# Enhancing energy awareness through the analysis of thermal energy consumption

Andrea Acquaviva[†], Daniele Apiletti[†], Antonio Attanasio[† §], Elena Baralis[†],
Federico Boni Castagnetti[‡], Tania Cerquitelli[†], Silvia Chiusano[†], Enrico Macii[†],
Dario Martellacci[‡], Edoardo Patti[†]
[†] Dipartimento di Automatica e Informatica, Politecnico di Torino, ITALY
[§] Istituto Superiore Mario Boella, Torino, ITALY – [‡] IREN Energia Torino, ITALY
[†] {name.surname}@polito.it – [§]attanasio@ismb.it – [‡]{name.surname}@gruppoiren.it

## ABSTRACT

Energy efficiency by means of reduction in wasteful energy consumption is a growing policy priority for many countries. Innovative systems should be designed to continuously monitor a smart city environment and provide all stakeholders the tools to improve energy efficiency. This paper presents the EDEN platform, designed to collect and analyze thermal energy consumption of residential and public building heating systems. EDEN is being deployed in a major Italian city and collects energy consumption measurements through an extensive smart metering grid involving thousands of buildings. EDEN also collects and analyzes indoor climate conditions, and user feedbacks, such as their thermal comfort perception, by means of an ad-hoc social network. Collected data are further enriched with temporal and spatial information at different abstraction levels and meteorological data available as an open source data set. Several technical Key Performance Indicators (KPIs) have been defined to inform users on their building thermal energy consumption, while user-friendly KPIs present energy savings or over-consumptions in an informative fashion.

## 1. INTRODUCTION

In the last few years, the interest in urban data computing is continuously growing both in the industrial and research domains, as well as in the Public Administration. Industries are attracted by the business opportunities arising from the design, implementation, and exploitation of novel technologies and applications to effectively support all the crucial aspects of Smart Cities management. Researchers, instead, are interested in the challenging issues coming from the application of innovative data management and mining techniques to new and more complex fields. Innovative systems should be designed to continuously monitor a smart city environment and suggest new ways to improve the quality of life within an urban environment, for both citizens and the Public Administration. A complete overview of the key challenges of urban computing from the computer scientists perspective is presented in [25]. Among the large variety of applications available in the context of smart cities, this paper focuses on energy consumption, and specifically on thermal energy consumption in buildings during the winter period. The goal is to improve energy infrastructures and reduce energy consumption, and the associated costs, by suggesting energy-saving strategies to users and by providing better information to the different people involved in the energy management roles.

Energy efficiency is a growing policy priority for many countries around the world, as governments seek to reduce wasteful energy consumption and encourage the use of renewable sources. The International Energy Agency (IEA) has estimated that in terms of primary energy consumption, buildings represent roughly 40% of total final energy consumption in most countries. The amount of this energy used for heating and cooling systems is about 60% in the residential sector and 45% in the service one [12].

Important research activities have been carried out to use database management systems and exploratory data mining techniques in the field of storage and analysis of energy data to evaluate the efficiency of buildings. The proliferation of sensor networks for monitoring indoor and outdoor environmental parameters [16, 19] has brought to the facility managers huge archives of measures with temporal and spatial references. Research contributions on these large data volumes have been carried out for: (i) supporting data visualization and warning notification [17, 20, 24]; (ii) efficient storing and retrieval operations based on NoSQL databases [19, 23]; (iii) discovering anomalous behaviors using clustering algorithms [6, 24], Support Vector Machines (SVM) [9] and outlier detection [21]; (iv) characterizing consumption profiles among different users [2, 9, 20]; identifying the main factors that increase energy consumption (e.g., floors and room orientation [10], location [9, 14]).

In this paper we describe the Energy Data ENgagment platform, EDEN, designed to monitor and analyze thermal energy consumption of heating systems for enhancing user energy awareness. EDEN collects data from smart meters deployed in thousands of buildings in Turin, a major Italian city. EDEN also collects and analyzes indoor climate conditions by means of temperature sensors installed in a subset of the monitored buildings. Thermal comfort perception and user feedbacks on indoor climate conditions are also collected by means of an ad-hoc social network. Collected data are further enriched with temporal and spatial information at different abstraction levels, and meteorological data available as an open source data set. Several technical and user-friendly Key Performance Indicators (KPIs) are defined within EDEN targeting different users. A *technical KPI* informs users on their building thermal energy consumptions, while a *user-friendly KPI* explains monetary savings or overconsumption by converting its value into the price of commonly purchased goods. EDEN is designed, developed and experimented within the context of a publicly-funded research project, including both academic and industrial partners that contribute to make it a live platform, with actual deployment and real data.

This paper is organized as follows. Section 2 discusses our vision towards enhancing energy awareness through the Energy Data ENgagment platform. Section 3 describes the main building blocks of the proposed system. For some blocks, we describe our first implementation to show both the feasibility and high potential of the proposed approach. Section 4 reports a preliminary analysis of thermal energy consumption for 2 school buildings and 6 residential buildings located in Turin. Section 5 draws conclusions and presents future developments of this work.

## 2. PLATFORM OVERVIEW

Figure 1 shows the overall architecture of the EDEN system. In this study we focus on an instance of EDEN tailored to an indoor heating monitoring system. However, the EDEN architecture can be easily tailored to different indoor monitoring contexts, such as electric cooling, and outdoor monitoring applications as well. It includes three main components, named *Data Platform*, *Publication Platform*, and *Social Platform*, briefly described below and detailed in the following sections.

EDEN is designed for the collection, storage, modeling, and analysis of a large amount of heterogeneous data to provide different levels of relevant knowledge. The aim is to make people aware of their energy and thermal consumptions, as well as encouraging them to pursue energy saving strategies. Collected data include energy consumption logs provided by thermal smart meters and indoor climate conditions monitored through indoor temperature sensors. In addition, data on the user thermal comfort perception of indoor climate conditions and user feedbacks are gathered through an ad-hoc social network. Heterogeneity in terms of formats, timings and sampling periods, and sources presented a challenge to the designers, also considering the changes over time of this factors, determined by contexts (e.g., smart meters update) or design improvements. To this aim, EDEN exploits a non-relational schema-free data warehouse, which allows coping with frequent changes in data formats without technological issues. This component will be detailed in Section 3.3.

Energy consumption data are collected by means of a large number of smart meters (4,000 as of December 2014) deployed in Turin (Italy) by IREN [13] to monitor thermal energy for district heating. IREN is a multi-utility company listed on the Italian Stock Exchange and operates in the sectors of electricity, thermal energy for district heating, gas, management of integrated water services as well as the collection and disposal of waste.

Data on energy consumption and on monitored indoor climate conditions, collected through sensors and smart meters, are stored in the *Data Platform* component. These data are enriched with spatial and temporal information at different granularity levels as well as with various meteorological conditions. The enriched dataset is stored in a datawarehouse and is managed by the *Publication Platform* component. Specifically, an informative dashboard is generated based on a selection of Key Performance Indicators (KPIs) to produce useful feedbacks to the different users and suggests ready-to-implement energy efficient actions or strategies. Mainly, the following two classes of KPIs have been proposed. (i) *Technical KPIs* allow informing users on the thermal energy consumption of their building, but also comparing the consumption between buildings in the same neighborhood, also in different time periods. Comparison can be performed under similar meteorological conditions. (ii) *Informative and user-friendly KPIs* present the results of the analysis on energy savings and overconsumption in an informative fashion, using simple and easily understandable comparisons according to the user profile. For example, let us consider the energy consumption of a secondary school, and suppose that we would like to improve students' energy awareness. An informative and user-friendly KPI can provide the school energy savings in terms of energy needed for heating the gym for a given number of days. Alternatively, it can explain the possible monetary savings in terms of commonly purchased goods (e.g. average number of pizzas that could have been purchased by saving on energy consumption).

The *Social Platform* component is a digital and social platform which will be developed as a social network where users can share their feedbacks and their perceptions of indoor thermal comfort (e.g. too hot, too cold, or comfortable). Furthermore, it provides visibility of both technical and informative KPIs. The aim is ehnancing energy awareness and stimulate sustainable behaviors to optimize energy consumption.

The EDEN platform also includes the knowledge extraction block for discovering interesting associations among thermal energy consumptions, indoor climate conditions, meteorological conditions, and user perception of indoor thermal comfort in the form of association rules [1]. Association rules represent a powerful exploratory data mining approach able to discover interesting and hidden correlations in the data.

Finally, a subset of interesting and open data (e.g., KPI values) will be published in the *Smart Data Platform* to improve both individual and collective energy awareness. The Smart Data Platform exploited in EDEN is the Yucca Smart Data Net [18] developed by the Piedmont Region (Italy).

## 3. PLATFORM COMPONENTS

In this section we describe the main components of the proposed EDEN platform, which are currently under development.

### 3.1 Data platform

Remote measurements of energy consumption are collected by IREN [13], an Italian energy-provider company, by means of gateway boxes installed in monitored buildings. Each gateway includes a GPRS modem with an embedded programmable ARM CPU. An ad-hoc software has been developed to execute the following activities: sensor management, GPRS communication, remote software update, data collection scheduling, and collected data sending to a remote server.

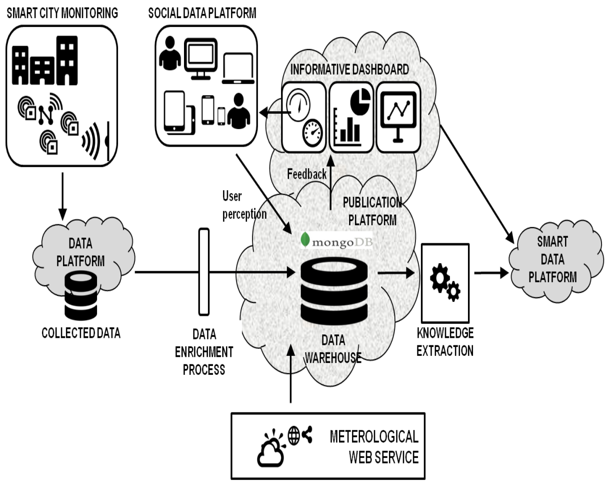


**Figure 1: The** EDEN **system architecture.**

Each gateway has in charge the management of all the sensors deployed in its building. Thermal energy is measured under different aspects, such as instantaneous power, cumulative energy consumption, water flow and corresponding temperatures. Furthermore, gateways also collect indoor temperature and the status of the heating system.

A cloud architecture is used for storing and processing all the monitored data. As of December 2014, there are about 4 thousands monitored buildings, each generating about 2,000 data frames per day. Thus, EDEN needs to manage a growing base of at least 8 million data frames per day. The gateways send the data frame to the cloud architecture, where a firewall first authenticates the data sender and then assigns each data frame to one of four dispatchers to guarantee the system reliability. Each dispatcher delivers the frame to a cluster of computers including different processing servers where data are stored in an HDFS distributed file system. The dispatcher is able to recognize if the process server has stored the frame correctly and in that case it sends the ACK to the gateway which can send the next data frame.

Each processing server elaborates the received data and stores the result in an Oracle database. The logical model of the database includes the following three tables: (i)The *Building* table contains the main features characterizing each building such as address and volume; (ii) the *Sensor* table stores the list of sensors located in each building and the main characteristics for each sensor (e.g., unit of measure, description, sensor type and model, etc.); (ii) the *History* table stores the collected measurements for all sensors. On average, every 5 minutes a data frame is received from each building. Then, corresponding data are stored in many records, with one record for each measurement value.

To efficiently perform the management of a large volume of collected data, different strategies have been adopted (e.g., data sharding, distributed map-reduces, and data replication).

## 3.2 Data integration and enrichment

Data collected through the smart meters are aggregated and enriched with additional contextual information acquired from external open data sources. More specifically, to analyze the *temporal distribution* of thermal energy consumption, the following time granularities are considered: day, month, 2-month, 3-month, 6-month time periods. Moreover, each day is classified as holiday or not, and the measurement time is aggregated into the corresponding *daily time slot* (morning, afternoon, evening, or night).

In Italy, heating systems are operated only from October 15th to April 14th, hence times periods outside this range were not considered. In addition, since the heating systems under monitoring within EDEN are operated at fixed time slots, each aggregation (morning, afternoon, evening) includes only the time slots when the system is actually on (e.g., morning from 6:00a.m. to 11:59a.m, afternoon from 12:00p.m to 6:59p.m., evening from 7:00p.m to 10:00p.m.).

To analyze the *spatial distribution* of thermal energy consumption, different space granularities are also considered beyond the building addresses. In addition, each *address* is mapped to the corresponding geographical *coordinates* (longitude and latitude degrees), *neighborhood* and *city district* including that neighborhood. While the address is an information recorded for the monitored building, the geographical coordinates and both the neighborhood and district names corresponding to the address are added as additional contextual features to the repository. We exploited the Google Maps APIs [11] for geocoding street addresses. Furthermore, topological information about neighborhood names and districts are integrated in the repository as well. The latter have been retrieved from [22]. Topologies are used to graphically analyze the most significant spatial trends in thermal energy consumption data and were encoded in GeoJSON, which is a standard format for encoding a variety of geographic data structures.

The above data were also enriched with meteorological information collected from the web. Specifically, historical meteorological data were taken from the Weather Underground web service, which gathers data from Personal Weather Stations (PWS) registered by users. For the city of Turin more than twenty PWS are distributed throughout the territory and about 4 of them are directly located inside the area considered in this study. The decision to use data from PWS is motivated by the fact that they reflect with high accuracy the real conditions registered in their neighborhood, as opposed to other services that provide estimated values with respect to a wider area. Although the measurement frequency can be easily set by the user for each PWS (and can vary over time), the average value for the ones we considered was about 5 minutes. Data were collected for the period going from October 2012 to April 2013. More specifically, each measurement includes the air temperature (expressed in degree Celsius), the relative humidity (percentage), the precipitation level (mm), the wind speed (km/h) and the sea level atmospheric pressure (hPa). The date and time of each measurement is also included.

## 3.3 Data warehouse

While the data collection from smart meters exploits an Oracle database, due to the fixed and constant nature of those measurements, enriched data is much more variable and heterogeneous, and its analysis requires a different technological solution. To this aim, enriched data are modeled into a document-oriented distributed data warehouse providing rich queries, full indexing, data replication, horizontal scalability and a flexible aggregation framework, including a distributed map-reduce engine. The current database em-

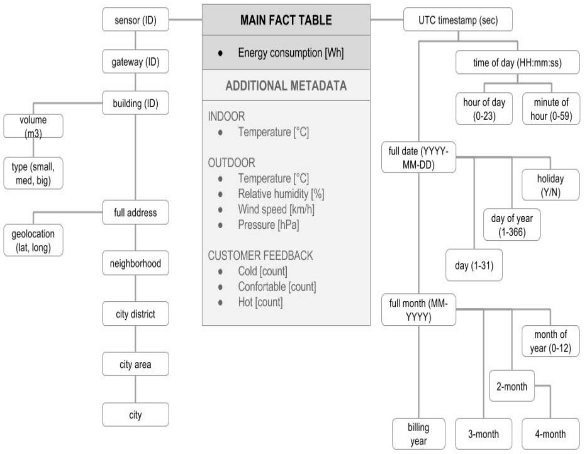


**Figure 2: The** EDEN **data warehouse design.**

powering EDEN analytics inside the Publication Platform is MongoDB [7], and to our purpose it is actually exploited as a data warehouse: periodically, sensor-collected data and social-platform data are enriched, integrated and loaded into a MongoDB collection.

Following best practices in data warehouse design, data are de-normalized and redundant information is added to each record (document) to speedup read performance by avoiding join operations (which are not sopported by MongoDB), and resulting in fast querying and KPI computation. The model design aims at providing a human-readable document format, hence the choice of long, self-descriptive field names, with sub-documents for each separate aspect of the record, from user feedbacks to geo-location, through smart meter measurements and other contextual informations. Such structured choice helps in coping with heterogeneity, but presents a main drawback in disk space usage: each field name is re-written within each document, together with all the redundant information that enrich the measurement. However, the low cost of disk space nowadays makes it an acceptable issue, also considering that no image or video data are currently included.

In Figure 2 the data warehouse conceptual model is presented: the fact table consists of a main measure, the energy consumption in a 5-minute time period, and some additional metadata coming from indoor sensors, outdoor PWSs, and the social data platform collecting customer feedbacks. Two hierarchies are defined: a time-related hierarchy and a place-related one. The former provides many different blends of time spans, from minutes to months and years. The latter starts from the physical sensors inside each monitored building and builds up to the whole city, with the building volume and the geolocation coordinates as related features included in the document.

Some metadata, in particular weather data and customer feedbacks, may require some additional pre-processing during the data loading phase because of different time spans: e.g., a customer feedback given at a certain point in time may be considered valid for a longer period than the specific 5-minute of a single data warehouse document, and weather data may be unavailable for a specific point in space. The solution adopted in EDEN supposes that customer feedback in terms of indoor environment comfort has a temporal validity of 30 minutes, which is distributed from 15 minutes before the feedback is provided and 15 minutes after. Hence, a customer reporting a very cold indoor environment at midnight, is associated with 5-minute documents from 23:45 (included) to 00:15 (excluded). Weather data associated with a specific building and address are computed as a distance-based weighted mean of the values provided by the three nearest PWSs. The weight is inversely proportional to the distance from the PWS to the building location, hence three equally distant PWSs would have the same weight in determining the outdoor values of a given building.

In the following, a sample MongoDB document from the designed data warehouse is provided. Subdocuments have been extensively used to group similar fields together. Some fields deem special attention:

- The customer feedback fields that identify too cold, too hot and comfortable indoor environments are the count of the collected feedbacks in the 30-minute time span as previously described.
- The customer comments are a list of text strings provided as status description on the social data platform; this allows us to exploit text mining techniques to associate keywords to measurement values, by building upon the text search features of MongoDB. This issues will be addressed as a future development of the current implementation.
- The billing period spans over two different years: October-November is the first 2-month (billing and operational) period and so the December-January 2-month period spans two calendar years, hence the choice to be verbose and use values such as ‘2-2014-2015‘.

```
{
    _id: ObjectId(...),
    energy_consumption: 0.12,
    indoor: {
        temperature: 21.2
},
    outdoor: {
        temperature: 15.6,
        relative_humidity: 70.0,
        wind_speed: 5.0
    },
    feedback: {
        cold: 2,
        comfortable: 12,
        hot: 1,
        comments: ["nice sunny winter day",...]
    },
    place: {
        sensor:  {id: 123456, model:"..."},
        gateway: {id: 234567, model:"..."},
        building: {
            id: 345678,
            volume: 1234,
            type: "med"
        },
        address: {
            full: "corso Castelfidardo 39, 10129, ...",
            street_name: "Castelfidardo",
            street_number: "39",
```

```
            coordinates: [7.6600778, 45.0632518],
            ...
        },
        neighborhood: "Crocetta",
        city_district: "Circoscrizione I",
        city: "Torino"
    },
    time: {
        UTC_timestamp: 1419266446.0,
        day: {
            time: "16:40:46",
            minute: 40,
            hour: 16,
            slot: "afternoon"
        },
        date: {
            full: "2014-12-22",
            day: 22,
            day_of_year: 356,
            holiday: "N"
        }
        month: "12-2014",
        month_of_year: 12,
        2month: "2-2014-2015",
        3month: "1-2014-2015",
        4month: "1-2014-2015",
        billing_year: "2014-2015"
    }
}
```

Finally, the data model design addresses horizontal scalability and replication choices.

Horizontal scalability is obtained by exploiting data sharding, i.e., storing documents across multiple distributed machines by dividing the collection and distributing its data over multiple servers, or shards. As the size of the data increases, EDEN only needs to add more machines to scale and support the demand of a higher number of read and write operations. Each shard processes fewer operations as the cluster grows, and the amount of data that each server needs to store is reduced.

MongoDB provides automatic sharding and the key design choice is the attribute whose values partition the collection documents, i.e., the shard key. In EDEN the sharding is performed using a hash-based partitioning on the value of the building ID field. The shard key choice is motivated by KPIs that are typically computed by grouping measurements per building, and the number of buildings grows with the expansion of the EDEN framework, hence it is a natural scaling indicator. Hash-based partitioning has been chosen over the range-based partitioning approach to ensure that data are evenly distributed across the machines in the cluster, since no range queries are performed on the building ID.

Replication is obtained by exploiting MongoDB replica sets to provide redundancy and high availability. With multiple copies of data on different servers, replication avoids data loss from a single server failure. Currently, in EDEN each replica set consists of a primary server, a secondary server and an arbiter. All writes go to the primary server, while the secondary server can be exploited to increase the read capacity at the cost of possible inconsistency. However, this is not an issue in EDEN since KPIs for the dashboards can wait to be updated after the secondary has caught up the updates from the primary, which usually happens within seconds.

## 3.4 KPIs definition

The EDEN system performs the KPI analysis tailored to different users to gain insights on the integrated data. In Business Intelligence, the analysis of Key Performance Indicators (KPIs) is an established methodology [15]. KPIs help organizations define and measure progress towards organizational goals by monitoring the most significant achievements. In our context, KPIs are quantitative indicators of thermal energy consumptions. To apply KPI analyses to data coming from a real scenario, we defined *technical KPIs* and *informative and user-friendly KPIs*. The aim of KPI generation is to produce useful feedbacks to enhance energy awareness for different types of users. We identified four different operational roles representing users of the EDEN system: (i) the *Energy Manager* is responsible for the energy services provided. He/She needs to access summary and high-level information in order to grasp the overall picture of the energy situation of the city district under observation. He/She requires dashboards showing KPIs at a higher level of granularity (e.g., city district). (ii) The *Energy Analyst* is an expert in energy consumption. He/She is interested in analyzing the complete streams of collected data to observe and understand the observed phenomenon, analyze the different components and identify possible causes. He/She needs to inspect a significant volume of data to understand the anomaly. (iii) The *Consumer* represents the building condos administrators or the public administration (as in the case of public schools), whose interest is to assess the efficiency of the heating system, as well as to get a feeling of virtuous behaviors that should/could lead to energy savings while maintaining the desired level of indoor confort. He/She only needs to visualize a few indicators, possibly presented in a clear and intuitive way. (iv) The *Users living in the building* are interested in mantaining indoor wellness and understand how their behaviors affect energy consumption and they can achieve a significant reduction of their energy expenditure. Presented data should be informative and, at the same time, easy to understand.

For users living in the building we define two user-friendly KPIs that measure virtuous behaviors (i.e. energy savings) in terms of (i) energy needed for heating the given building for a given number of days, or (ii) kilograms of bread or number of pizzas that can be purchased with the savings.

The technical KPIs aims at evaluating the energy consumption at different levels: from the single building to the neighborhood, and from hours and days to months. In EDEN four technical KPIs have been identified.

- Building KPI. Average energy consumption indicator of the building per unit of volume, i.e., total energy consumption of the building divided by the building total volume. This KPI can be also normalized according to the degree days and to the known indoor temperature.
- Neighborhood KPI. Average energy consumption indicator of the buildings in the same neighborhood per unit of volume.
- Building-type KPI. Average energy consumption indicator of the buildings of the same type and in the same neighborhood per unit of volume.

- Climate KPI. Average energy consumption indicator of the buildings of the same type and in the same neighborhood per unit of volume, considering only energy consumption during specific outdoor conditions (temperature range).

These KPIs are computed on different time scales, in particular: hourly, for each daily time slots, daily, monthly, and on N-month periods.

Rich queries, indexing and map-reduces are the data warehouse features exploited to compute KPIs. Specifically, fields frequently used by KPI queries such as building IDs are indexed, and map reduces are exploited to perform KPI computation. Let consider a simple KPI such as the first of the list, and for the sake of simplicity, suppose the temporal scope and normalizations are removed (their implications will be discussed later). The equivalent SQL query to extract the Building KPI would be as follows.

```
select sum(energy_consumption)/building_volume
from fact_table, dimension_table1, ...
where <join fact and dimension tables>
group by building_id, building_volume
```

In EDEN such KPI is computed by exploiting map, reduce and finalize functions of MongoDB, as follows. The map function determines the key and value pairs emitted by each processed document: the key is similar to the group by SQL clause, and in this case it corresponds to the building ID, whereas the value is a more complex object, since to compute an average we need to carry over both operands, the consumption sum and the building volume. Hence, we put these two values into the value object returned (emitted) by the map function.

```
function() {
    key = this.place.building.id;
    value = {
        ec: this.energy_consumption,
        vol: this.place.building.volume
    };
    emit(key, value);
}
```

The reduce function receives a list of values from the map functions having the same key, hence we have a list of objects containing the energy consumption (*ec*) and the building volume (*vol*), and we need to sum all the *ec* values of the list. The building volume is the same for all values, since they refer to the same building (the building id is the map reduce key).

```
function(key, values) {
    reduced_value = {
        ec: 0,
        vol: values[0].vol,
    };
    for (var i=0; i<values.length; i++) {
        reduced_value.ec += values[i].ec;
    }
    return reduced_value;
}
```

After the reduce phase we have a list of value objects, one for each building id (key), containing the total energy consumption and the building volume. The finalize function adds to each object in this list the average value, which is the final result and corresponds to the desired KPI.

```
function (key, value) {
    value.ec_vol = value.ec/value.vol;
    return value;
};
```

The provided example is computed over the whole collection and return total cumulative results since the beginning of the data collection. The temporal scope can be introduced by exploiting two approaches: (i) a specific query filtering undesired time periods can be passed to the map reduce MongoDB command, thus limiting the computation to a specific time span, or (ii) a more complex key can be used involving compound building ID and time periods. The latter is particularly useful to save pre-aggregated results in a collection similarly to materialized views. For instance a simple compound map-reduce key such as the concatenation of the building ID and the date (YYYY-MM-DD) of the measurement would automatically provide day-level aggregations and would require a small change in the map function only. In EDEN then, monthly KPIs are computed directly by querying the daily KPIs collections, hence building a tree of map-reduces that are fed by lower-level lesser-aggregated results and feed higher-level map-reduces in the tree.

Current advantages of the map-reduce KPI approach include a natively distributed computation, that allows horizontal scaling and load balancing among the nodes of the MongoDB cluster. We are currently analyzing further improvements on the EDEN KPI computation framework that include incremental map-reduces, which are an obvious approach due to the nature of the data loading, and the exploitation of the MongoDB aggregation framework. Furthermore, the ability to add new fields to the documents allow us to easily implement new KPI computations as they are required, even if the database does not natively support join operations. Indeed, the actual join is performed as a preprocessing step during the data enriching phase.

Finally, MongoDB also provides native support for geospatial querying, that is exploited in EDEN to compute KPIs involving the neighborhood besides the administrative boundaries. For instance, to query all the measurements associated with buildings within a given distance from a specific point in space, the following snippet of code can be added to an existent query.

```
{
   'address.coordinates': {
      $geoWithin: {
        $center: [ [7.6600778, 45.0632518], 0.01]
      }
   }
}
```

This limits the results to the measurements in a radius of approximately 0.01 degrees (roughly 1 km) from the point at the given longitude and latitude coordinates.

### 3.5 Smart data platform

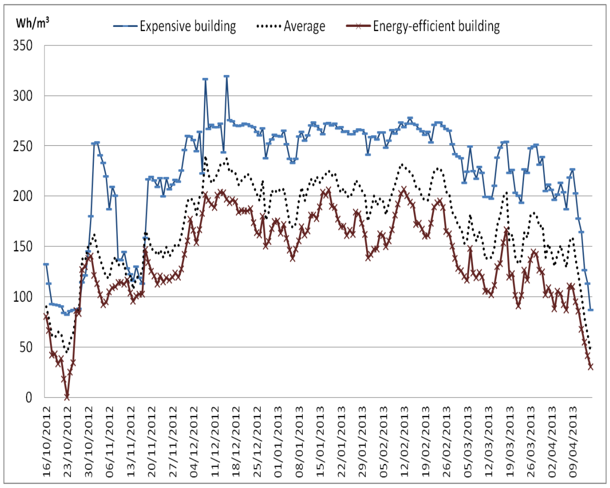


**Figure 3: Residential buildings: Daily energy consumption per unit of volume ($Wh/m^3$) .**

The EDEN system will publish a subset of collected data and results of the analysis in the Yucca Smart Data Platform (SDP) [18]. Specifically, a portion of the data showed to users through the informative dashboard, a subset of user's feedbacks and indoor thermal comfort perception data, and interesting knowledge items extracted from the enriched data collection. The Yucca SDP is a Big Data store developed and maintained by CSI Piemonte [8]. Based on the Open Data paradigm, it gives individuals and organizations the opportunity to publicly share their data under a license that allows anyone to freely use them. It enables the interconnection of geographically distributed applications, social networks, objects and systems. The Yucca SDP supports different protocols to receive and send data, such as HTTP, MQTT, RTSP, WebSocket and OData REST APIs. It also provides some basic functionalities of data enrichment, aggregation, filtering, pattern matching and windowing.

# 4. EXPERIMENTAL RESULTS

We performed a preliminary analysis of energy consumption on a real dataset using the EDEN platform. We considered 2 school buildings and 6 residential buildings, all located in two neighboring districts in Turin, within a circular area of 1 km of radius. Values were measured roughly every 5 minutes. The full time period depends on the availability of measurements for each building. For the residential buildings, measures are available from 2012 to 2014. To consider a complete winter period we analyzed the period from October $15^{th}$, 2012 to April $14^{th}$, 2013. For the first school (named school A), instead, the time period is from November $28^{th}$, 2013 to April $30^{th}$, 2014. For the second school (named school B), it is from October $1^{st}$, 2012 to March $14^{th}$, 2013.

Firstly, the daily energy consumption per unit of volume ($Wh/m^3$) has been computed for each residential building, together with the daily average consumption among all buildings. Figure 3 shows the average consumption profile, and the profiles of an expensive building and an efficient one.

Since the time periods available for the two school buildings are different, also in duration, a further processing has been performed to compare their energy efficiency: the consumption has been normalized with respect to the total degree days measured for the same time length. This measure represents the different external temperatures that influence the daily energetic demand for heating. We computed the total degree days as the sum of all the positive differences between a reference indoor temperature (i.e., 20 °Celsius) and the average daily temperature taken from the ARPA weather archives [3]. Results are reported in Table 1. As shown in Table 1, the daily energy consumption in school B is much greater than in school A, with a difference of about 254 kWh. However, a higher value of average degree days can also be observed (12.37 °C of school B versus 10.97 °C of school A). The last row in Table 1 shows the energy consumption per unit of volume divided by the total degree days. The total consumption normalized with respect to the degree days is still higher, but the difference is much smaller. In fact, if we had 1690 degree days for school B (like in school A), the total energy consumption per unit of volume unit would have been only 31.04 $Wh/(m^3 \times °C) \times$ 1690 °C= 52458 $Wh/m^3$, rather than 63336 $Wh/m^3$, which is much closer to the 50373 $Wh/m^3$ of school A.

# 5. CONCLUSIONS AND FUTURE WORKS

This paper presented a preliminary implementation of the EDEN platform to enhance energy awareness. As of December 2014, IREN has installed thousands of thermal smart meters in buildings in Turin, a major Italian city. EDEN components and design choices that led to the Data Platform and the Publication Platform have been discussed, with the aim of efficiently collect and analyze data on energy consumption. The Data Platform collects all the monitored data, while the Publication Platform includes pre-processed data enriched with spatial and temporal information at different abstraction levels, as well as meteorological data available in open source datasets. We also designed and implemented different technical and user-friendly KPIs to provide informative dashboards targeting different users.

We are currently implementing an ad-hoc social platform where users are proactively engaged in the act of generating data related to their perception of thermal comfort, as well as useful feedbacks on thermal energy consumption of the buildings where they live or work. The social platform will also show to users both technical and user-friendly KPIs on energy consumptions (savings or overconsumption) in an informative fashion.

Since the collected data easily scale towards very large datasets, the problem of discovering interesting and hidden correlations for these huge data collections becomes challenging. We are currently designing an innovative scalable algorithmtailored to enriched data managed by EDEN to efficiently perform the association rule mining on a huge energy consumption dataset [5, 4].

# 6. ACKNOWLEDGMENTS

The research leading to these results has partially received funding from the Piedmont Region under the POR FESR 2007/2013 n. 281-79 (EDEN Project).

The authors wish to thank colleagues and other partners involved in the EDEN project (i.e., CSP Innovazione nel ICT, COMMITWORLD S.r.l., Capetti Elettronica s.r.l., Consorzio TOP-IX, Experientia, Sisvel Technology s.r.l.) for their advices and fruitful discussions.

| | School A | School B |
|---|---|---|
| Volume [m$^3$] | 4480 | 4480 |
| Time period | 11/28/2013 – 04/30/2014 | 10/01/2012 – 14/03/2013 |
| Total energy consumption per unit of volume [Whm$^3$] | 50,373 | 63,336 |
| Daily energy consumption [Wh] | 1,465,390 | 1,719,658 |
| Daily consumption per unit of volume [Wh/m$^3$] | 327.10 | 383.85 |
| Average degree days [ °C] | 10.97 | 12.37 |
| Total degree days (in the given time period) [°C] | 1690 | 2040.4 |
| Total normalized consumption [Wh/(m$^3$×°C)] | 29.81 | 31.04 |

**Table 1: School buildings: Energy consumption normalized per unit of volume and degree days**

## 7. REFERENCES


[1] R. Agrawal, T. Imielinski, and Swami. Mining association rules between sets of items in large databases. In *ACM SIGMOD 1993*, pages 207–216, 1993.

[2] O. Ardakanian, N. Koochakzadeh, R. P. Singh, L. Golab, and S. Keshav. Computing electricity consumption profiles from household smart meter data. In *EDBT/ICDT Workshops'14*, pages 140–147, 2014.

[3] ARPA. Piedmont Region. *Regional Agency for the Protection of the Environment. Available at http://www.arpa.piemonte.it/english-version Last access: December 2014.*

[4] E. Baralis, T. Cerquitelli, S. Chiusano, and A. Grand. P-mine: Parallel itemset mining on large datasets. In *Workshops Proceedings of the 29th IEEE International Conference on Data Engineering*, pages 266–271, 2013.

[5] E. Baralis, T. Cerquitelli, S. Chiusano, and A. Grand. Scalable out-of-core itemset mining. *Inf. Sci.*, 293:146–162, 2015.

[6] E. Baralis, T. Cerquitelli, and V. D'Elia. Modeling a sensor network by means of clustering. In *18th International Workshop on Database and Expert Systems Applications*, pages 177–181, 2007.

[7] K. Chodorow and M. Dirolf. *MongoDB: the definitive guide*. O'Reilly Media, 2010.

[8] CSI website. *http://www.csipiemonte.it/. Last access on December 2014.*

[9] S. Depuru, L. Wang, V. Devabhaktuni, and P. Nelapati. A hybrid neural network model and encoding technique for enhanced classification of energy consumption data. In *Power and Energy Society General Meeting*, pages 1–8, 2011.

[10] C. Filippín and S. F. Larsen. Analysis of energy consumption patterns in multi-family housing in a moderate cold climate. *Energy Policy*, 37(9):3489 – 3501, 2009.

[11] Google Maps. *Available at http:/maps.google.it Last access: December 2014.*

[12] IEA. Energy efficiency indicators. 2014.

[13] IREN website. *http://www.gruppoiren.it/index.asp. Last access on December 2014.*

[14] S. R. Iyer, V. Sarangan, A. Vasan, and A. Sivasubramaniam. Watts in the basket?: Energy analysis of a retail chain. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, pages 4:1–4:8. ACM, 2013.

[15] R. Kimball and M. Ross. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. John Wiley & Sons, Inc., 2nd edition, 2002.

[16] D. Kriksciuniene, T. Pitner, A. Kucera, and V. Sakalauskas. Sensor network analytics for intelligent facility management. In *Proceedings of the 6th International Conference on Intelligent Interactive Multimedia Systems and Services*, pages 212 –221, Amsterdam, 2013. IOS Press.

[17] D. Kriksciuniene, T. Pitner, A. Kucera, and V. Sakalauskas. Data analysis in the intelligent building environment. *International Journal of Computer Science and Applications*, Volume 11, 2014.

[18] Piedmont Region. ITALY. *SMARTDATANET. Available at http://www.smartdatanet.it/ Last access: December 2014.*

[19] L. Pitt, P. Green, and B. Lennox. A sensor network for predicting and maintaining occupant comfort. In *Environmental Energy and Structural Monitoring Systems*, pages 1–6, 2013.

[20] M. Santamouris, G. Mihalakakou, P. Patargias, N. Gaitani, K. Sfakianaki, M. Papaglastra, C. Pavlou, P. Doukas, E. Primikiri, V. Geros, M. Assimakopoulos, R. Mitoula, and S. Zerefos. Using intelligent clustering techniques to classify the energy performance of school buildings. *Energy and Buildings*, 39(1):45 – 51, 2007.

[21] J. E. Seem. Using intelligent data analysis to detect abnormal energy consumption in buildings. *Energy and Buildings*, 39(1):52 – 58, 2007.

[22] Turin GeoPortal. *Available at http://www.comune.torino.it/geoportale/ Last access: December 2014.*

[23] J. van der Veen, B. van der Waaij, and R. Meijer. Sensor data storage performance: SQL or NoSQL, physical or virtual. In *Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on*, pages 431–438, June 2012.

[24] D. Wijayasekara, O. Linda, M. Manic, and C. Rieger. Mining building energy management system data using fuzzy anomaly detection and linguistic descriptions. *Industrial Informatics, IEEE Transactions on*, 10(3):1829–1840, Aug 2014.

[25] Y. Zheng, L. Capra, O. Wolfson, and H. Yang. Urban computing: Concepts, methodologies, and applications. *ACM Trans. Intell. Syst. Technol.*, 5(3):38:1–38:55, Sept. 2014.