Future Generation Computer Systems 92 (2019) 1066–1079

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

An open IoT platform for the management and analysis of energy data

Fernando Terroso-Saenz *, Aurora González-Vidal, Alfonso P. Ramallo-González, Antonio F. Skarmeta

Department of Information and Communications Engineering, Computer Science Faculty, University of Murcia, Spain

HIGHLIGHTS

- IoT platform for the management of energy data in buildings.
- Includes several inner features to support data analytics in the energy domain.
- Based on the open IoT initiative FIWARE.
- Evaluated in a real pilot with comprising several buildings.

ARTICLE INFO

Article history:
Received 15 March 2017
Received in revised form 24 May 2017
Accepted 23 August 2017
Available online 6 September 2017

Keywords:
IoT platform
Energy consumption
FIWARE
Data mining

ABSTRACT

Buildings are key players when looking at end-use energy demand. It is for this reason that during the last few years, the Internet of Things (IoT) has been considered as a tool that could bring great opportunities for energy reduction via the accurate monitoring and control of a large variety of energy-related agents in buildings. However, there is a lack of IoT platforms specifically oriented towards the proper processing, management and analysis of such large and diverse data. In this context, we put forward in this paper the IoT Energy Platform (IoTEP) which attempts to provide the first holistic solution for the management of IoT energy data. The platform we show here (that has been based on FIWARE) is suitable to include several functionalities and features that are key when dealing with energy quality insurance and support for data analytics. As part of this work, we have tested the platform IoTEP with a real use case that includes data and information from three buildings totalizing hundreds of sensors. The platform has exceed expectations proving robust, plastic and versatile for the application at hand.

1. Introduction

Several reports claim that residential and commercial buildings represent around 30%–40% of the overall energy consumption in Europe and in the United States [1,2]. Because of this, buildings are known to be the largest end-use energy contributor followed by transport and industry, and therefore they are a clear target for potentially reducing global energy consumption substantially.

Despite being great consumers, there is some evidence that shows that public and private buildings have not fully exploited all opportunities available to increase their energy efficiency. On the contrary, they suffer from a rather substantial energy waste that is partly due to inefficient heating, cooling, lighting and other power system (equipment) [3], due to bad use of the systems (behavior) [4] and due to poor fabric efficiency [5]. Although the implementations of measurements to improve the first or the third category can be rather expensive, it has been seen that soft measurements that focus on the change of behavior of buildings' users are cheap, but yet, can contribute greatly to the reduction of energy use [6].

In order to address the aforementioned inefficiencies due to lack of understanding on how the systems should be operated and other behavioral related aspects in the building sector, one could consider the use of Information and Communication Technologies (ICT) and, more specifically, of the Internet of Things (IoT). This new paradigm that also exists at the domestic level could be used as an instrument to make a realization of the so called *Smart Building*. In fact, it is foreseen that from 2 to 3 houses out of 10 will be equipped with up to 500 smart devices in the near future [7].

The installation of smart meters and In Home Energy Displays to make households aware of their energy consumption is not new [8,9]. The adoption of these devices seems to be an opportunity to exploit them for the reduction of energy use when looking at the available scientific literature (will be detailed later). However, one may also think that the technological effort to deploy such systems may be substantial and become a barrier to achieve this level of technification of the buildings. Nevertheless this technification seems to be happening naturally.

* Corresponding author.
E-mail address: fterroso@um.es (F. Terroso-Saenz).

http://dx.doi.org/10.1016/j.future.2017.08.046
0167-739X/© 2017 Elsevier B.V. All rights reserved.

The large amounts of IoT data that will be coming from buildings in the near expected future will have to be analyzed to reveal insights that could help to obtain, expose and understand knowledge from buildings. In turn, this derived knowledge should be able to help to achieve meaningful energy saving strategies and interventions in the targeted buildings [10].

These wealth of information about energy use, offers a great opportunity according to some literature on energy feedback that suggests that intelligent feedback, (that with an extra larger of computation over simple observation) is an effective technique for the reduction of energy demands via behavioral change [11]. Only with a platform capable of making this possible, the implementation of this new paradigm will be successful.

In the IoT ecosystem, several platforms have emerged providing support from the sensorization stage to the stage of management and storage of the data in different forms [12]. In that sense, one of the most large-scale affords is the FIWARE platform, a key initiative of the Future Internet Public–Private Partnership (PPP) to create a well-aligned set of open enablers to receive, process, contextualize and publish IoT data from and for smart cities including from city-wide information to dwelling specific data.[1]

Despite all the reasons exposed before, little efforts have been made so far in order to adapt such platforms to building energy management. This energy ecosystem comprises a set of particularities that should be targeted in a specific manner. After analyzing the few examples of studies that have tried to tackle this problem, one can see that it exists a pressing need to apply different data mining techniques in the building energy domain mainly focusing on consumption prediction and pattern discovery or failure tolerance [13]. Thus, IoT energy platforms should include functions for data analysis among their features.

Although giving insighting knowledge behind data is an instrumental aspect of the wealth produced by the IoT, existing platforms are still limited when it comes to integrate data processing and analytic techniques suitable for IoT ecosystems [14]. This is a fundamental limitation of the state of the art as it is key to ensure that the platform will work on the new paradigm of providing tailored, real-time energy feedback to people. This also includes features to support the easy extension of platforms to allocate new data mining techniques comprising common steps in the data mining process. Examples of such features are built-in data-cleaning mechanisms for data pre-processing and storage solutions that would facilitate the execution of online and offline data mining algorithms.

All the aforementioned limitations have motivated us to envision, design, develop and validate what we called the IoT Energy Platform (IoTEP). The key strength of IoTEP is that it is, to our knowledge, the first holistic solution to large scale building energy data management from IoT.

Unlike existing IoT platforms, IoTEP is mainly oriented to support and ease the analysis of large amounts of heterogeneous energy data. A simplified overview of the platform IoTEP is shown in Fig. 1 representing its key features.

To begin with, IoTEP has been designed to easily retrieve either the most up-to-date readings of each sensor within a building, or to retrieve the historic data from such sensors. By means of these two types of access, the platform facilitates the application of both online and offline data analyses over the collected data. As we will see on further sections, this functionality is implemented with two FIWARE storage components, the ORION context broker and COMET. For both enablers, a NGSI-based information model has been defined in order to homogenize all the measured energy-related data.

[1] https://catalogue.fiware.org/.

Secondly, a real-time data cleaning module has been designed as a built-in component of IoTEP. With this, sensor readings are filtered by discarding potential outliers before injecting them in the storage components. This ensures a more efficient use of the resources. For this feature, we have followed a Complex Event Processing (CEP) approach that allows the real-time processing of event streams.

In addition to the above mentioned features, the platform includes also a mechanism to detect volatility changes in the incoming energy data. This mechanism intends to perceive meaningful shifts in such data that might need to re-launch the data-mining services that run within the platform.

Finally, IoTEP features a novel mechanism to automatically identify high-level areas in a building with certain energy-related similarities by means of clustering techniques. The benefit of these virtual areas is twofold. Firstly, they provide alternative representations of the energy status of a building beyond its physical structure; and secondly, they can help in the performance of other data mining analyses by reducing redundancies and defining different granularity levels in the captured sensor data.

Summarizing, the platform presented in this paper intends to be the first stage towards the full adaptation of the IoT paradigm in the retrieval, management and, above all, analysis of energy data in buildings. Considering the need of developing tools that are able to provide personalized real time feedback to change behaviors, and with them, have the potential to reduce energy use, IoTEP is intended to become the stepping stone for the development of such tools.

The paper is structured as it follows: Section 2 provides an overview of the state of the art in this research area. Section 3 looks into the IoT energy platform, including its architecture and its functional modules. Section 4 provides an evaluation of some of the features of the platform; and Section 5 concludes the paper with some final remarks and conclusions.

2. Related work

The present work is based upon two different lines of research, the management of energy data and the implementation of IoT platforms. Consequently, an overview of both lines is put forward in this section.

2.1. Energy data management systems

During the last years, some initiatives within the cloud computing domain have been made to intelligently manage energy data of buildings. In that sense, Zhou et al. [15] described a model for big-data energy management ranging from the collection and pre-processing of data to its further analysis and the final exposition to services. However, it only provides a theoretical approach.

From a practical perspective, the Dynamic Demand Response (D^2R) platform [16] makes use of public and private clouds combined with infrastructure and platform as a service for data storage. This platform was extended with *Cryptonite*, a repository to store sensitive *Smart Grid* data [17]. Then, different classes of data-driven forecasting models were generated on top of the whole platform with the purpose of carrying out energy prediction among others.

ElasticStream also provides a prototype solution for energy data management and analysis. In this case, the proposed mechanism transfers energy data to a cloud platform for further analysis on the basis of rate changes in the input data streams [18]. Moreover, Vastardis et al. [19] described a centralized architecture to monitor energy consumption in houses including features of pattern-matching related to the behavioral habits of the target users.

In the work of Ozadowicz [20], the authors propose different approaches to calculate the power demand related to energy

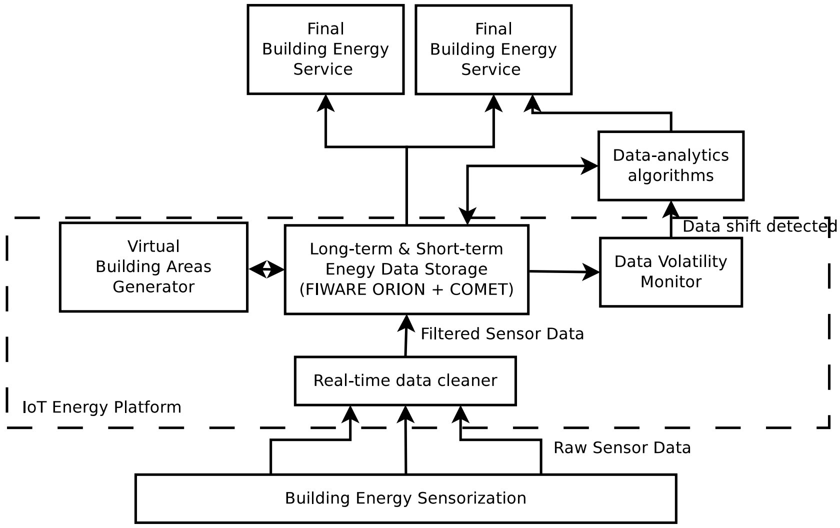


Fig. 1. Conceptual view of the IoT Energy Platform (IoTEP).

consumption using time-driven and event-driven mechanisms for Building Automation and Control Systems. Their Building Energy Management Systems (BEMS) implementation is realized with an IoT platform, introduced by Echelon Corp that includes chips, stacks, communication, application interfaces (API) and management software. Their approaches to calculate the energy demand are based in time (fixed or sliding length of the time windows with the possibility of overlapping) and in events (occupancy).

The MultiAgent System (MAS) named SAVES (Sustainable multiAgent systems for optimizing Variable objectives including Energy and Satisfaction) defined in [21] is used in [22] regarding actual occupant preferences and schedules, actual energy consumption and loss data measured from a real test bed building at the University of Southern California in order to predict energy consumption at different levels (frequency of prediction and device aggregation).

Other works provide energy data management solutions without focusing on analytic aspects. This is the case of the Virtual SCADA architecture for cloud computing (VS-Cloud) that encompasses Cloud Computing for energy data storage [23]. VS-Cloud mainly focuses on the orchestration of components in Smart Grids and the safety storage of sensitive data executed actions, incidents or alarms. Therefore, its domain of application is more related to risk management.

Similarly, the work in [24] proposes an automation platform for energy monitoring. However, such platform does not provide any particular feature to support energy data analytics as it focuses more on the definition of control strategies for energy saving.

Unlike the aforementioned initiatives, our work provides a holistic energy data management and analysis solution. Our platform also follows an open approach by relying on the well-established FIWARE initiative. In that sense, the present work includes explicit features like data volatility monitoring and outliers detection to ease the deployment of data mining algorithms and other services over of the stored data.

FIWARE brings other advantages with respect to previous solutions: firstly, the whole platform orchestration is done by means of lightweight RESTful APIs, that facilitate its further extension; and secondly, the definition of an information model compliant with NGSI standard allows to come up with a homogeneous view of the energy-related data within a building. This feature is key to exploit the potential of gathering energy data. What we propose here is not only an archive of data, but a comprehensive flexible and powerful tool that will serve as the breeding ground for the creation of context-aware tailored energy feedback platforms that could be realized at a scale never considered before, even reaching national levels.

2.2. IoT platforms

The Internet of Things paradigm is the second pillar of this initiative. All the literature indicates that small devices connected to the internet in buildings will be the norm in the near future. With the right algorithms and communication mechanisms, this situation will enable the monitoring and characterization of energy behaviors and energy consumption in buildings.

The need of effective instantiation of IoT under realistic conditions has generated a varied ecosystem of methodologies and tools taking the form of integrated IoT platforms. In that sense, it is possible to find several surveys in the literature that review existing proprietary and open-source platforms in the IoT ecosystem [12,14,25]. Other important aspects like data ownership, security and privacy [26] or data storage [25] have been also deeply studied in the IoT domain. The reader is referred to this sources to expand on the state of the art.

According to such reviews, some relevant IoT platforms follow a similar open-source and centralized approach along with heterogeneous sensor support like IoTEP. This is the case of Nimbits[2] that provides an open source Java library for developing Java, Web and Android solutions to connect to a Nimbits Server. This backend part enables simple processing of the collected data based on rules. However, it does not comprise any advanced data-analytics support. ThingSpeak[3] features the acquisition, visualization and analysis of data but this is done by means of the proprietary Matlab tool, what may make more difficult the popularization of the platform.

One feature frequently neglected by existing IoT platforms is the support of built-in data mining features able to generate new useful knowledge from the collected and stored data [14]. In real IoT deployments, this processing and analysis task has been frequently done by third-party services. However, integrating certain data mining functionalities as built-in features of platforms would provide a great benefit in a wide range of domains, for example: quick statistics, easy to generate digests or sanity checks. In

2 https://www.nimbits.com/.
3 https://thingspeak.com/.

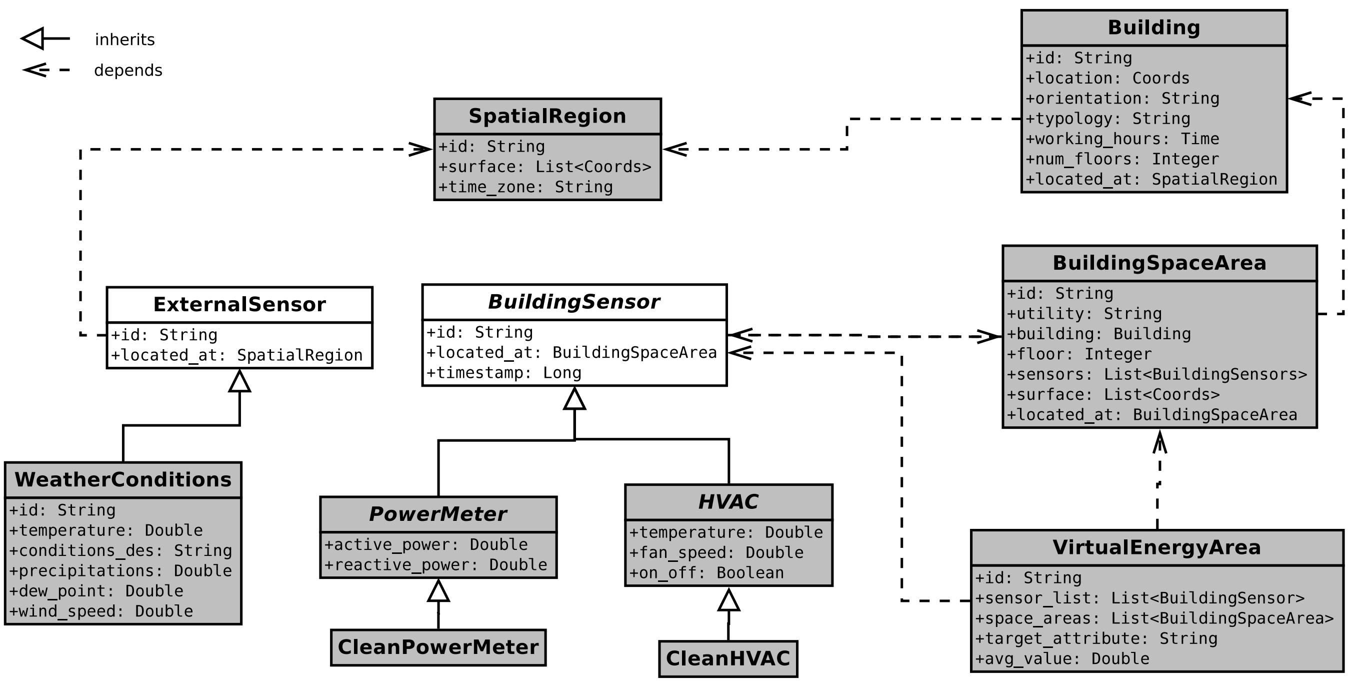


Fig. 2. IoTEP information model.

that sense, only a few IoT platforms actually include native data-analytics features. As a matter of fact, SensorCloud[4] enables a simple interface for common operations like smoothing, filtering and interpolation whereas GroveStreams[5] provides some real-time data analytics mechanisms. However, none of them support sensor heterogeneity nor follow an open source approach like IoTEP does.

As for the energy ecosystem, several research lines have already stated the feasibility and suitability of data analysis in order to increase energy awareness within a building [13]. In that sense, IoTEP provides one of the first steps towards such a data-mining enrichment by providing several features fully focused on easing the analysis of IoT energy data namely, real-time data cleaning, data volatility detection and data reduction procedures.

Finally, our work is enclosed within the FIWARE architecture. The high-level goal of this architecture is to build the Core Platform of the Future Internet, introducing an innovative infrastructure for cost-effective creation and delivery of versatile digital services, providing high QoS and security guarantees. In that sense, FI-LAB [27] conforms live instances of generic enablers, available to developers for free experimentation within this technology.

Some initiatives have started to profit from FIWARE in several domains. One of the most ambitious works is the application on [28] which established a world-wide semantic interoperability solution combining the NGSI, which is part of the core of the FIWARE initiative, and oneM2M context interfaces. Apart from that, [29] demonstrated the suitability of the FIWARE paradigm to compose Future-Internet applications by means of the integration of generic enablers. In a similar manner, [30] put forward a semantic mechanism to integrate data from different types of devices by also using FIWARE components. Finally, in a more functional domain, [31] made use of certain enablers, like ORION context broker, to create a cloud-based gesture recognition application. Also, [32] describes a sensor management for seaports based on the FIWARE platform. It is therefore possible to say that our work seems to be one of the first efforts to make use of FIWARE enablers in the building energy domain, and furthermore in the energy domain in general.

3. IoT Energy Platform (IoTEP)

This section explains in detail the proposed IoTEP solution. Since the management of the energy data is its key feature, we firstly describe the information model used to define all the data within the IoTEP ecosystem; next, we put forward the specific architecture of the platform that deals with the energy data according to the model.

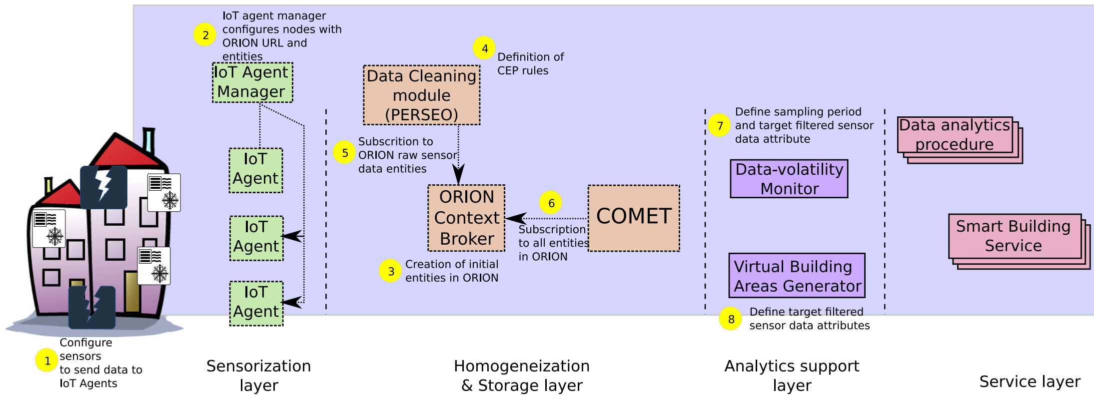
3.1. Information model

One of the first steps towards the realization of IoTEP was to define a common information model for the whole platform. Such a model must be compliant with the NGSI information model commonly accepted in the FIWARE ecosystem, what facilitates interconnection with other models and other users. This information model follows an entity–attribute approach where entities represent real or virtual elements of interest. Each entity has a type what allows to define type-based hierarchies. In this way, an entity has its own defined attributes and the inherited ones from its ancestors. The IoTEP information model is depicted in Fig. 2. The model design follows the UML class notation with two types of relationships, inheritance and dependence. Each of them is represented by a different arrow in the figure. Whilst inheritance indicates that the child element comprises all the attributes of its parent element, the dependence relationship indicates that an instance of the element at the arrow's origin contains an attribute referencing at one or more instances of the element at the arrow's destination.

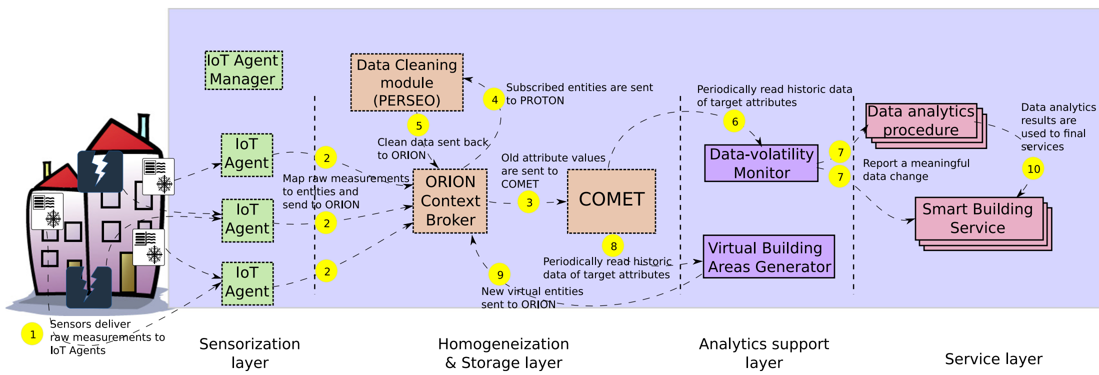
Focusing on the content of the model, one can find among its components three key elements related to the energy ecosystem of a building by means of NGSI entities.

To begin with, the entity *building* models the target building. Several operational and architectonic details of the building are included as attributes on this entity. Examples of information in

4 http://www.sensorcloud.com/.
5 https://grovestreams.com/.



(a) IoTEP workflow for bootstrap stage.



(b) IoTEP workflow for operational stage.

Fig. 3. Platform general workflow.

this section are: opening hours or building use (e.g., company headquarters, university faculty, etc.) but also physical relevant attributes such as fabrics, windows, orientation, and so forth. Moreover, the *spatial region* entity defines the geographic region containing the building. This entity would help to link together buildings located in similar geographic regions that, as a consequence, might share certain energy-related characteristics. The inner structure of a building is represented with the *building space area* entity. This entity gathers the different spatial areas within a building (e.g., classrooms, corridors, halls, landings, etc.). Furthermore, a recursive structure of these areas can be made with their *located at* attribute to represent, for example, that a classroom is inside a *teaching zone*.

This way of introducing data about the buildings and the spaces will made the communication between a Building Information Modeling (BIM) platforms and the IoTEP platform straight forward, what would facilitate the transfer of information among members of a given team.

The second group of entities refers to the energy sensors deployed in the building and the data they collect. This is modeled by means of the *building sensor*, *power meter* and *hvac* entities. Each entity includes the set of attributes monitored by the corresponding energy sensor along with other metadata (e.g., location of the sensor or timestamp of each observation). The *clean* version of these entities refer to the sensor data generated after the data filtering process as described in Section 3.2.2.

The third group of entities focus on representing sensors that are not necessarily within the infrastructure of the building but that may provide useful when collecting energy data. This is the case, for example, of weather stations reporting conditions of the building site. As Fig. 2 shows, this is defined by means of the *external sensor* and *weather conditions* entities.

Finally, only the entities in gray in Fig. 2 have instances stored in ORION and COMET as we will see later.

3.2. Platform architecture

The proposed IoTEP has been structured in four different layers in an incremental approach (this is shown in Fig. 3). In the upcoming sections, a detailed description of each layer is given.

3.2.1. Sensorization layer

This layer is in charge of connecting physical devices or actuators that are going to provide energy data to the platform. Once this is done, it maps the collected data to the NGSI entities of the information model (described in the previous section) and sends the mapped information to the upper homogenization and storage layer.

For the realization of this layer, we have made use of the FIWARE IoT Agent enabler [33]. In a nutshell, this enabler allows to automatically perform the aforementioned data mapping. Different types of this enabler support transport protocols to connect to the physical devices like MQTT.[6] or Lightweight M2M (LwM2M)[7].

Consequently, during the bootstrapping phase of the platform, a set of IoT Agents are configured with the NGSI entity type associated to each of its associated sensor by means of the IoT Agent Manager (see Fig. 3(a)). In particular, power meters deployed in the target building are mapped to the *power meter* entity type whereas HVAC devices are mapped to the *hvac* one. Furthermore, we developed an ad-hoc agent to parse the weather conditions coming

6 http://mqtt.org.

7 http://openmobilealliance.org/iot/lightweight-m2m-lwm2m/.

from an external third-party weather service to the *weather conditions* entity on a regular basis. During the operational phase (see Fig. 3(b)) each time an IoT Agent receives the raw measurements from a physical device, it *inflates* the entity instance associated to the device in upper layer by means of a RESTfull API, in the homogenization and storage layer (will be described in the next section).

3.2.2. Homogenization and storage layer

In this layer, all the collected energy data from the previous layer is conveniently stored in a uniform solution. This way, this layer addresses the heterogeneity of the incoming energy-related data. Moreover, it contains real time data cleaning stage what ensures the quality of the data collected.

Sensor data repository. Regarding the energy-related data storage, this has been achieved by integrating two FIWARE components.

Firstly, ORION context broker [34] implements a publish–subscribe store providing data access by means of the NGSI-10 API [35]. In IoTEP, this enabler stores the entity instances of the information model. By means of the NGSI update operation, IoT Agents in the sensorization layer update the sensor entities' attributes in real time with the new readings from the devices.

Secondly, the COMET enabler [36] is used for supporting access to historic time series data extending the ORION functionality. In that sense, COMET adheres to the same information model, thus, it does not require any further data harmonization process. It incorporates an ad-hoc API to retrieve raw historical sensor data along with several built-in simple aggregation functions over such data (e.g., provide the sum, min or max of the collected observations for a specific time period).

During the bootstrapping phase of the platform, ORION is initiated with the *static* attributes of the entities in the information model (e.g., 'identifier', 'located at' or 'orientation' attributes) and COMET subscribes in ORION to the *dynamic* attributes of the entities to receive each new value (see Fig. 3(a)).

Sensor data cleaning. Concerning the data quality assurance, we developed a data cleaning module to remove the outliers that might be contained in the raw measurements from the sensors. In that sense, outliers have been reported to be the most prominent quality issue of energy data [37,38].

This module had two key requirements. To begin with, the data cleaning process must be done in a timely manner in order to avoid potential bottlenecks. Furthermore, in an IoT ecosystem we should expect a great variety of data formats and structure. Thus, such data cleaning should be done after data homogenization in order to simplify the overall computational cost of the cleaning stage.

In order to cope with the time-processing constraints, we opted for following the Complex Event Processing (CEP) paradigm to develop a real-time data cleaning module. CEP focuses on timely processing streams of information items, so-called events, by filtering, aggregation or pattern discovery using predefined rules following the event–condition–action paradigm [39]. In the present setting, the incoming events are the readings from the energy sensors, the conditions to be detected are whether a reading should be considered or not an outlier and the action of the final insertion of the cleaned data in the storage structure of the platform.

For the outlier definition, we followed a strategy based on quartiles with fences [40]. In brief, such a strategy extracts the median, the lower $\mathcal{Q}_1$ and upper quartiles $\mathcal{Q}_3$ (aka 25th and 75th percentiles) along with the interquartile range $\mathcal{IQ}(=\mathcal{Q}_3-\mathcal{Q}_1)$ of the data set under study. On the basis of such statistics, two fences are defined,

- Lower outer fence: $\mathcal{Q}_1 - 3 \times \mathcal{IQ}$
- Upper outer fence: $\mathcal{Q}_3 + 3 \times \mathcal{IQ}$

This way, a measurement beyond such fences is considered an *extreme outlier*.

The translation of this strategy to CEP allows to calculate such fences incrementally and update their boundaries each time that a sensor pushes in new data. In particular, two types of CEP rules were defined. The first one comprises the rules in charge of computing for each sensor the aforementioned statistics with respect to each of its parameters. For the sake of clarity, the pseudocode of the CEP rule in charge of calculating the fences for power meter sensors is shown here and it looks as it follows:

```
CONDITION PowerMeter.groupBy(id).within(t_int^clean) as A
ACTION    new PowerMeterStats(A.id,
          calculateLowerOuterFence(A.active_energy))
          calculateLowerOuterFance(A.reactive_energy),
          calculateUpperOuterFence(A.active_energy)
          calculateUpperOuterFence(A.reactive_energy))
```

where `groupBy` and `within` are two sliding windows. While `groupBy` splits the stream of power-meter data with respect each particular device, `within` defines a time window to retain the last power-meter data generated during the last t_{int}^{clean} time units. After this, the action part of the rule, generates a new `power meter stats` event comprising the percentiles for each sensor's attribute considering the data included in the time window. It is important to note that this rule would fire each time that new power meter data is injected into the CEP system.

The second set of rules performs the actual extreme outliers detection. Again, there is one rule per sensor type in charge of this task. The pseudocode of the CEP rule to detect the outliers in the power meter data is shown next,

```
CONDITION PowerMeter as A
          AND PowerMeterStats as B
          AND A.id = B.id
          AND A.active_energy ∈ [B.active_energy_lowerFence,
          B.active_energy_upperFence]
          AND A.reactive_energy ∈ [B.reactive_energy_lowerFence,
          B.reactive_energy_upperFence]
ACTION    new CleanPowerMeter(A.id, A.timestamp, A.located_at,
          A.active_energy, A.reactive_energy)
```

Describing it briefly, this rule fires each time that a new power-meter reading is received. The condition part of the rule matches such reading with its associated statistics and checks whether each parameter is contained in its own fences. If that is the case, the reading is considered that has been cleaned. As a result, the action part creates a new `clean power meter` event with the pre-processed data.

A very similar approach is followed for the HVAC data but, this time, using the thermostat temperature attribute of this type of sensor in order to give rise to `clean hvac` events.

The implementation of this CEP mechanism has been made with the Perseo FIWARE enabler [41]. This component incorporates a CEP engine and an SQL-based event processing language to define and execute the CEP rules. Furthermore, it leverages the publish-subscription capabilities of ORION. This way, the engine receives each entity instance, which data has been just updated in ORION, as incoming events; and the cleaned events generated by the rules, automatically update their associated entities in ORION (Fig. 3(b)). Hence, during the bootstrapping phase (see Fig. 3(a)) this component is configured with the rules to be executed and the list of entities in ORION to subscribe (in this case, *power meter* and *hvac* entities).

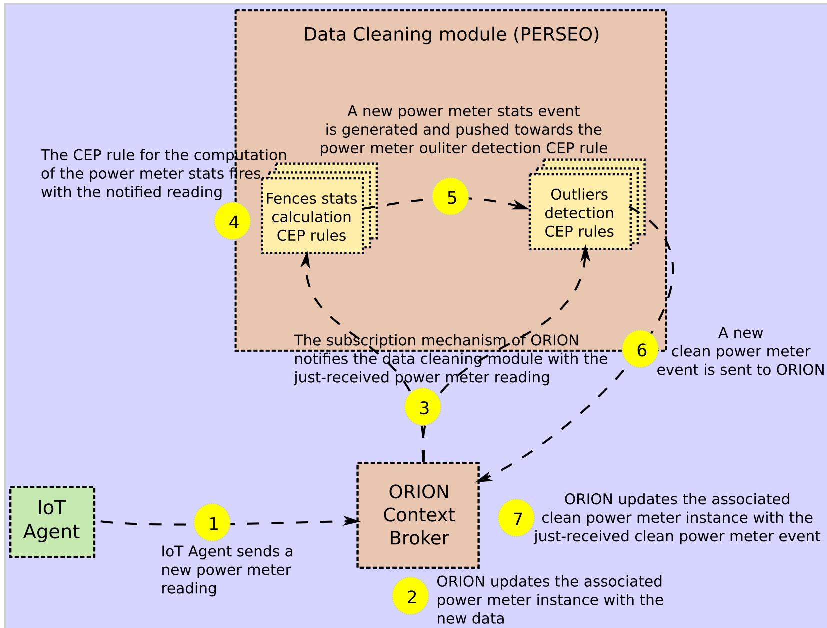


Fig. 4. Workflow of the cleaning of power meter readings.

Finally, Fig. 4 shows an illustrative example of the workflow of the CEP cleaning mechanism and its connection with the sensor data repository. As this figure depicts, each raw sensor reading coming from the IoT Agents is initially stored in ORION by updating its associated *building sensor* instance. In the figure's scenario, a new power-meter reading will update the *power meter* instance representing the sender's sensor (steps 1 and 2 in the figure).

Next, ORION automatically notifies to the data cleaning module the new reading (step 3). This notification fires the two types of CEP rules described before (steps 4 and 5). At the end, the module outcome takes the form of a `clean power meter` event that updates the associated *clean power meter* instance in ORION. This *clean power meter* instance represents the cleaned version of the power meter sensor updated in step 2. Moreover, we should note that all the aforementioned interactions occur following a push-style communication enabling the real-time processing.

3.2.3. Analytics support layer

The third layer of the platform embraces all the functionalities of the platform to provide support for data mining services that can run on top of the platform. In particular, two features have been included in this layer, an energy data volatility detector and a virtual entities generator.

Virtual energy building areas generator (VEBAG). The amount of data that we are able to collect in smart buildings by means of large sensor networks sometimes does not increase the *information volume* because of redundancy. Depending on its nature, this redundancy is treated using different approaches: redundancy detection, data compression, feature extraction, and some others [42].

IoTEP works under the hypothesis that a clever way to reduce the number of variables taking part in the models can not only decrease the computation costs but also increase the accuracy on predictions and classification. In this way, the creation of abstract entities will be justified from the data analytics side, based on the assumption of the existence of this redundancy.

Therefore, the goal of the VEBAG module is the creation of high level entities that preserve as much information as possible in the data set but yet, reducing the volume of it. In this case, we want to create virtual areas comprising several *building space areas*, finding patterns in the energy-related use and defining these virtual areas according to such information to optimize the content of information.

To do so, we aggregate each attribute per energy device daily. This aggregation can be easily done with the built-in RESTful aggregation functions provided by COMET within the homogenization and storage layer. That way, we can represent each device as a time series having one attribute measurement per day and with this, it is possible to find a clustering algorithm that groups every attribute of the time series finding some distinctions between them, like DBSCAN or longitudinal k-means.

Once every device is assigned to a cluster or virtual area, the generator computes the mean of the elements of each cluster to get an average measurement. Finally, each generated cluster is stored in the storage layer as an instance of the *virtual energy area* entity (see Fig. 3(b)). In that sense, this generator is launched on a regular basis or when certain data shifts are detected in the data by the data volatility monitor (described in the next section). Fig. 5 depicts an illustrative example of this process given the building's floor.

Firstly, Fig. 5(a) shows the distribution of room-based building space areas along with their HVACs. It should be recalled that each of these areas and sensors will be stored as different instances in IoTEP. Furthermore, the figure also shows an example of a possible time-series plot of the regulated temperature for each HVAC for illustration purposes.

Next, Fig. 5(b) shows the *virtual energy areas* generated on the basis of the aforementioned temperature time series. As we can see, the six initial room-based building space areas have been merged into three instances of *virtual energy areas* by grouping together the HVACs with similar time series. This way, rooms 4, 5 and 6 and their associated HVACs have been merged into a single area (*virtual energy area 3* in the figure).

All in all, the generation of these virtual energy areas enables the platform to provide multiple views of the energy status of a building. In a low-level setting, we can monitor energy parameters from a single-sensor point of view. Over such simple view, we can also extract energy parameters related to a particular building spatial area (e.g., room, corridor and the like) by simple aggregation

(a) HVACs and room-based building space areas.

(b) Virtual Energy Areas generated based on the HVACs' temperature time-series.

Fig. 5. Example of generation of *virtual energy areas* considering the HVAC temperature in a building floor.

using the *building spatial area* instances. Finally, *virtual energy area* instances enrich the energy awareness by providing an extra layer of perception that is not constrained by the building architectural structure. This way, it is possible to monitor building areas with similar energy behaviors simultaneously.

Data volatility monitor. In order to come up with real energy-aware services, the monitoring of certain energy parameters of a building becomes paramount. This includes detecting either abnormal energy consumption related to building spaces or an abnormal temperature setting related to HVACs.

For that goal, the data volatility monitor focuses on computing the current rate of change of each energy sensor parameter included in the storage layer. This is done in three steps.

Firstly, we extract the historic data set of the target energy parameter for a particular sensor with respect to a pre-defined time period t_{int}^{vol} from COMET. Then, the average rate of change among pairs of consecutive observations of the attribute is computed. Finally, if such averaged value is substantially different that the historic rate of change of that attribute then an alarm is triggered. For the sake of clarity, the pseudo-code of this process is shown in Algorithm 1.

Algorithm 1: Data volatility calculation.

Input: Type, identifier and energy parameter of the monitored sensor ($sensor_{type}$, $sensor_{id}$, $sensor_{attr}$), time interval under study (t_{int}^{vol}) and historic rate of change of the considered parameter for the target sensor (rh_{attr}^{sensor}).
Output: Data volatility alarm, if any.

/* Historic data extraction */
$\mathcal{D} \leftarrow$ get_COMET_raw_historic_data($sensor_{type}$, $sensor_{id}$, $sensor_{attr}$, t_{int}^{vol})
/* Average data-rate change calculation */
$d_{prev} \leftarrow 0$ $r_{avg} \leftarrow 0$ $n \leftarrow 0$
for each $d \in \mathcal{D}$ **do**
  $r \leftarrow |d - d_{prev}|$
  $r_{avg} \leftarrow r_{avg} + \frac{d - r_{avg}}{n}$
  $d_{prev} \leftarrow d$ $n \leftarrow n + 1$
/* Meaningful data-rate change detection */
if $r_{avg} >> rh_{attr}^{sensor}$ **then**
  return data_volatility_alarm($sensor_{type}$, $sensor_{id}$, $sensor_{attr}$, r_{avg})

This alarm is received by the final energy services on top of the platform and the VEBAG module. If this module receives a set of consecutive alarms related to the same energy parameter in a short period of time then it might indicate that the energy similarities in between building areas have changed. In order to capture such shift, VEBAG re-launches the clustering process to reconfigure the virtual areas related to such energy factor. In that sense, this monitor is endlessly executed every t_{int}^{vol} time units in order to keep a continuous control over the sensor data streams.

Finally, we would like to notice that this last mechanism along with the CEP data cleaning described in Section 3.2.2 might provide some clues to building operators about data inconsistencies due to sensor interferences. In particular, the data cleaning module can remove readings that are not consistent with the normal operation of a sensor whereas the data volatility mechanism can also detect abnormal disturbances in the data rate change of a sensor reporting that something unusual is happening.

3.2.4. Service layer

Although not that central when considering the architecture of the platform here developed, the Service Layer is the last level of the IoTEP. This layer serves as interface between the IoTEP and the user, that could be anything from a building services manager to the back end of a smartphone application.

At this level, the data analytics procedures can be invoked and their results visualized. Also, smart-building services that may be the norm when the smart-building paradigm is fully established will be nested at this level of the IoTEP platform, and will allow features such as advanced HVAC predictive control, home automation, fuel poverty evaluation, sick building syndrome diagnostics, risk situations for vulnerable people (as in heat waves), smart tariff strategies, and many others.

4. Validation of the platform

In order to test the feasibility of the proposed platform, IoTEP has been instantiated in a real pilot that allowed us to evaluate functionalities of the new platform. Here we provide some details of the evaluation scenario.

4.1. Pilot description

IoTEP was instantiated at the University of Murcia, Spain. During the last three years, this university has carried out an ambitious plan to monitor and control its buildings' infrastructures distributed across the university premisses. The number of buildings monitored and the automated services have increased quickly in the last years, what serves well the purpose of testing the plasticity of the platform presented in this paper. It should be noted that the sensorization of the buildings at the University of Murcia was done independently of this project, so the fact that the platform was able to allocate the data coming from all the sensors was already a proof of its validity.

In this context, IoTEP was used as the main enabler of an energy efficiency campaign at three cases, namely the Faculty of Chemistry and two multi-disciplinary research and technological transfer centers within the university. Details of the three buildings are provided in Table 1.

Lastly, the evaluation of IoTEP covered a three-month winter campaign from 01/10/2016 to 28/02/2017.

Platform configuration. IoTEP was installed in a centralized server with CentOS 6.7 as operating system, 8 GB RAM and 250 GB hard disk. Besides, Table 2 sums up the configuration of the inner parameters of the platform. It should be reminded that t_{int}^{clean} defines the time interval used by the CEP cleaning mechanism to compose the quartile fences (Section 3.2.2) whereas t_{int}^{vol} indicates the length of the time series considered by the data volatility mechanism to infer meaningful data shifts (see Algorithm 1).

Before the deployment of IoTEP in the pilot, a full covering of energy related variables was done in the buildings under study. After preliminary evaluations, it was discovered that there are three families of data that are fundamental to understand the energy behavior of the building users and heat losses of the envelopes. The three families are: building characteristics, energy streams and building state.

The building characteristics are the physical description of the building. Detailed blueprints of the building were obtained from the department of estates of the university together with detailed plans of constructions. This information together with visual inspections carried out by the members of our team have allowed as to have a rather full description of the condition of the building thermal envelope. With this, it was possible to use building physical models to analyze and predict the heat flows of the building and therefore the energy performance of the fabrics.

About the second family, we were able to monitor in real time with a sampling period of 10 min the operation of more than 200 conditioning units in real time. This included the status of the machines (on/off) and the set point temperatures. It was also possible to obtain the technical characteristics of the machines, what together with the rest of the data allowed us to have a rather accurate proxy of specific power consumption in real time. To contextualize this individual power consumption, the total power consumption of the building was also measured.

Finally, it was needed to know what the conditions on the interior of the spaces of the building were. For this, we monitored in real time the temperature of more than 200 spaces. This temperatures are in accordance with the data taken from the conditioning systems what allowed us to create virtual control volumes/zones in which to evaluate energy flows.

Table 1
Use case building characterization.

	Faculty of chemistry (FC)	Technological transfer center (TTC)	Research center (RC)
Location (coords)	38.02, −1.16	37.72, −1.09	38.02, −1.17
Orientation	south-west	south-west	south-west
Surface area	1500 m^2	3323 m^2	1000 m^2
Floors	6	4	2

Table 2
IoTEP parameters setting.

Parameter	Description	Value
t_{int}^{clean}	Time window length for sensor stream fence calculation	30 days
t_{int}^{vol}	Time period for data volatility calculation	2 hours

Table 3
Information model entities distribution per building.

Entity	Number of instances		
	FC	TTC	RC
Spatial region	1	1	1
Building	1	1	1
Building space area	344	16	10
HVAC	239	0	4
Clean HVAC	239	0	4
Power meter	1	13	4
Clean power meter	1	13	4
Weather conditions	1	1	1

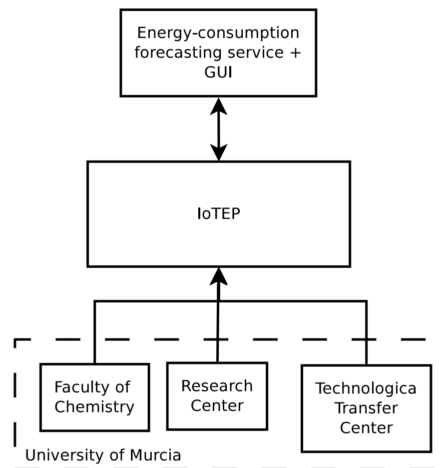


Fig. 6. Representation of the IoTEP pilot evaluation.

The IoTEP was created in such manner that it allows to allocate all this information in two ways: in the form of data stream, and in the form of "static" information. In this way, the description of the building is allocated on the *building* entity previously described. The characteristics of the conditioning system and the data stream can be placed on the *HVAC* and *power meter* entities created for this purpose.

This comprehensive set-up fully monitors the most important energy related aspects of the building, what could be a two-bladed sword. In principle, this allows to do high level reasoning on the data with the high added value that this represents; however, such a large flow of data may render the infrastructure slow and inefficient with such a heterogeneous data. With the solution proposed in this paper we overcome the problems, leading to a platform that, because of the efficient handling of data inherited from FIWARE, allows for the true real time comprehensive data analysis of buildings. With the advantages that this represents.

As a result of this study, Table 3 shows the distribution of instances of the entities of the IoTEP information model stored in ORION per building.

4.2. Pilot objectives

The goal for this testing campaign was to develop a new service able to predict the next-day energy consumption of each of the three buildings, and with this to evaluate the framework we present at all the different levels. However, it should be reminded that this is only an example of the variety of features that could be implemented on IoTEP. The service tested would be instrumental for the department of estates of the university in order to plan energy-saving actions and advanced versions of model predictive control.

As Fig. 6 shows, this service was developed on top of IoTEP i.e. on the service layer shown in Section 3.2.4, by using its functionalities. It was implemented as a web application allowing the control of some of the IoTEP features by the buildings manager to carry on decisions according to data analysis results. Consequently, this application acts as a dashboard that allows users to control the platform and access the aforementioned energy consumption service (see Fig. 7).

In terms of access of the inner features of IoTEP the application includes the following actions,

- Firstly, it is possible to visualize the most recent readings of the HVAC devices per each room of the building. For this feature, the application makes use of the ORION component of the platform.
- Secondly, it is also possible to visualize the HVAC data given a time range defined by the user. For this purpose, the application leverages the raw historic data extraction method of COMET.
- Moreover, this dashboard also allows to control and visualize the results of the *virtual energy areas* generation of the platform (VEBAG module). In that sense, the user can also select the clustering method, and the number of clusters will be selected automatically by the Calinski–Harabasz index.

Finally, the energy consumption prediction service was also integrated in this application. On this way, building managers have full control over all the data analytic process starting from data visualization, aggregation and clustering to the final energy prediction procedure. This integration allows to perform such prediction for several granularity levels targeting from single devices, space areas or *virtual energy areas*. This multi-faced prediction is a key innovation aspect of the application.

For the evaluation of the platform, we studied the suitability and feasibility of the multi-layered view of the energy-related information proposed by IoTEP by means of the *virtual energy areas* generation. Additionally, we also studied the accuracy of the

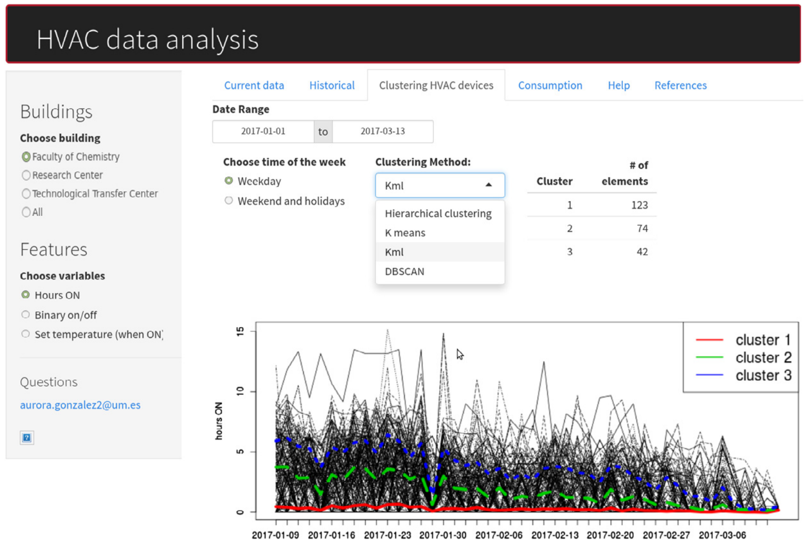


Fig. 7. IoTEP dashboard and energy consumption prediction service.

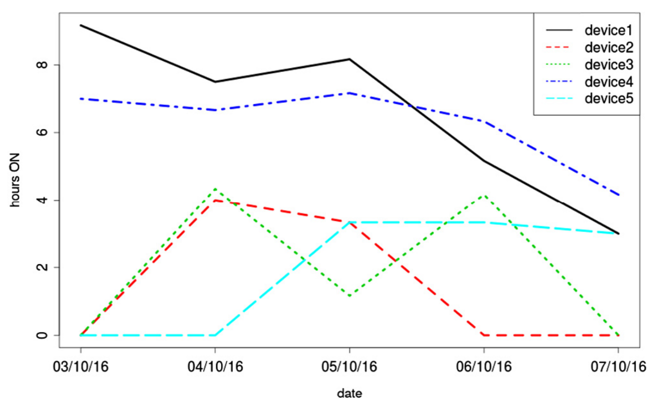


Fig. 8. Time series of 5 HVAC devices.

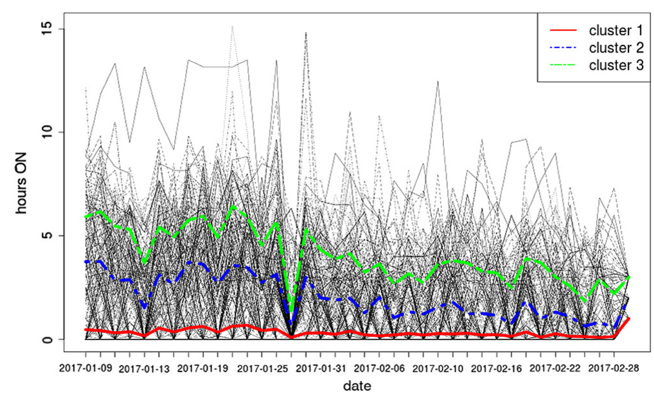


Fig. 9. Cluster evolutions.

energy prediction service when such areas are included as the target entities.

For the generation of these areas, the daily aggregation of data made by the VABAG module was based on counting the hours that each device is tuned on during the day (24 h). As an example, the number of hours that five devices were on during five days is shown in Fig. 8

For the clustering of such aggregated data, we relied on the k-means algorithm [43], but as mentioned before, more algorithms can be used for this purpose. We arbitrarily selected 3 clusters, but a different number can be selected if needed. In Fig. 9 we show the three evolutions of the groups of HVAC within FC that this algorithm identified for working days during the period of study. That way, we found rooms in this building with high use pattern (cluster 3, comprising 47 devices), rooms with little use (cluster 1 with 118 HVACs) and rooms presenting an intermediate frequency of use of the HVAC system (cluster 2 with 74 HVACs). The separation of these clusters could be the first step to an intervention strategy to modify the behavior of big consumers.

In the same way, and looking at the infrastructure level, we represent 239 values taken from the HVAC devices into 3 variables providing a 98.7% reduction of data.

Regarding the energy-prediction service, it makes its prediction according to the previous HVAC grouping within FC. Hence, we compare its performance with the use of the raw data set and in combination with environmental variables. Being the inputs and outputs of the model identified, we followed the next steps [44]: Being the inputs and outputs of the model identified, we followed the next steps [44]:

1. Standardization of inputs
2. Splitting the data into training (75%) and test set (25%)
3. Validation: 10-fold cross validation and 5 repetitions over the training data set using several models: random forest, artificial neural networks and support vector regression.
4. Evaluation: Using the RMSE metric to evaluate the models and its coefficient of variation for comparison.

The scenarios to compare are based on the different inputs to consider:

- "Hours on" average per cluster of the previous day
- Weather predictions from Weather Underground API.[8]
- Raw HVAC data (every HVAC device daily usage)
- Both average per cluster and weather predictions

As we can see in Table 4, with a really reduced number of inputs (only 3 variables), for every model we obtain very good results compared to the others. That way, the use of clusters for creating

8 https://www.wunderground.com/.

Table 4
RMSE (and CV-RMSE) of the different models and inputs.

Model	HVAC clusters	Weather	Raw HVAC	Clust + Weath
RF	0.32 (10.53)	0.513 (17.74)	0.358 (11.83)	0.356 (11.76)
SVM	0.316 (11.03)	0.635 (22)	0.446 (14.76)	0.461 (15.23)
BRNN	0.281 (9.48)	0.423 (14.63)	0.347 (11.47)	0.398 (13.15)
# Inputs	**3**	23	239	26

higher level entities is proved to be useful. Although this is a rather arbitrary method, we prove with this that the platform serves to host algorithms for data analysis and prediction on a very versatile way

Comparative results. In the work [22], CV-RMSE is used in order to validate their results. They are evaluating both aggregated (total) and disaggregated (cooling and ventilating) energy consumption in a daily, weekly and monthly basis. When we compare our results with theirs, we are obtaining 6% less of variance for the RMSE, which is very satisfactory.

In addition, the Recommended Values for Baseline Model from ASHRAE Guideline 14 [45] account for the CV-RMSE smaller than 30% for daily predictions which we reach with ease (our best performance returns a 9.48 %, see Table 4).

To sum up, with this small example we show what can be implemented on the service layer of the IoTEP. With this, we intend to prove how rather complex methods can be implemented on a simple way in our platform. Also, we have shown an example of reducing data volume taking advantage of data redundancy reduction doing clustering. For this specific example we have taken three clusters as an arbitrary number and we have shown that total energy can be predicted with them. This was done as it evaluates all the features of the platform that we show in this paper, but many other applications and examples can be developed following the principles shown in Section 2.

4.3. Lessons learnt

From this first deployment of IoTEP, we can draw up some remarks.

Firstly, the results of the preliminary sensorization study of pilot were easily integrated in the IoTEP information model. This allowed to homogenized all such results in a common format and showed the versatility of the model.

Secondly, the integration of data mining support procedures as part of the platform made possible the easy development of a final service for energy data mining. In that sense, developers only needed to focus on the actual functionality of the service related to the prediction algorithms since other important tasks of the data analysis like data pre-processing or clustering were already provided by the platform.

Finally, the idea of providing a multi-layered view of the energy status of a building by means of clustering techniques has proved its suitability in the energy prediction service in two aspects. From a data-mining point of view, it reduces the redundancy of data and, thus, making up lightweight models. From a more functional point of view, the level of abstraction that the virtual energy areas provide might help building managers to better understand certain energy behaviors within the building.

All in all, this pilot has helped us to confirm that the integration of data analytics support features as part of the IoT platform is currently a key requirement in the energy domain. This enables the development of more sophisticated energy-aware services in a fast-pace process what seem to be the next natural step towards a more efficient energy-literate society.

5. Conclusions

Due to the importance of the building sector in the end-use energy consumption, it becomes a foremost task to achieve meaningful energy savings that will reduce this energy use in reality.

Despite the fact that IoT technologies have been widely used for the realization of the smart building concept, the simple sensorization of buildings is not enough to make a housing stock that consumes fewer energy resources a reality. IoT is also required to properly process, manage and, above all, analyze the energy-related data that would help to develop final energy-aware services targeting the energy efficiency goal.

In this context, several multi-purpose IoT platforms already provide generic solutions to manage IoT data. However, there is a lack of platforms in this field focusing on (1) the household energy domain and (2) providing support for data analytics. As a result, the present work shows an IoT Energy Platform (IoTEP) that covers the two aforementioned needs by following an open approach based on FIWARE enablers. IoTEP provides several functionalities oriented to the data analytics domain like the CEP data cleaning module or the times series storage along with functionalities for the correct energy management like the data volatility monitoring or the virtual energy areas detector that will allow with personalized energy feedback for the improvement of energy behavior.

Lastly, the platform has been instantiated in a real use case having a large energy sensor network. In that sense, one of the key novelties of IoTEP is that the virtual areas detection has proved to be of great help when it comes to develop an end-use energy prediction service over the platform, but many other services could be implemented with trivial computational effort under this paradigm.

Regarding further work, IoTEP has been developed re-using several open source components that are orchestrated following lightweight RESTfull calls what allows other scientists and engineers to contribute to this platform, opening the door to crowd sourced development. Consequently, new modules and enablers can be smoothly integrated in the existing architecture. In that sense, the integration of other types of sensing approaches beyond mote-class sensors, like crowdsensing, it foreseen as future actions in the platform. This would allow to capture and analyze other forms of human behavior also relevant for the building energy domain.

Acknowledgments

This paper has been made possible thanks to the support of the European Commission through the H2020-ENTROPY-649849, the Spanish National Project CI-CYT EDISON (TIN2014-52099-R) and MINECO TIN2014-52099-R project (grant BES-2015-071956) granted by the Ministry of Economy and Competitiveness of Spain (including ERDF support). Ramallo-González would like to thank the program Saavedra Fajardo (grantnumber220035/SF/16) funded by Consejería de Educación y Universidades of CARM, via Fundación Séneca-Agencia de Ciencia y Tecnología de la Región de Murcia.

References

[1] Odyssee Mure project, Energy efficiency trends in buildings in the EU, lessons from the ODYSSEE MURE project, 2012. http://www.odyssee-mure.eu/publications/br/energy-efficiency-in-buildings.html.

[2] Buildings Energy Data Book, Buildings share of U.S. primary energy consumption stats. http://buildingsdatabook.eren.doe.gov/TableView.aspx?table=1.1.3.

[3] GeSI, GeSI smarter2020 report. http://gesi.org/SMARTer2020.

[4] D. Coley, T. Kershaw, M. Eames, A comparison of structural and behavioural adaptations to future proofing buildings against higher temperatures, Build. Environ. 55 (2012) 159–166. http://dx.doi.org/10.1016/j.buildenv.2011.12.011. http://www.sciencedirect.com/science/article/pii/S0360132311004276.

[5] Project iNSPiRe - Development of Systemic Packages for Deep Energy Renovation of Residential and Tertiary Buildings including Envelope and Systems, 2014. http://inspirefp7.eu.

[6] M. Vellei, S. Natarajan, B. Biri, J. Padget, I. Walker, The effect of real-time context-aware feedback on occupants heating behaviour and thermal adaptation, Energy Build. 123 (2016) 179–191. http://dx.doi.org/10.1016/j.enbuild.2016.03.045. http://www.sciencedirect.com/science/article/pii/S0378778816301992.

[7] J.A. Stankovic, Research directions for the internet of things, IEEE Internet of Things J. 1 (1) (2014) 3–9. http://dx.doi.org/10.1109/JIOT.2014.2312291.

[8] S. Darby, Smart metering: What potential for householder engagement? Build. Res. Inf. 38 (5) (2010) 442–457. http://dx.doi.org/10.1080/09613218.2010.492660.

[9] V. Desley, B. Laurie, M. Peter, The effectiveness of energy feedback for conservation and peak demand: A literature review, Open J. Energy Effic. 2013 (2013).

[10] M.V. Moreno, A.F. Skarmeta, L. Dufour, D. Genoud, A.J. Jara, Exploiting IoT-based sensed data in smart buildings to model its energy consumption, 2015 IEEE International Conference on Communications, ICC, 2015, pp. 698–703. http://dx.doi.org/10.1109/ICC.2015.7248403.

[11] N. Simcock, S. MacGregor, P. Catney, A. Dobson, M. Ormerod, Z. Robinson, S. Ross, S. Royston, S.M. Hall, Factors influencing perceptions of domestic energy information: Content, source and process, Energy Policy 65 (2014) 455–464. http://dx.doi.org/10.1016/j.enpol.2013.10.038. http://www.sciencedirect.com/science/article/pii/S0301421513010604.

[12] M. Zdravković, M. Trajanović, J. Sarraipa, R. Jardim-Gonçalves, M. Lezoche, A. Aubry, H. Panetto, Survey of Internet-of-Things platforms, in: 6th International Conference on Information Society and Techology, ICIST 2016, Vol 1, ISBN: 978-86-85525-18-6, 2016, pp. 216–220, Kopaonik, Serbia. https://hal.archives-ouvertes.fr/hal-01298141.

[13] M. Molina-Solana, M. Ros, M.D. Ruiz, J. Gmez-Romero, M. Martin-Bautista, Data science for building energy management: A review, Renew. Sustain. Energy Rev. 70 (2017) 598–609. http://dx.doi.org/10.1016/j.rser.2016.11.132. http://www.sciencedirect.com/science/article/pii/S1364032116308814.

[14] J. Mineraud, O. Mazhelis, X. Su, S. Tarkoma, A gap analysis of internet-of-things platforms, Comput. Commun. 89 (2016) 5–16.

[15] K. Zhou, C. Fu, S. Yang, Big data driven smart energy management: From big data to big insights, Renew. Sustain. Energy Rev. 56 (2016) 215–225. http://dx.doi.org/10.1016/j.rser.2015.11.050. http://www.sciencedirect.com/science/article/pii/S1364032115013179.

[16] Y. Simmhan, S. Aman, A. Kumbhare, R. Liu, S. Stevens, Q. Zhou, V. Prasanna, Cloud-based software platform for big data analytics in smart grids, Comput. Sci. Eng. 15 (4) (2013) 38–47. http://dx.doi.org/10.1109/MCSE.2013.39.

[17] A. Kumbhare, Y. Simmhan, V. Prasanna, Cryptonite: A secure and performant data repository on public clouds, 2012 IEEE Fifth International Conference on Cloud Computing, 2012, pp. 510–517. http://dx.doi.org/10.1109/CLOUD.2012.109.

[18] A. Ishii, T. Suzumura, Elastic stream computing with clouds, 2011 IEEE 4th International Conference on Cloud Computing, 2011, pp. 195–202. http://dx.doi.org/10.1109/CLOUD.2011.11.

[19] N. Vastardis, M. Kampouridis, K. Yang, A user behaviour-driven smart-home gateway for energy management, J. Ambient Intell. Smart Environ. 8 (6) (2016) 583–602.

[20] A. Ożadowicz, J. Grela, An event-driven building energy management system enabling active demand side management, in: Event-Based Control, Communication, and Signal Processing, EBCCSP, 2016 Second International Conference on, IEEE, 2016, pp. 1–8.

[21] L. Klein, J.-y. Kwak, G. Kavulya, F. Jazizadeh, B. Becerik-Gerber, P. Varakantham, M. Tambe, Coordinating occupant behavior for building energy and comfort management using multi-agent systems, Autom. Constr. 22 (2012) 525–536.

[22] N. Li, J.-y. Kwak, B. Becerik-Gerber, M. Tambe, Predicting HVAC energy consumption in commercial buildings using multiagent systems, Proceedings of the 30th International Symposium on Automation and Robotics in Construction and Mining, ISARC, 2013.

[23] C. Alcaraz, I. Agudo, D. Nunez, J. Lopez, Managing incidents in smart grids a la cloud, 2011 IEEE Third International Conference on Cloud Computing Technology and Science, 2011, pp. 527–531. http://dx.doi.org/10.1109/CloudCom.2011.79.

[24] M.V. Moreno, B. beda, A.F. Skarmeta, M.A. Zamora, How can we tackle energy efficiency in iot basedsmart buildings? Sensors 14 (6) (2014) 9582–9614. http://dx.doi.org/10.3390/s140609582. http://www.mdpi.com/1424-8220/14/6/9582.

[25] A. Botta, W. de Donato, V. Persico, A. Pescap, Integration of cloud computing and Internet of Things: A survey, Future Gener. Comput. Syst. 56 (2016) 684–700. http://dx.doi.org/10.1016/j.future.2015.09.021. http://www.sciencedirect.com/science/article/pii/S0167739X15003015.

[26] J. Zhou, Z. Cao, X. Dong, A.V. Vasilakos, Security and privacy for cloud-based IoT: Challenges, IEEE Commun. Mag. 55 (1) (2017) 26–33. http://dx.doi.org/10.1109/MCOM.2017.1600363CM.

[27] T. Zahariadis, A. Papadakis, F. Alvarez, J. Gonzalez, F. Lopez, F. Facca, Y. Al-Hazmi, FIWARE Lab: Managing resources and services in a cloud federation supporting future internet applications, IEEE/ACM 7th International Conference on Utility and Cloud Computing, 2014, pp. 792–799. http://dx.doi.org/10.1109/UCC.2014.129.

[28] E. Kovacs, M. Bauer, J. Kim, J. Yun, F. Le Gall, M. Zhao, Standards-based worldwide semantic interoperability for IoT, IEEE Commun. Mag. 41 (2016).

[29] S. Sotiriadis, E.G.M. Petrakis, S. Covaci, P. Zampognaro, E. Georga, C. Thuemmler,An architecture for designing Future Internet (FI) applications in sensitive domains: Expressing the software to data paradigm by utilizing hybrid cloud technology, 13th IEEE International Conference on BioInformatics and Bio-Engineering 2013, pp. 1–6. http://dx.doi.org/10.1109/BIBE.2013.6701578.

[30] F. Ramparany, F.G. Marquez, J. Soriano, T. Elsaleh, Handling smart environment devices, data and services at the semantic level with the FI-WARE core platform, 2014 IEEE International Conference on Big Data, Big Data, 2014, pp. 14–20. http://dx.doi.org/10.1109/BigData.2014.7004417.

[31] A. Preventis, K. Stravoskoufos, S. Sotiriadis, E.G.M. Petrakis, Personalized motion sensor driven gesture recognition in the FIWARE cloud platform, 14th International Symposium on Parallel and Distributed Computing, 2015, pp. 19–26. http://dx.doi.org/10.1109/ISPDC.2015.10.

[32] P. Fernndez, J.M. Santana, S. Ortega, A. Trujillo, J.P. Surez, C. Domnguez, J. Santana, A. Snchez, Smartport: A platform for sensor data monitoring in a seaport based on fiware, Sensors 16 (3) (2016). http://dx.doi.org/10.3390/s16030417. http://www.mdpi.com/1424-8220/16/3/417.

[33] Telefonica I+D, IoT Agent documentation, 2017. http://fiware-iot-stack.readthedocs.io..

[34] Telefonica I+D, ORION context broker documentation, 2017. http://fiware-orion.readthedocs.io.

[35] Open Mobile Alliance (OMA) Specification, NGSI Context Management, 2010. http://www.openmobilealliance.org/release/NGSI/V1_0-20100803-C/OMA-TS-NGSI_Context_Management-V1_0-20100803-C.pdf.

[36] Telefonica I+D, COMET documentation, 2017. http://fiware-sth-comet.readthedocs.io..

[37] W. Chen, K. Zhou, S. Yang, C. Wu, Data quality of electricity consumption data in a smart grid environment, Renew. Sustain. Energy Rev. 75 (2017) 98–105. http://dx.doi.org/10.1016/j.rser.2016.10.054. http://www.sciencedirect.com/science/article/pii/S1364032116307109.

[38] A. Ramallo-González, New method to reconstruct building environmental data, in: Buildign Simulation International Conference BS2015, University of Bath, 2015.

[39] O. Etzion, P. Niblett, Event processing in action, first ed., Manning Publications Co., Greenwich, CT, USA, 2010.

[40] NIST/SEMATECH, e-Handbook of Statistical Methods, 2012. http://www.itl.nist.gov/div898/handbook/.

[41] Telefonica I+D, PERSEO official repository, 2017. https://github.com/telefonicaid/perseo-fe..

[42] H. Hu, Y. Wen, T.S. Chua, X. Li, Toward scalable systems for big data analytics: A technology tutorial, IEEE Access 2 (2014) 652–687. http://dx.doi.org/10.1109/ACCESS.2014.2332453.

[43] C. Genolini, B. Falissard, KmL: k-means for longitudinal data, Comput. Stat. 25 (2) (2010) 317–328. http://dx.doi.org/10.1007/s00180-009-0178-4.

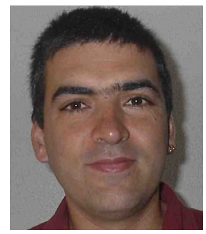
[44] A. Gonzlez-Vidal, V. Moreno-Cano, F. Terroso-Senz, A.F. Skarmeta, Towards energy efficiency smart buildings models based on intelligent data analytics, Procedia Computer Science 83 (2016) 994–999. http://dx.doi.org/10.1016/j.procs.2016.04.213. The 7th International Conference on Ambient Systems, Networks and Technologies, ANT 2016 / The 6th International Conference on Sustainable Energy Information Technology, SEIT-2016 / Affiliated Workshops. http://www.sciencedirect.com/science/article/pii/S1877050916302460.

[45] ASHRAE, Measurement of Energy and Demand Savings, ASHRAE, 2002.

Fernando Terroso-Sáenz graduated from the University of Murcia with a degree in Computer science in 2006. He also received the master's degree in Computer Science at the same university in 2010. Since 2009, he has been working as a grant student in the Department of Information Engineering and Communications of the University of Murcia where he has published several papers in national and international conference proceedings. His research interests include complex event processing, ubiquitous computing and fuzzy modeling.

Aurora Gonzalez Vidal graduated in Mathematics from the University of Murcia in 2014. In 2015 she got a fellowship to work in the Statistical Division of the Research Support Service, where she specialized in Statistics and Data Analysis. In 2015, she started her Ph.D. studies in Computer Science, focusing her research on Data Analysis for Energy Efficiency and studied a Master in Big Data. Her research covers machine learning, data mining, and time series segmentation.

Ramallo-González completed his Ph.D. in Building Physics at the University of Exeter with a scholarship from the Wates Foundation. He has worked as post-doctoral researcher on two EPSRC funded projects in the department of Architecture and Civil Engineer of the University of Bath. Currently he is a Savedra-Fajardo Research Fellow in the Faculty of Computer Science at the University of Murcia, and PI of the project ThermaSim.



Antonio F. Gómez-Skarmeta received the MS degree in Computer Science from the University of Granada and BS (Hons.) and the Ph.D. degree in Computer Science from the University of Murcia. He is a Full Professor in the same Department and University. He has worked on different research projects at regional, national and especially at the European level in areas related to advanced services like multicast, multihoming, security and adaptive multimedia applications in IP and NGN networks.