes done in the document since created.

Revision	Date	Description	Author (Organization)
1.0	02/02/2020	First version of the document	Alfonso Ramallo (UMU)
2.0	08/02/2021	Internal revision	Antonio Skarmeta (UMU)
3.0	24/02/2021	Review 1	Eirini Christoulaki (KaMa)
3.1	25/02/2021	Review 2	Aristotelis Ntafalias, Panos Papadopoulos (VERD)
4.0	26/02/2021	Final revision	Alfonso Ramallo, Antonio Skarmeta (UMU)

Executive Summary

This document describes the strategy of project PHONIX with respect to the Data Management Plan. The document gives a description of the main data that will be collected on the pilots, and it describes the strategies to follow a Fair data philosophy. More details about how the data will be handled in terms of equipment and security aspects are also given in this document.

The document shows how UMU pilot and ARDEN Pilot have a great volume of data, but the most interesting point about the summary of the data that will be available in this project is how heterogeneous it is.

This document also presents the strategies for FAIR data. In the current situation of research, it is the norm to facilitate the access to data from research and innovation to increase the impact of the projects. In the same way, the availability of the data allows in many cases to replicate the studies, making possible the verification of results, or the extrapolation of them with another data that can complete the studies. This document shows how the data in this investigations will be made public when possible, and also that it is considered to allocate resources to make open-access the scientific publications that are produced within PHOENIX. When gold open access is not possible then green open access will be targeted.

This deliverable also outlines concepts on data security and ethical aspects, although with respect to this last issue Deliverables 9.1 and 9.2 should be the main text.

Disclaimer

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 893079, but this document only reflects the consortium's view. The European Commission is not responsible for any use that may be made of the information it contains.

Table of Contents

Table of Contents *5*

1 Data Summary *6*

1.1 UMU **8**

1.2 ODINS **9**

1.3 MIWenergía **12**

1.4 KaMa **13**

1.5 Suite5 ..... **14**

1.6 ARDEN **14**

1.6.1 Pilot Site #1 Rediscovery Centre 14

1.6.2 Other Pilot Sites 19

1.6.3 Common Data 19

1.7 LTU Demo Site **20**

2 Fair data *21*

2.1 Making data findable, including provisions for metadata **21**

2.2 Making data openly accessible **23**

2.3 Making data interoperable **24**

2.4 Increase data re-use (through clarifying licences) **24**

3 Allocation of resources *25*

4 Data Security *26*

5 Ethical aspects *27*

6 Other *27*

Appendix Data attributes for UMU Pilot *28*

1 Data Summary

This project aims at demonstrating the possibility of upgrading existing buildings, so that they become smarter. Part of that *smartisation* is the capability of devices to send information (data) over the internet. This implies that the data plays an important role in this project. The data used in the project will be collected from devices (electronic) such as sensors, or appliances with the capability of sending information over the internet.

An important part of the project is to develop services that will help building users, occupants and also services providers. It is important for this reason that the data reflects occupants' uses and patterns, and so will be done within PHOENIX project.

Most data in this project will come from electronic devices. However, there will be surveys and other data investigation methods (such as web analytics) that will serve on the purpose of this task. All these data will be securely stored on local servers protected under pseudo-anonymization and password protection.

The data collected in this project will be used to test the services developed. However, the project has been structured in work packages that maximise its success and make the best use of resources. This implies that there will be efforts on developing the intelligence and the services before data from the upgrades is collected. To make sure that the progress in these fronts is done, it is likely that already published data from other groups will be used, as well as historical data collected from the partners of the consortium prior to this project.

With respect to the volume of the data, the project has five pilots with one or more buildings in each one. Each building will have a variety of devices that will collect information in different forms and of different volumes. For this, it is expected that the volume of project's data will be rather high. To cope with this volume without compromising security, local resources (servers), with available capacity enough to deal with the data, have been allocated.

There are three main groups of data that will be generated and used within PHOENIX. These are:

- **Survey data**: Including information about demographics, intentions, norms and believes among others.

- **Device data**: These data will come from the pilots and will be, reported by the devices into the platform of PHOENIX. These data include streams of physical variables, behavioural patterns and other observations, and they constitute the main input for the analytic/monitoring layers of the PHOENIX solution.

- **Analytical data:** These data will be produced within PHOENIX thanks to the algorithms and processes that will be developed in the project in order to achieve the expected functionalities.

These include monitoring of buildings, reports and alarms of building operation, messages and alerts from grid interaction among others.

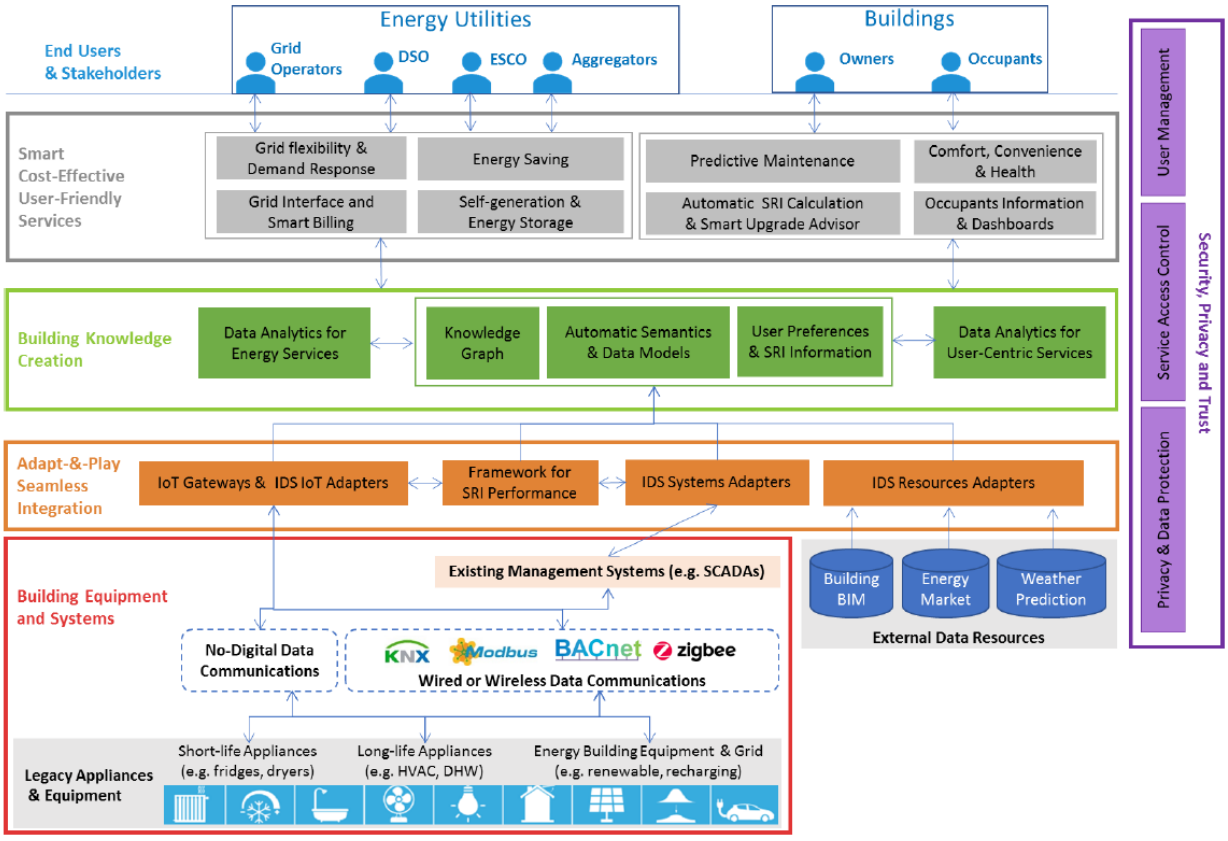

Figure 1. PHOENIX Architecture.

The architecture of PHOENIX in its current version and the components included help identifying a set of data formats that will be present in the project. An example of the expected payback, given by the request under the RESTful (Automated Programmable Interface) API can be seen here for a stream of a physical quantity:

Component: IoT-Broker based on FIWARE architecture

Type/nature of data produced: Sensors/Metering

Format/Protocol/Standard: REST protocol

Structure/Model: JSON format

Potential reuse: Sensor produced data could be used in R&D projects of smart building applications based on energy efficiency, renewable energy harvesting, security, access control, inside-mobility, etc.

Estimated size: The amount of data produced within the project will be very high, because in the buildings there are hundreds of devices sending sensor data every 10 minutes.

Example of JSON format for temperature sensor:

```
{
    "name": "Temperature",
    "type": "Float",
    "value": "29.2"
}
```

Purpose of Data: The purpose of data collation in buildings is to validate PHOENIX framework and their components in real scenarios with real-time data according to the use cases proposed in the project.

Potential reuse: Sensors' data can be useful for researchers and developers of new solutions (e.g. applications and services) oriented towards improving the use of buildings and its equipment.

More details regarding the data that will be used at each one of the pilots can be seen in the following tables.

1.1 UMU

sensor_type	building_area
energy_power_electricity	chemistry, pleiades
electricity	pleiades
hvac	chemistry, pleiades
ambient_temperature	pleiades
ambient_luminosity	pleiades
ambient_humidity	pleiades
ambient_co2	pleiades
ambient_presence	pleiades
external_temperature_weather_station	pleiades
wind_speed_weather_station	pleiades

wind_direction_weather_station	pleiades
solar_radiation_weather_station	pleiades
weather_obs	chemistry
weather_pred	chemistry

With respect to building data:

Building ID	chemistry, pleiades
Construction year	chemistry, pleiades
Building type	chemistry, pleiades
Building size	chemistry, pleiades
Windows to wall ratio	chemistry, pleiades
Building regulations	chemistry, pleiades
Consumption baseline	chemistry, pleiades
Sensor ID (link with sensor data)	chemistry, pleiades
Total number of sensors	chemistry, pleiades
Internal average temperature	chemistry, pleiades
Internal average humidity	chemistry, pleiades
Occupants per room/building	chemistry, pleiades

The attributes of the data steams can be seen in the Appendix.

1.2 ODINS

Weather data will be gathering from the online service of weatherbit.io site.

Data type and description

`lat`: Latitude (Degrees)
`lon`: Longitude (Degrees) `sunrise`: Sunrise time (HH:MM)
`sunset`: Sunset time (HH:MM) `timezone`: Local IANA Timezone
`station`: Source station ID `ob_time`: Last observation time (YYYY-MM-DD HH:MM)
`datetime`: Current cycle hour (YYYY-MM-DD:HH) `ts`: Last observation time (Unix timestamp)
`city_name`: City name `country_code`: Country abbreviation
`state_code`: State abbreviation/code `pres`: Pressure (mb)
`slp`: Sea level pressure (mb) `wind_spd`: Wind speed (Default m/s) `wind_dir`: Wind direction (degrees)
`wind_cdir`: Abbreviated wind direction
`wind_cdir_full`: Verbal wind direction `temp`: Temperature (default Celcius) `app_temp`: Apparent/"Feels Like" temperature (default Celcius)
`rh`: Relative humidity (%)
`dewpt`: Dew point (default Celcius) `clouds`: Cloud coverage (%)

`pod`: Part of the day (d = day / n = night) `vis`: Visibility (default KM)
`precip`: Liquid equivalent precipitation rate (default mm/hr)
`snow`: Snowfall (default mm/hr)
`uv`: UV Index (0-11+)
`aqi`: Air Quality Index [US - EPA standard 0 - +500] `dhi`: Diffuse horizontal solar irradiance (W/m^2) [Clear Sky]
`dni`: Direct normal solar irradiance (W/m^2) [Clear Sky]
`ghi`: Global horizontal solar irradiance (W/m^2) [Clear Sky]
`solar_rad`: Estimated Solar Radiation (W/m^2) `elev_angle`: Solar elevation angle (degrees)
`h_angle`: Solar hour angle (degrees)

The next example shown the JSON format of the data gathered from weatherbit.io service:

Example Response (JSON):


```
{
   "data":[
      {
         "wind_cdir":"NE",
         "rh":59,
         "pod":"d",
         "lon":"-78.63861",
         "pres":1006.6,
         "timezone":"America\/New_York",
         "ob_time":"2017-08-28 16:45",
         "country_code":"US",
         "clouds":75,
         "vis":10,
         "wind_spd":6.17,
         "wind_cdir_full":"northeast",
         "app_temp":24.25,
         "state_code":"NC",
         "ts":1503936000,
         "h_angle":0,
         "dewpt":15.65,
         "uv":2,
```

```
                "aqi":45,
                "station":"CMVN7",
                "wind_dir":50,
                "elev_angle":63,
                "datetime":"2017-08-28:17",
                "precip":0,
                "ghi":444.4,
                "dni":500,
                "dhi":120,
                "solar_rad":350,
                "city_name":"Raleigh",
                "sunrise":"10:44",
                "sunset":"23:47",
                "temp":24.19,
                "lat":"35.7721",
                "slp":1022.2
            }
        ],
        "count":1
    }
```


1.3 MIWenergía

MIWenergía has two different scenarios: residential and office building. Residential data consumption is available from the (Distribution Systems Operator) DSO smart meter, which we have access because the project users are clients of the company. Regarding the office building, it is not our client thus there is no accessible data, also the smart meter covers the consumption for the whole building but not for each office.

The data will be gathering:

Energy Meters	Parameters
Smart Meter User 1	kWh
Smart Meter User 2	kWh
Smart Meter User 3	kWh
Smart Meter User 4	kWh

The structure for residential data will be as:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	CCHId	CCHDatoCUPS	CupsId	TimeStamp	CCHDatoFechaHora	CCHDat	CCHDatoTipoMed	CCHDatoMedidaEntrante	CCHDatoCalidadE	CCHDatoMedidaS	CCHDatoCalidadS	CCHDatoReactivaC1	CCHDato
2	31875	ES00xxxxxxxxxxxx1EM	35137	60553	28/11/2020 1:00:00	0	null	397	null	0	null	270	null
3	31875	ES00xxxxxxxxxxxx1EM	35137	60554	28/11/2020 2:00:00	0	null	214	null	0	null	129	null
4	31875	ES00xxxxxxxxxxxx1EM	35137	60555	28/11/2020 3:00:00	0	null	251	null	0	null	154	null
5	31875	ES00xxxxxxxxxxxx1EM	35137	60556	28/11/2020 4:00:00	0	null	266	null	0	null	162	null
6	31875	ES00xxxxxxxxxxxx1EM	35137	60557	28/11/2020 5:00:00	0	null	232	null	0	null	132	null
7	31875	ES00xxxxxxxxxxxx1EM	35137	60558	28/11/2020 6:00:00	0	null	186	null	0	null	119	null
8	31875	ES00xxxxxxxxxxxx1EM	35137	60559	28/11/2020 7:00:00	0	null	185	null	0	null	117	null
9	31875	ES00xxxxxxxxxxxx1EM	35137	60560	28/11/2020 8:00:00	0	null	287	null	0	null	132	null
10	31875	ES00xxxxxxxxxxxx1EM	35137	60561	28/11/2020 9:00:00	0	null	295	null	0	null	141	null
11	31875	ES00xxxxxxxxxxxx1EM	35137	60562	28/11/2020 10:00:00	0	null	297	null	0	null	138	null
12	31875	ES00xxxxxxxxxxxx1EM	35137	60563	28/11/2020 11:00:00	0	null	633	null	0	null	190	null
13	31875	ES00xxxxxxxxxxxx1EM	35137	60564	28/11/2020 12:00:00	0	null	1235	null	0	null	137	null
14	31875	ES00xxxxxxxxxxxx1EM	35137	60565	28/11/2020 13:00:00	0	null	840	null	0	null	143	null
15	31875	ES00xxxxxxxxxxxx1EM	35137	60566	28/11/2020 14:00:00	0	null	302	null	0	null	116	null
16	31875	ES0021000006230501EM	35137	60567	28/11/2020 15:00:00	0	null	1400	null	0	null	165	null

Figure 2. Example of data table in MIWENERGIA Pilot.

Different kind of sensors will be installed to upgrade the intelligence of the legacy equipment but those will be decided in future stages of the project. Yet, it is planned that sensors will be deployed in order to get the following data: smart meters for energy consumption at building and office level (kWh), ambient sensors for temperature, humidity and luminance (ºC, %, lux), occupancy, HVAC control devices, smart plugs (kWh, on/off), etc.

1.4 KaMa

The main data log for the Pilot-Case-KaMa (residential building) is devoted to electrical energy consumption in each apartment. Historical data is extracted based on calculations after taking into account consumers' habits via questionnaires (available for comparison to track Pilot's improvement during the PHOENIX project). Sensor types and related parameters measured for each flat are described in the following tables. Additionally, electrical consumption and weather conditions are measured to describe the general building's conditions.

Each Apartment (No1 - No8)

Sensor_Type	Parameter/ Usage
Smart_meter_1	General consumption
Smart_meter_2	Washing machine consumption
Smart_meter_3	Refrigerator consumption
Smart_differential_Thermostat	Solar thermal collectors
Smart_Thermostat	Heating/Cooling energy from HVAC
Luminance_Meter	Ambient/ Room luminance

General Building

Sensor_Type	Parameter/ Usage
Smart_meter_1	General building consumption
Smart_meter_2	EV Charger consumption
Smart_meter_3	PV Energy generation
Weather Station	Air temperature; Luminance; Relative humidity; Wind speed; Barometric pressure, Dew point

1.5 Suite5

Suite5 is leading the work in WP2 towards the definition of project foundations. Questionnaires were circulated in an anonymous way (through EU Survey Platform) and the feedback (aggregate statistics) will be available in the public deliverables "D2.1 Business, market & regulatory requirements" and "D2.2 Social barriers and enablers, building use"

On the other hand, Suite5 is responsible for the delivery of non-energy services in WP5. The results of the algorithms and processes occurring in the project to achieve the expected functionalities for the provision non-energy services will be defined in the following months. In any case, and following the data management principles defined in the project, no personalized information will be made available and only aggregate/anonymized results will be made available.

1.6 ARDEN

1.6.1 Pilot Site #1 Rediscovery Centre

The Rediscovery Centre is controlled by a Delta Controls Building Management System (BMS) system with an Entelliweb front end visualisation platform. The following tables list the BMS sensors and inputs for environmental and temperature sensors and for energy maters (gas, heat and electricity).

Environmental and Temperature Sensors	
sensor_type	**Parameter**
CHP_Flow Temp	Temperature 10K3A1 -40-150 °C AIC
CHP_Retn Temp	Temperature 10K3A1 -40-150 °C AIC

Boiler_Flow Temp	Temperature 10K3A1 -40-150 °C AIC
Boiler_Retn Temp	Temperature 10K3A1 -40-150 °C AIC
HeatPump_Flow Temp	Temperature 10K3A1 -40-150 °C AIC
HeatPump_Retn Temp	Temperature 10K3A1 -40-150 °C AIC
Stove_Flow Temp	Temperature 10K3A1 -40-150 °C AIC
Stove_Retn Temp	Temperature 10K3A1 -40-150 °C AIC
Buffer_Tank Temp 1	Temperature 10K3A1 -40-150 °C AIC
Buffer_Tank Temp 2	Temperature 10K3A1 -40-150 °C AIC
Buffer_Flow Temp	Temperature 10K3A1 -40-150 °C AIC
Buffer_Retn Temp	Temperature 10K3A1 -40-150 °C AIC
LPHW Header_Flow Temp	Temperature 10K3A1 -40-150 °C AIC
LPHW Header_Retn Temp	Temperature 10K3A1 -40-150 °C AIC
Ground_Rad Flow Temp	Temperature 10K3A1 -40-150 °C AIC
Ground_Rad Retn Temp	Temperature 10K3A1 -40-150 °C AIC
First_Rad Flow Temp	Temperature 10K3A1 -40-150 °C AIC
First_Rads Retn Temp	Temperature 10K3A1 -40-150 °C AIC
Outside Temp	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_Gnd 1	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_Gnd 2	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_Gnd 3	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_Gnd 4	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_Gnd 5	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_Gnd 6	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_Gnd 7	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_Gnd 8	Temperature 10K3A1 -40-150 °C AIC
Garden_Temp_Gnd 9	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_Gnd 10	Temperature 10K3A1 -40-150 °C AIC
Room_Co2 1	Co2 0-10V 0 to 2000ppm AIC
Room_Co2 2	Co2 0-10V 0 to 2000ppm AIC
Room_Co2 3	Co2 0-10V 0 to 2000ppm AIC
Room_Co2 4	Co2 0-10V 0 to 2000ppm AIC
Room_Co2 5	Co2 0-10V 0 to 2000ppm AIC

Room_TVOC 1	TVOC 0-10V 0 to 100 AIC
Room_TVOC 2	TVOC 0-10V 0 to 100 AIC
Room_TVOC 3	TVOC 0-10V 0 to 100 AIC
Room_TVOC 4	TVOC 0-10V 0 to 100 AIC
Room_TVOC 5	TVOC 0-10V 0 to 100 AIC
Garden_Temp_Gnd 11	Temperature 10K3A1 -40-150 °C AIC
Boilerhouse_Temp	Temperature 10K3A1 -40-150 °C AIC
Cafe_Temp_Gnd 13	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_1st 15	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_1st 16	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_1st 17	Temperature 10K3A1 -40-150 °C AIC
Garden_Temp_1st 18	Temperature 10K3A1 -40-150 °C AIC
Garden_Temp_1st 19	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_1st 20	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_1st 21	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_1st 22	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_1st 23	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_1st 24	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_1st 25	Temperature 10K3A1 -40-150 °C AIC
Tank Rm_Temp_1st 26	Temperature 10K3A1 -40-150 °C AIC
Room_Temp_1st 27	Temperature 10K3A1 -40-150 °C AIC
WS_Baromic Pressure	Baromic Pressure 4-20mA 600-1100 hPA AIC
WS_Rain Intensity	Rain Intensity 4-20mA 0-200mm AIC
WS_Outside Temp	Temperature 4-20mA -52 to 60°C AIC
WS_Outside Humidity	Humidity 4-20mA 0 to 100 %rh AIC
WS_Wind Direction	Wind Direction 4-20mA 0-360 Deg AIC
WS_Wind Speed	Wind Speed 4-20mA 0-60 m/s AIC

Energy Meters	
Heat Meter Data	
HeatPump HM_ENERGY WATT-HR	Watt-Hour
CHP HM_ENERGY WATT-HR	Watt-Hour

Boiler HM_ENERGY WATT-HR	Watt-Hour
Solar HM_ENERGY WATT-HR	Watt-Hour
Ground HM_ENERGY WATT-HR	Watt-Hour
First HM_ENERGY WATT-HR	Watt-Hour
ThermalStore HM_ENERGY WATT-HR	Watt-Hour
Gas Meter Data	
Main Gas	TBC
Boiler Gas	TBC
CHP Gas	TBC
Electricity Meter Data	
Main ESB (50001)	
50001.AI4	L1 Voltage
50001.AI5	L2 Voltage
50001.AI6	L3 Voltage
50001.AI7	L1 Amps
50001.AI8	L2 Amps
50001.AI9	L3 Amps
50001.AI10	Power L1
50001.AI11	Power L2
50001.AI12	Power L3
50001.AI13	Power Total kW
50001.AI14	Power Reactive kvar
50001.AI15	Power Factor
50001.AI16	Frequency
50001.AI17	Energy kWh
Kitchen DB (50002)	
50002.AI13	Power Total kW
50002.AI17	Energy kWh
Bin Store DB (50003)	
50003.AI13	Power Total kW
50003.AI17	Energy kWh
Container DB (50004)	

50004.AI13	Power Total kW
50004.AI17	Energy kWh
FAP (50005)	
50005.AI13	Power Total kW
50005.AI17	Energy kWh
RWHP (50006)	
50006.AI13	Power Total kW
50006.AI17	Energy kWh
Comms Panel (50007)	
50007.AI13	Power Total kW
50007.AI17	Energy kWh
Spare_1 (50008)	
50008.AI13	Power Total kW
50008.AI17	Energy kWh
Distribution Board (50009)	
50009.AI13	Power Total kW
50009.AI17	Energy kWh
Spare_2 (50010)	
50010.AI13	Power Total kW
50010.AI17	Energy kWh
Workshop DB (50011)	
50011.AI13	Power Total kW
50011.AI17	Energy kWh
MCC (50012)	
50012.AI13	Power Total kW
50012.AI17	Energy kWh
Platform Lift (50013)	
50013.AI13	Power Total kW
50013.AI17	Energy kWh
EXL DB (50014)	
50014.AI13	Power Total kW
50014.AI17	Energy kWh

CHP (50015)	
50015.AI13	Power Total kW
50015.AI17	Energy kWh
Photovoltaic (50016)	
50016.AI13	Power Total kW
50016.AI17	Energy kWh
Lighting (50017)	
50017.AI13	Power Total kW
50017.AI17	Energy kWh
General Services (50018)	
50018.AI13	Power Total kW
50018.AI17	Energy kWh
Power (50019)	
50019.AI13	Power Total kW
50019.AI17	Energy kWh

1.6.2 Other Pilot Sites

The other pilot sites contain various isolated legacy systems (electricity meters, solar inverters, EV chargers) from which data will be collected. Existing legacy systems will be augmented by IoT sensors.

sensor_type	**Parameter**
Main Electricity	kWh
Indoor temperature (for different zones)	C
Solar generation	kWh
EV charging	kWh
Battery charge	kWh
Gas	M3
Plug loads	kWh

1.6.3 Common Data

In addition to site specific data collected from the sites, common data for the pilots will be collected as listed below.

Data_type	Parameter
Electricity Market Price	€/kWh
External temperature	C
Heating degree days	Days.C

1.7 LTU Demo Site

The data from LTU site is divided into the following parts: 1) Smart meters for both electricity and District Heating from the energy provider Skellefteå Kraft (SKR). This smart meters send data to SKR and through them the data will be received in the project. This includes an overall reading of general consumption for the entire building together in terms of "Electricity consumption" in kWh and for "District Heating" both primary and secondary side. Part of the hot water is also used for the ventilation pumps and the readings are inclusive of this consumption.

For electricity consumption individual consumption can be read and retrieved as well since the residents have signed the consent form. This is viewable through the online SKR portal as well. The "District Heating" includes both space heating for the building as well as the domestic hot water heating. For the heating system, there is currently no historical values but these will be collected from now onwards once the APIs are in place. These values can be viewed currently through a software called "Automate". It is possible to view trends for example for one hour once this is enabled. This includes the two pumps for the heating system and heat regulator. Parameters included are temperature measurements. It is the same for the heat exchanger in the heating system.

There is possibility to use three existing temperature devices that are sometimes used by the building management company to measure temperature in different parts/rooms of the building which log the data in the local device. These are not connected to the Internet, and the data is retrieved when the devices are collected after a certain period. The sampling rate can be set as preferred in these devices.
For the ventilation, there are no historical values currently. These values can be viewed by connecting to a BMS like Styrportalen where some trends can be viewed for the past hour once data logging is enabled. The parameters include, temperature, efficiency of the reheater, pressure of the ventilation system (used to control the air pressure) and speed of the fan.

An excel file is attached for heating system part which is generated through Automate software and KTC the company that provides the devices, this includes a sample tag list of all parameters and values that may be available for this system.

2 Fair data

2.1 Making data findable, including provisions for metadata

Efficient data management is vital for the success of projects. This is even more important for large-scale integrating projects such as PHOENIX, which includes a large number of data streams coming from five different pilots that push data –of many different types- into the platform.

This document comprises the considerations on how to manage the data and information created during the project. As the data related to scientific publications will be considered under the open data philosophy, the partners of the project will be encouraged to make available the data needed to replicate the results. With respect to deliverables, all data provided on them will be treated depending on the case. Data coming with personal information will be anonymised and presented only on an aggregated manner. Regardless of its use, all data will be stored in a secured, software platform, which will be accessible with the protection of passwords.

With the purpose of making more understandable the kind of data that is used in the project, an excel file has been created in which all partners register the data they have in terms of historical records, and that data that has been collected thanks to the start of PHOENIX. This file has a series of columns and the consortium describes the following attributes of each stream of data:

- **Start date**: Time at which the data started being collected.
- **End date**: Time at which the data stopped being collected.
- **Sampling period**: Time elapsed between data readings.
- **Completeness:** Percentage of data that is missing.
- **Physical quantity:** Type of variable being collected (can be categorical).
- **Unit:** Physical unit of the quantity**.**
- **Device:** Device that has gathered the data.

- **Scope:** Area from which the data is being captured (Whole building, room, socket).
- **Responsible:** Person responsible for the data.
- **Format:** Format in which the data has been stored. It could be file, or database.
- **Sensitive:** Flag describing if the data contains sensitive information.
- **Comments:** Other comments that were not reflected in the previous fields.

Regarding the representation of the data that PHOENIX is going to use, there are different technologies that will be used. Among these alternatives, there are open and standard initiatives, such as NGSI and NSGI-LD, which are promoted by FIWARE and ETSI to allow a homogeneous way of representation of heterogeneous information based on JSON documents. NGSI, in its current version, represents contextual entities following a straightforward approach. Contextual entities are composed of the following properties: an identifier, a type, and a series of attributes, as we can see in the following example:

```
{
   "id":"device-001A",
   "type": "device",
   "name": "SensorFacultad",
   "dateCreated": "20201-01-01T10:00:00",
   "dataProvider": {
      "@type": "Organization",
      "name": "Universidad de Totana"
    }
   "source": "Sensor Temperatura y Humedad",
   "belongsInfrastructure": 34,
   "isInEnclosure": "be-SalaDibulibu",
   "category": "sensor",
   "deviceClass": "Clase 0",
   "controlledProperty": ["temperature","humidity"],
   "parametersOfMeasurements":["temperatureParameters","humidityParameters"],
   "value":"t=21.2;h=35.4",
   "serialNumber": "IA0201E16000100052",
   "macAddress": "AA:BB:CC:DD:EE:FF",
   "supportedProtocol": ["lora", "mqtt"],
   "owner": "myDevice Inc.",
```

```
    "brandName": "myDevice",
    "modelName": "myModel",
    "manufacturerName": "myDevice Inc.",
}
```

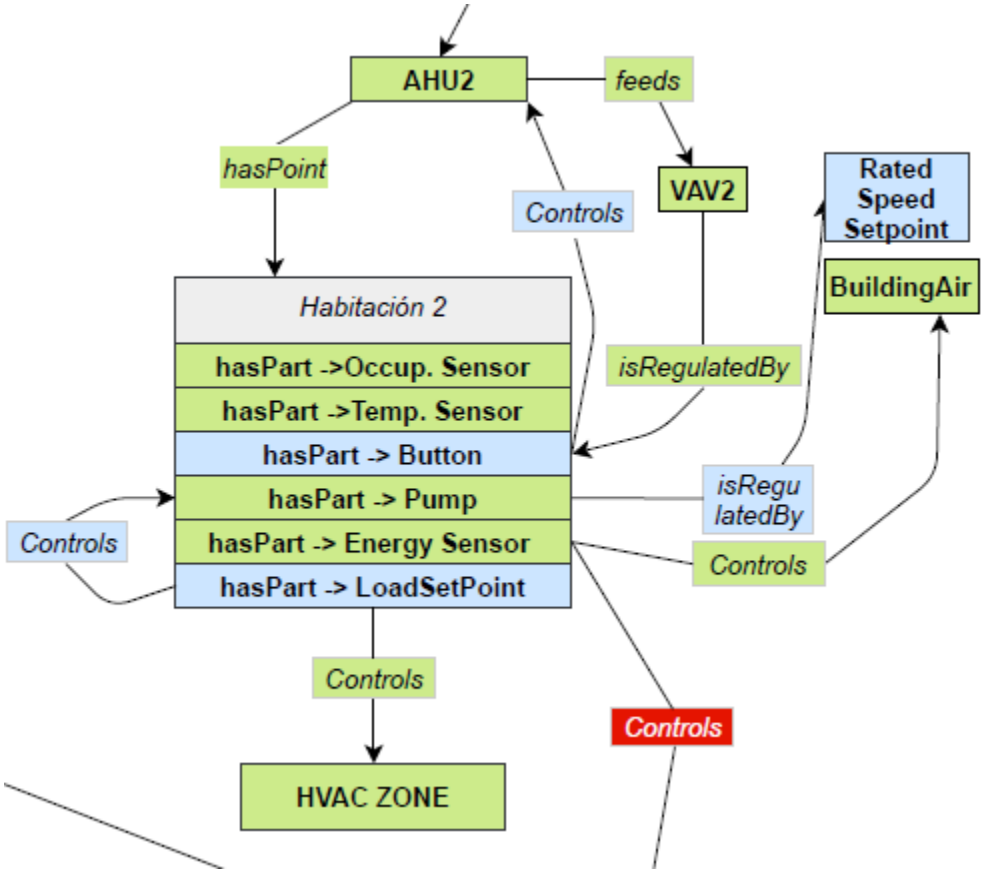

Figure 3 Example of data model using BRICKS a candidate to model the data in PHOENIX to facilitate interoperability.

2.2 Making data openly accessible

PHONIX is not part of ORD. However, when the situations are adequate, the publication of the data will be considered. For that, data repositories will be considered. In these cases, the data used for generating the scientific results will be the ones published.

In case the data intended to be public, once filtered and checked against any privacy conflict, will be stored on the local repositories, and will be provided to researchers and other third parties under request. This will allow us to do a closer follow up of the propagation of the data and the impact of the research. When requested, the data will be transferred on csv format, as it tends to be the most versatile one. For longer collaborations, access to APIs of the project will be provided.

With respect to the deliverables of the project, the documents have different visibility depending on the information they include, which is already pre-defined in the "Deliverables with type and dissemination level" tables in the Consortium Agreement. Some of them are considered as confidential and some others are public. The latter will be available on the website in PDF format.

2.3 Making data interoperable

The data will be put on a canonical format using the ontology selected by the project, which is likely to be an evolution of the BRICKS schema. The use of this data model will facilitate the interoperability of data among consortium partners and also with third parties. Data that cannot enter into the categories of BRICKS will be given a metadata that will be sufficient to describe what they are and how to use it.

The description of the data models used, either if they belong to standardised schemas or if they have been created ad-hoc, will be available as part of the content of several deliverables. Any third party wanting to use the data generated by the project, will be pointed to these documents to facilitate the use of the data.

As regards the algorithms of intelligence and other methodologies, they will be developed with an open source philosophy, and they will be made public when the authors considered it appropriate.

2.4 Increase data re-use (through clarifying licences)

The data that is made public will not be given any restrictive licence to facilitate its use for research and innovation. It is expected that some data will be published under more restrictive licenses, but each case will be studied in detail. The data which is public will be accessible during the length of the project on project's website.

With respect to data quality assurance, the UMU team has developed a mechanism to evaluate data quality[1]. This methodology for ensuring quality of data has been considered as an important part of this project due to its nature. According to the increasing number of IoT devices, the amount of data gathered in present days is rather large and continuously growing. The availability of new sensors presents in IoT devices, data sources from internet resources and open data platforms

[1] Tomás Alcañiz, Aurora González-Vidal, Alfonso P. Ramallo and Antonio F. Skarmeta. Quality of Information within Internet of Things Data on *Data Integrity and Quality*, edited by Dr. Santhosh Kumar. Publisher: IntechOpen (2021).

provide new possibilities for innovative applications and use-cases that may be unprecedented. However, the dependence on data for the provision of services, creates the necessity of assuring the quality of data, so as to ensure the viability of the services, and furthermore, to check how sound the information within that data is. A richer context in which the quality of information can be measured is provided by geo-localised data, having services that are more advanced. In order to support the extraction of valuable information, several metrics have been defined as indicators of the quality of data will be used on a quantifiable, fast, reliable, and human understandable way. The metrics are based on tested statistical indicators, and can be used on single-sensor and multi-sensor scenarios. Statistical analysis, machine learning algorithms and contextual information are some of the methods to create the quality indicators. The development of a general metric on the basis of several data-quality facets allows data-consumers to determine the quality of the data and, in case of finding an anomaly, determining if it is due to a system failure or an external phenomenon by examining the data's intrinsic patterns. These metrics have been applied in two real scenarios, smart packing and environmental sensing for smart buildings, and in both cases the methods have been proven as internally consistent, fast, efficient, and effective for representing the quality of the data.

3 Allocation of resources

Making data FAIR will represent a cost on resources for the project. Cost associated to open access –to-research data are eligible as part of the Horizon 2020 grant (if compliant with the conditions of the Grant Agreement), therefore the creation of a data management plan is considered.

With respect to papers published under open access license, resources will be allocated in management and dissemination of the general budget. For those specific of a given work package, the resources will be allocated on the budget of that work package under "other costs". This is because no specific budget has been allocated for this, so the costs are implicitly integrated on the work load of the partners. As it is seen in Figure 3, open access can come on the form of *green open access* that with the help of research portals such as Research Gate, give visibility to accepted manuscripts after the journal embargo, eliminating the cost of gold open access.

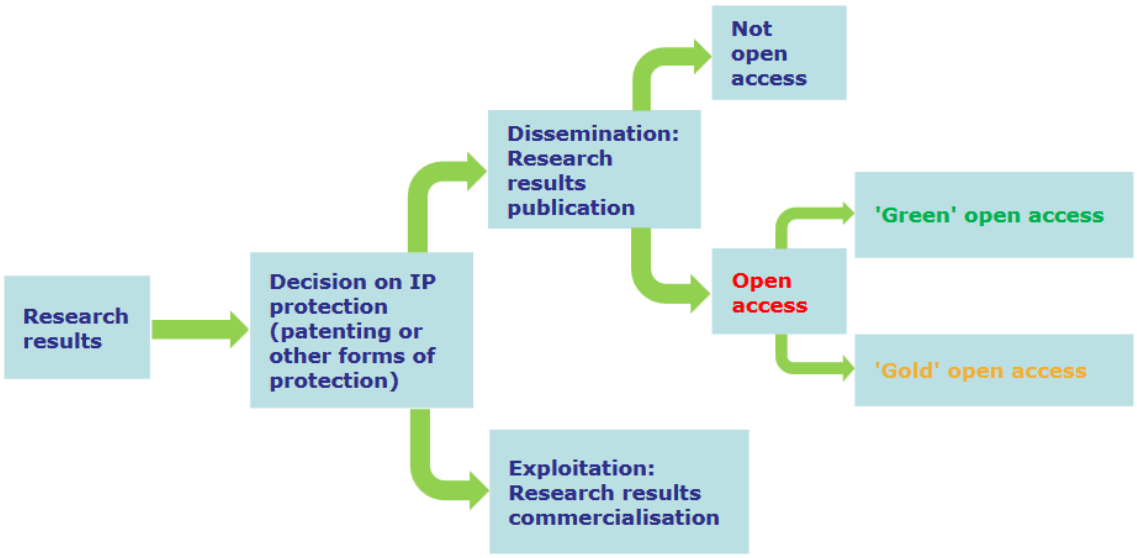

Figure 4. Schematic view of the path to open access included in European IPR Helpdesk Fact Sheet Open access to publications and data in Horizon 2020: Frequently Asked Questions (FAQ) 2014.

Another resource for data management for this project is the appointing of the Data Protection Officer (Julián Valero Torrijos UMU) who will supervise the management of data during the length of the project. Also within the consortium, Josiane Xavier (SAGOE) will have the role of Quality Assurance and Data Manager and will be involved in all processes concerning data. These human resources dedicated to data management are already allocated within project PHOENIX. Nevertheless, the person responsible for coordinating PHOENIX data management process will be the Data Manager, the General Assembly will be the ultimate decision-making body of the Consortium, responsible for taking major strategic decisions with respect to data management if necessary. The General Assembly will also promote consensus in case of conflict and, if no consensus can be found, will take decisions according to the procedures and rules defined in the Consortium Agreement.

4 Data Security

Data access and sharing activities will be implemented in compliance with the privacy and data collection rules and regulations, as they are applied both nationally and in the EU, as well as with H2020 rules. Concerning the results of the project, these will become publicly available based on the IPRs as described in the Consortium Agreement.

Due to the nature of the data involved, some of the results that will be generated by each project phase will be restricted to authorized users, while other results will be publicly available. One

possibility would be to ask users to pre-register for the purpose of using the system and will then need to authenticate them against a user database. If successful, the users will have roles associated with them. These roles will determine the level of access that a user will be given and what they will be permitted to do.

The data from sensors and other devices will be stored on a secured location that performs periodic back-ups to recover the information from an interval smaller than a week.

5 Ethical aspects

A detailed data security plan has been elaborated for this project, maximising sensitive information protection and privacy. More information about this plan has been provided in Deliverables 9.1 and 9.2 of this project.

6 Other

The guideline document "FAIR Data Management in Horizon 2020" has been used for the development of PHOENIX's Data Management Plan.

Appendix Data attributes for UMU Pilot

sensor_type	energy_power_electricity
sensor_id	322964003
building_area	chemistry
attribute_to_read	**device_ID;timestamp;Active_energy**;Reactive_inductive_energy;Reactive_Capacitive_energy;Reactive_energy;Active_power_ph1;Reactive_power_ph1;Aparent_power_ph1;Active_power_ph2;Reactive_power_ph2;Aparent_power_ph2;Active_power_ph3;Reactive_power_ph3;Aparent_power_ph3;Current_phase1;Current_phase2;Current_phase3;Voltage_ph1ph2;Voltage_ph2-ph3;Voltage_ph3ph1;Voltage_ph1;Voltage_ph2;Voltage_ph3
Sampling Rate	~10 minutes

sensor_type	energy_power_electricity
sensor_id	→ CVM-Mini (room 1,16) → CVM-Mini (room 1.18) → CVM-Mini (room 1,16) → CVM-Mini (room 1.18)) → CVM-Mini (room SEMA 2 – 1.9) → CVM-Mini (room SEMA 4 – 1.7) → CVM-Mini (room 1.9) → CVM-Mini (room 1.7) → Schneider PM710 #1 (general) → Schneider PM710 #2 (CP(+3) CLIMA 2- PROD.R (floor 3, at zone B)) → Schneider PM710 #3 (CP(+2) CLIMACRAI.R (floor 2, at CRAI)) → Schneider PM710 #4 (CP(+5) CLIMA 1-PROD.R (floor 5, at zone A))
building_area	pleiades

attribute_to_read	**device_ID;timestamp;Active_energy**;Reactive_inductive_energy;Reactive_Capacitive_energy;Reactive_energy;Active_power_ph1;Reactive_power_ph1;Aparent_power_ph1;Active_power_ph2;Reactive_power_ph2;Aparent_power_ph2;Active_power_ph3;Reactive_power_ph3;Aparent_power_ph3;Current_phase1;Current_phase2;Current_phase3;Voltage_ph1ph2;Voltage_ph2-ph3;Voltage_ph3ph1;Voltage_ph1;Voltage_ph2;Voltage_ph3
Sampling Rate	~2 minutes

sensor_type	electricity
sensor_id	335544938-→ CVM-1D (lightning) (first half room 0.5) → CVM-1D (lightning) (second half room 0.5) → CVM-1D (lightning) (first half rooms 0.3 y 0.4) 335544944 → CVM-1D (fuerza rooms 0.3 y 0.4) 335544965 → CVM-1D (4.2 (lightning)) 335544966 → CVM-1D (4.3 (lightning)) → CVM-1D (4.2) → CVM-1D (4.3) 335545030 → CVM-1D room 0.5 → CVM-1D (teachers' area rooms 0.3, 0.4 y 0.5) → CVM-1D (lightning) (second half rooms 0.3 y 0.4) → CVM-1D (1,16 (lightning group)) → CVM-1D (1.18 (lightning group)) 335545101 → CVM-1D (1.9 (lightning grupo)) 335545102 → CVM-1D (1.7 (lightning grupo)) → CVM-1D (R2 (1.1)) → CVM-1D (R3 (1.2)) → CVM-1D (R4 (B1.1.008)) → CVM-1D (R5 (B1.1.007)) → CVM-1D (R6 (B1.1.006)) → CVM-1D (A1 – 1.6 (lightning red)) 335545109→ CVM-1D (A2 – 1.1 (lightning red))
building_area	pleiades
attribute_to_read	**device_ID;timestamp;Active_energy**;Reactive_inductive_energy;Reactive_Capacitive_energy;Reactive_energy;Aparent_energy;Active_power;Reactive_power;Aparent_power;Voltage;Electricity
Sampling Rate	~ 2 minutes

sensor_type	hvac
sensor_id	322962918, 322962828, 322962838, 322963044, 322963048, 322963142, 322963130, 322963146, 322962874, 322962878, 322962888, 322962886, 322962904, 322962860, 322963074, 322963086, 322963090, 322962958, 322962946, 322962980, 322962982, 322963002, 322963160, 322963168, 322963190 *(these HVACs are installed in different rooms of the building that will be controlled by the Energy Patrol videogame)*
building_area	chemistry
attribute_to_read	**device_ID;timestamp;ON_OFF**;Fan_speed;shades_config;**temperature;r egulation**
Sampling Rate	~10 minutes

sensor_type	hvac (Not all on this list have data available) 335545141 --> (room 02.ED136.B1.0.000)8 335545143 --> (room 0.8) 335545145 --> (room 0.8) 335545147 --> (room 0.4) 335545149 --> (room 0.4) 335545151 --> (room 02.ED136.B1.0.052) 335545153 --> (room 02.ED136.B1.0.051.1) 335545155 --> (room 02.ED136.B1.0.050) 335545157 --> (room 02.ED136.B1.0.055) 335545159 --> (room 02.ED136.B1.0.051.1) 335545161 --> (room 02.ED136.B1.0.051) 335545163 --> (room 02.ED136.B1.0.047) 335545165 --> (room 02.ED136.B1.0.047) 335545167 --> (room 02.ED136.B1.0.026) 335545169 --> (room 02.ED136.B1.0.026) 335545171 --> (room 02.ED136.B1.0.025) 335545173 --> (room 02.ED136.B1.0.024) 335545175 --> (room 02.ED136.B1.0.011) 335545177 --> (room 02.ED136.B1.1.042) 335545179 --> (room 02.ED136.B1.1.041.2) 335545181 --> (room 02.ED136.B1.1.041.3) 335545183 --> (room 02.ED136.B1.1.043) 335545185 --> (room 02.ED136.B1.1.041.1) 335545187 --> (room 02.ED136.B1.1.044) 335545189 --> (room 02.ED136.B1.1.045) 335545191 --> (room 02.ED136.B1.1.046) 335545193 --> (room 02.ED136.B1.1.047) 335545195 --> (room 02.ED136.B1.1.049)

	335545197 --> (room 1.19) 335545199 --> (room 1.19) 335545201 --> (room 1.18) 335545203 --> (room 1,16) 335545205 --> (room 1.17) 335545207 --> (room 2.18) 335545209 --> (room 2.19) 335545211 --> (room 2.19) 335545213 --> (room 02.ED136.B1.2.013) 335545215 --> (room 02.ED136.B1.2.014) 335545217 --> (room 2.14) 335545219 --> (room 2.15) 335545221 --> (room 02.ED136.B1.1.008) 335545223 --> (room 02.ED136.B1.1.007) 335545225 --> (room 02.ED136.B1.1.006) 335545227 --> (room 1.9) 335545229 --> (room 1.10) 335545231 --> (room 1.1) 335545233 --> (room 1.2) 335545235 --> (room 1.7) 335545237 --> (room 1.8) 335545239 --> (room 1.6) 335545241 --> (room 2.1.1) 335545243 --> (room 2.3.1) 335545245 --> (room 2.1) 335545247 --> (room 2.7.2) 335545249 --> (room 2.7) 335545251 --> (room 02.ED136.B1.2.024) 335545253 --> (room 02.ED136.B1.2.025) 335545255 --> (room 2.7.1) 335545257 --> (room 2.3) 335545259 --> (room 2.4) 335545261 --> (room 2.9.2) 335545263 --> (room 2.9) 335545265 --> (room 2.9.1) 335545267 --> (room 2.13.1) 335545269 --> (room 2.13) 335545271 --> (room 3.6) 335545273 --> (room 3.6) 335545275 --> (room 02.ED136.B1.3.013) 335545277 --> (room 3.1) 335545279 --> (room 3.2) 335545281 --> (room 02.ED136.B1.3.008) 335545283 --> (room 3.4) 335545285 --> (room 3.5) 335545287 --> (room 02.ED136.B1.3.015) 335545289 --> (room 3.10) 335545291 --> (room 3.11) 335545293 --> (room 3.12) 335545295 --> (room 4.2) 335545297 --> (room 4.3) 335545299 --> (room 4.3) 335545301 --> (room 4.1)

	335545303 --> (room 4.6) 335545305 --> (room 4.7) 335545307 --> (room 4.8) 335545309 --> (room 4.5) 335545311 --> (room 4.5) 335545313 --> (room 4.9) 335545315 --> (room 4.10)
building_area	pleiades
attribute_to_read	device_ID;timestamp;ON_OFF;Fan_speed;shades_config;temperature;r egulation
Sampling Rate	~ 10 minutes

sensor_type	ambient_temperature
sensor_id	(Not all on this list have data available) "335545326": "temperature propel", "335545328": "temperature return", "335545330": "external temperature", "335545341": "temperature propel", "335545342": "temperature return", "335545343": "external temperature", "335546173": "temperature propel informativa 1", "335546174": "temperature propel informativa 2", "335546175": "temperature propel informativa 3", "335546176": "temperature return/ambient informativa 1", "335546177": "temperature return/ambient informativa 2", "335546178": "temperature return/ambient informativa 3", "335546345": "external temperature", “335546346": "temperature return", "335546371": "temperature propel", "335547105": "temperature maceta 2", "335547106": "temperature maceta 1", "335547107": "temperature ambient 2", "335547653": "temperature ambient 1", "35544455": "temperature room 4.2", "335544458": "temperature room 4.3", "335544461": "temperature room 4.4", "335544472": "temperature room 4.11", "335544479": "temperature room 4.6", "335544488": "temperature room 4.7", "335544491": "temperature room 4.8", "335544499": "temperature 1 room 4.5", "335544506": "temperature room 4.9", "335544517": "temperature room 4.10", "335544522": "temperature room 3.1", "335544529": "temperature room 3.2", "335544532": "temperature room 3.6", "335544542": "temperature room 3.3",

	"335544547": "temperature room 3.4", "335544554": "temperature room 3.5", "335544557": "temperature room 3.10", "335544563": "temperature room 3.11", "335544569": "temperature room 3.12", "335544575": "temperature room 3.13", "335544605": "temperature room 2.1", "335544613": "temperature room 2.7", "335544628": "temperature room 2.3", "335544639": "temperature room 2.4", "335544648": "temperature room 2.9", "335544655": "temperature room 2.10", "335544660": "temperature room 2.14", "335544672": "temperature room 2.13", "335544685": "temperature room 2.15", "335544692": "temperature room 2.16", "335544699": "temperature room 2.17", "335544720": "temperature room 2.18", "335544721": "temperature room 2.19", "335544736": "temperature room 1.1", "335547743: "temperature room 1.7", "335547759: "temperature room 1.7", "335547765: "temperature room 1.8", "335547771: "temperature room 1.8", "335547778: "temperature room 1.8", "335547785: "temperature room 1.9", "335547791: "temperature room 1.9", "335547798: "temperature room 1.9", "335547804: "temperature room 1.10", "335547810: "temperature (room 1.10", "335547817: "temperature room 1.10", "335547837: "temperature room 1.7", "335547856: "temperature room 1,16", "335547861: "temperature room 1,16", "335547871: "temperature room 1.17", "335547879: "temperature room 1.17", "335547886: "temperature room 1.18", "335547887: "temperature room 1.18", "335547901: "temperature room 1.19", "335547902: "temperature room 1.19", "335547903: "temperature room 1.19", "335547925: "temperature room B1.2.013", "335547926: "temperature room B1.2.013", "335547940: "temperature room B1.2.014", "335547950: "temperature room 2.18", "335547951: "temperature room 2.18", "335547965: "temperature room 2.19", "335547966: "temperature room 2.19", "335547967: "temperature room 2.19"
building_area	pleiades

attribute_to_read	timestamp; Value
Sampling Rate	~ 10 minutes

sensor_type	ambient_co2
sensor_id	"335547753": "co2 F0A4 room 1.7", "335547772": "co2 F0A9 room 1.8", "335547792": "co2 F0A6 room 1.9", "335547811": "co2 F0AB EA2 010V 1.10", "335547862": "co2 F0A5 room 1,16", "335547876": "co2 F0A0 room 1.17", "335547888": "co2 F0A2 room 1.18", "335547904": "co2 F0A1 room 1.19", "335547927": "co2 F0A8 room B1.2.013", "335547941": "co2 F0A3 room B1.2.014", "335547952": "co2 F0AA room 2.18", "335547968": "co2 F0A7 room 2.19"
building_area	pleiades
attribute_to_read	timestamp; Value
Sampling Rate	~ 2 minutes

sensor_type	external_temperature_weather_station
sensor_id	335546942
building_area	pleiades
attribute_to_read	timestamp; Value
Sampling Rate	~20 minutes

sensor_type	external_humidity_weather_station
sensor_id	335546943
building_area	pleiades

attribute_to_read	timestamp; Value
Sampling Rate	~ 20 minutes

sensor_type	wind_speed_weather_station
sensor_id	335546945
building_area	pleiades
attribute_to_read	timestamp; Value
Sampling Rate	~10 minutes

sensor_type	wind_direction_weather_station
sensor_id	335546946
building_area	pleiades
attribute_to_read	timestamp; Value
Sampling Rate	TBA

sensor_type	solar_radiation_weather_station
sensor_id	335546947
building_area	pleiades
attribute_to_read	timestamp; Value
Sampling Rate	~ 20 minutes