NRG4CAST
FP7-2012-NMP-ENV-ENERGY-ICT-EeB
Contract no.: 600074
www.nrg4cast.org

NRG4Cast

Deliverable D5.1

Decision Support System and Reasoning Framework

Editor:	Maja Škrjanc
Author(s):	Maja Škrjanc, Klemen Kenda, Marko Grobelnik, Dunja Mladenić JSI
Deliverable Nature:	Report (R)
Dissemination Level: (Confidentiality)[1]	Public (PU)
Contractual Delivery Date:	November 2014
Actual Delivery Date:	November 2014
Suggested Readers:	Energy managers, energy analysts, data mining and forecasting experts, data analysts
Version:	1.3
Keywords:	Predictive analytics, reasoning techniques, energy management framework, data mining, forecasting, information enrichment

[1] Please indicate the dissemination level using one of the following codes:
• **PU =** Public • **PP =** Restricted to other programme participants (including the Commission Services) • **RE =** Restricted to a group specified by the consortium (including the Commission Services) • **CO =** Confidential, only for members of the consortium (including the Commission Services) • **Restreint UE =** Classified with the classification level "Restreint UE" according to Commission Decision 2001/844 and amendments • **Confidentiel UE =** Classified with the mention of the classification level "Confidentiel UE" according to Commission Decision 2001/844 and amendments • **Secret UE =** Classified with the mention of the classification level "Secret UE" according to Commission Decision 2001/844 and amendments

Executive Summary

In this report, we introduce the IDDS framework for energy management systems - MSDA (Multimodal Stream Data Analytics) framework, which was developed within the NRG4Cast project and will be implemented as a NRG4Cast platform. It provides a framework for the dynamic prediction of energy demand, energy production and detection of complex events by employing various reasoning approaches.

MSDA follows a hybrid approach by combining both knowledge-driven and data-driven elements. The use of a hybrid approach ensured that the development of IDSS followed business drivers with technological support in order to provide appropriate enriched information to decision makers in order to help them in making better-informed decisions.

MSDA is based on a number of selected advanced analytical techniques, which provide information enrichment in several dimensions such as various types of predictions, complex event detection and information extraction. It combined multimodal data such as data streams from various types of sensors, textual information and social networks. Enriched information represents the input for selected reasoning techniques such as case-based reasoning and machine learning models (Neural Networks, Hoeffding Trees, SVM, hierarchical clustering, Linear Regression).

MSDA represent a flexible generic framework for energy management systems, which support different scenarios from energy management of different buildings, public lighting systems, energy production and electrical car scenarios based on reasoning and machine learning methods and techniques, which will be customized, combined and integrated into a single framework within the NRG4Cast project.

Following the *Introduction* section, the second section presents the MSDA framework and offers insight into logical layers of Information enrichment. Applications of advanced analytical methods are grouped by their level of new knowledge generation for the energy management domain. The third section adds a different view point of the NRG4Cast contribution while describing, in brief, the NRG4Cast architecture and maps information enrichment layers into the NRG4Cast architecture. It includes mapping of NRG4Cast previous and future deliverables onto the NRG4Cast architecture. We conclude the report with a summary and plans for future work.

Table of Contents

Executive Summary 3
Table of Contents 4
List of Figures 5
Abbreviations 6
1 Introduction 7
2 MSDA – Energy management IDSS Framework 9
2.1 Information enrichment layers 10
2.1.1 Information Enrichment layer: Multimodal data integration and fusion 11
2.1.2 Information Enrichment layer: Forecasting of consumption, production and price 13
2.1.3 Information Enrichment layer: Complex event processing 15
2.1.4 Information Enrichment layer: Knowledge generation 16
3 Architecture of the NRG4Cast Platform 19
4 Summary and future work 22
References 23

List of Figures

Figure 1: Layers of Information enrichment in MSDA DSS framework 9
Figure 2: MSDA framework 10
Figure 3: Multimodal data integration and fusion 12
Figure 4: Set of prediction models for single data source (consumption sensor) for two different prediction windows (15 min and 1 day) 14
Figure 5: Example of RCA analysis for IREN pilot- thermal plan at Reggio nell' Emilia 17
Figure 6: Architecture of the NRG4Cast Platform. 19

Abbreviations

WP	Work package, as stated in DoW
Q/A	Question/Answers
DM	Data Mining
IDSS	Intelligent Decision Support System
CEP	Complex Event Processing
RCA	Route Cause Analysis
CBR	Case Based Reasoning
NN	Neural Network
SVM	Support Vector Machine
LR	Linear regression
HT	Hoeffding Tree

1 Introduction

Energy production, sale, distribution and final consumption of energy are obvious stages in the energy management chain. Each stage of it exposes opportunities for implementation of energy efficiency procedures with a clear goal of selecting appropriate actions to accomplish the selected target of a 20% cut in Europe's annual primary energy consumption by 2020 [1,2]. The EC has proposed several measures to increase efficiency at all stages. In order to identify appropriate actions in different environments and at various complexity levels (city district, buildings, lighting systems), the key factor for successful decision making process is the capability of making decisions based on:

- correct/trusted information,
- content rich information,
- information provided at the appropriate time and
- presented in an appropriate manner for decision makers to absorb the information.

During the past years, we are facing a rapid development of analytically advanced methods applied on data streams to provide various capabilities such as prediction based on data streams, detecting complex events, triggering alarms and generating new knowledge from large volumes of data. There is a wide range of available solutions, which offer a wide range of approaches with applications based on data mining techniques [3, 4] and which are not so focused on specific details. Quite a lot of publications offer insight into the application of conventional data mining approaches to the energy management domain [5,6], but do not offer many specifics, which are crucial for successful implementation. On the other hand, many focused solutions exist (for example [7, 8,9,10,11,12]) that offer very specific solutions with very specific foci.
However, when designing such a system, one can encounter a number of different challenges both from the technical and business perspectives. Some of the main technical challenges include:

- the ability to handle vast amounts of data,
- extraction of new relevant knowledge and information from the multi-modal data sources:
    - which include sensor measurements of different phenomena (energy consumption, signal quality, weather data, energy prices),
    - news and tweets from the World Wide Web,
    - pieces of metadata relevant for a particular type of object (working time, schedules, energy certificate, etc.).

From the business perspective, development of the energy management framework must take into account requirements of different stakeholders in energy efficiency management. In order to provide valid information for each stakeholder, we need to provide, integrate and analyse information from different parts of the energy chain. Such information and appropriate dataneed to be merged and enriched in a sensible way, so that the framework can offer improvement of services for each stakeholder:

- For producers, it can yield an estimation (prediction) of power demand for different time windows and time frames.
- For different types of consumers, it can predict consumption for different time frames, taking into account different important features like weather.
- For energy distributors, it can optimize energy distribution as well as improve the current models on predicting energy prices in European energy spot markets.

Energy management framework should include an appropriate system architecture, which supports the above mentioned analysis and overcome technical challenges (data fusion, alignment, integration). It enables the development and analysis of various profiles of different consumers or prosumers:

- consumption profiles of different public buildings (offices, hospitals, schools),
- consumption profiles of residential buildings,
- operational profiles of public lighting systems,
- prediction and energy profiles in electric cars (or drivers) and profiles of battery charging.

With new knowledge, extracted from various consumption profiles interested stakeholders can extend their set of business actions, which can be used for introduction of new measures for energy savings or simply for acquiring more precise information on their consumption, which gives them better negotiating position in the energy market, when signing or extending contracts with energy distributors.

One of the challenges in energy management domain in IT community is to find the way of proper selection and integration of data and stream mining techniques to perform effective forecasting. Some of the solutions are working on preselected techniques like genetic algorithms (MOGA) and variations of Neural Networks [13] , while other approaches from industry [14] shows attempts to cover wide range of scenarios, including energy trading and also offers their services as web services.
Having scanned the energy management domain and usage of machine learning techniques, NRG4Cast project can offer services and solution, which could be well positioned in the market. A well will be interesting to see comparison of final results with other applied approaches.

This report is presenting NRG4Cast proposal for energy management IDSS framework and its implementation stage in NRG4Cast project.

2 MSDA – Energy management IDSS Framework

In the NRG4Cast project, we developed a generic DSS framework for management of energy production/consumption systems – **MSDA framework (Multimodal Stream Data Analytics).** It combines approaches from knowledge discovery in database and advanced analytical technologies, such as Data and Stream Mining, in developing a framework for intelligent decision support systems (IDSS). The aim of the proposed framework is to support decision making by recalling past information (history data), inducing "chunks" of domain knowledge from this information (information enrichment) and performing reasoning upon this knowledge in order to reach better-informed conclusions in different business situations.

The MSDA framework supports specific needs for monitoring, predicting, knowledge extraction and diverse reporting capabilities in the energy management domain for a number of different stakeholders: energy producers, energy distributors, big complex energy consumers (municipality, university campus, industrial complex), smaller private consumers and "prosumers" – electrical cars, which could be seen either as consumers of electricity or energy storage ("battery") for electricity. It allows the calibration for specific needs of different pilots and can also be tailored for other energy networks, such as gas distribution or water management networks.

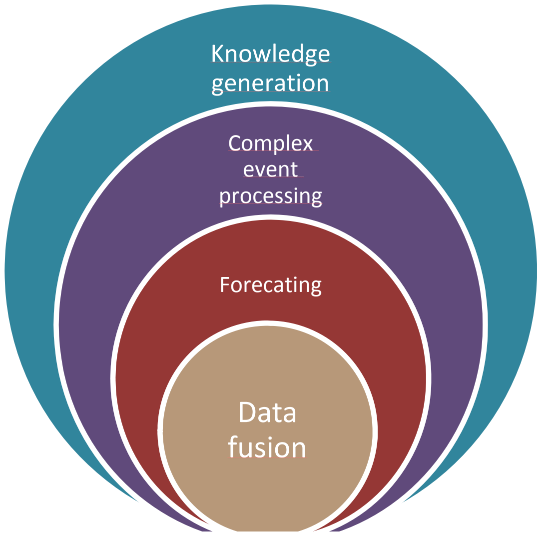

Figure 1: Layers of information enrichment in the MSDA DSS framework

It is designed in a way that allows flexibility at the level of information gathering (we can register various sensors or external sources) and presentation (reports on different summarization levels). MSDA framework offers a flexible data management environment, which can be calibrated for specific business scenario needs. NRG4Cast project covers very different business scenarios: energy production, energy management for public lighting systems, energy management of consumption on various types of consumers – electrical cars, buildings, campus. MSDA framework overcomes this challenge of analysing information at different summarization levels from various energy management scenarios, with the introduction of a flexible hierarchy (defined by the users) of consumption points – a point of measuring selected quantities. MSDA framework requires that an expert user first defines single **consumption points** corresponding to metadata, and later creates so-called **consumption centres** – which are defined by the users and are basically custom made groups of consumption points. Consumption points represent the most detailed level of an object for which we can measure, analyse and predict energy consumption. These objects are also described/presented in the framework with associated metadata (energy certificates for the buildings, size, purpose, etc.). For each of the consumption points and consumption centres, the MSDA framework enables various steps in information enrichment, as shown in Figure 1 , with respect to the available data. Information enrichment layers represent logical layers of applied advanced analytical techniques, which enable users to get correct and enriched information at the appropriate time in order to make better decisions, which are based on factual information.

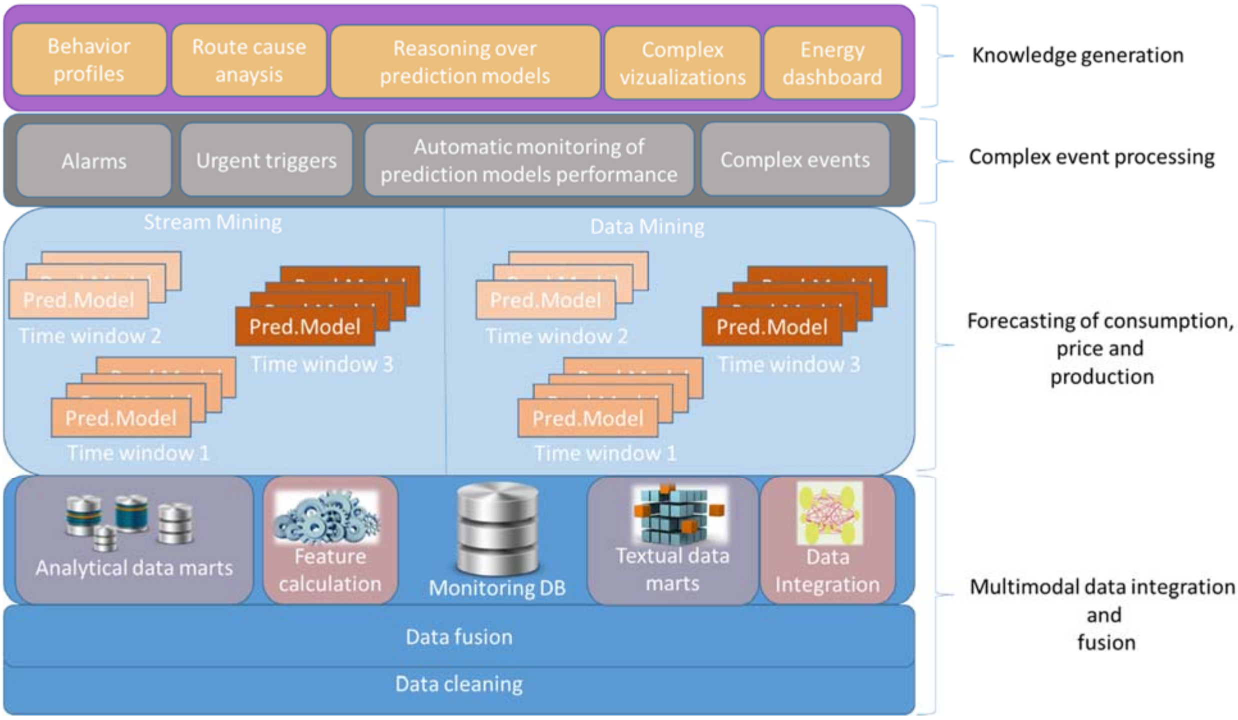

Figure 2: MSDA framework

While data models and the NRG4Cast system architecture supports generic functionalities for the energy management system as well as specific functionalities for different pilot scenarios, the MSDA framework presents generic logical layers of information enrichment with associated advanced analytics techniques. Figure 2 shows what type of functionalities each of the information enrichment layers contains. The information enrichment layers are described in more details in the following subsections.

2.1 Information enrichment layers

The MSDA framework includes the following information enrichment layers (see Figure 1):

- multimodal data integration and fusion,
- forecasting of consumption, production and price,
- complex event processing and
- new knowledge generation.

Each of the layers requires the application of appropriate advanced analytics techniques with respect to different data modalities (structured data, unstructured data, data streams) and also on different summarization levels. This is a challenging task since each of the techniques has its own specifics in terms of data preparation and result interpretation. The analytical results (various predictions, triggers) need to be properly interpreted and presented in a seamless manner so as to ensure that the final results should provide integrated, meaningful and valuable information to decision makers.

2.1.1 Information Enrichment layer: Multimodal data integration and fusion

The first level of Information Enrichment includes several basic functionalities, including data cleaning and integration of external data sources, such as weather services, in order to provide information about the environment. Along with the metadata regarding different objects, additional data is gathered to provide information from the environment of the object, like building occupancy, working hours, timetables.
It consists of the following steps:

- data cleaning (analytical approach using Kalman filters),
- data fusion (align data on the common denominator (recalculate values on a selected timestamp)),
- data modelling phase (integrate, merge and resample data from different sources),
- data enrichment (feature creation and calculation): holidays, day of the week, occupancy, working hours, etc.).

Each of the steps includes several different functionalities or substeps and therefore consists of several different building blocks, which are described in more detail in deliverables D2.2, D2.3, D2.4 and D3.1.

The data cleaning step is mostly focused on data streams from different types of sensors (consumption, weather, weather forecasts) and includes analytical techniques to correct the data.
The data fusion step includes functionalities to recalculate sensor measurements to a common denominator – the same timestamp and provide various types of aggregates. All the stream data sources for a single object need to be recalculated into the same data frequency in order to enable comparisons and feature calculations, which are crucial for later prediction/forecasting phases. For instance, for a consumption sensor, we need to align the measurements of consumption, weather measurements and weather forecasts and, usually, each of these sources have different frequencies.
The essence of the Data modelling step is to derive an appropriate data model to expose the multimodal data for monitoring and prediction functionalities. Research and development for designing data models, which enable monitoring and reporting, have found several solutions known for at least a decade and complies to standard Business Intelligence and Data models guidelines. However, in the case of NRG4Cast, we have to combine information from data streams and enable operations on streaming data, which means that we have to store the valuable information, but not the entire stream. Another concern, which does not comply with the standard approach of a data model design, is that we have to support data preparation for various prediction models, which is a challenge in itself. The usual approach to support this type of functionalities is to create dedicated datamarts, which include data at a more detailed level then the central database. In the MSDA framework, we decided for a distributed integration approach: we propose a central Monitoring database, which supports the usual monitoring functionalities and represents the central reporting database. MSDA framework specifies additional data structure, which is focused on data stream functionalities and contains more detailed data to allow stream operations and also data preparation for data/stream mining capabilities.

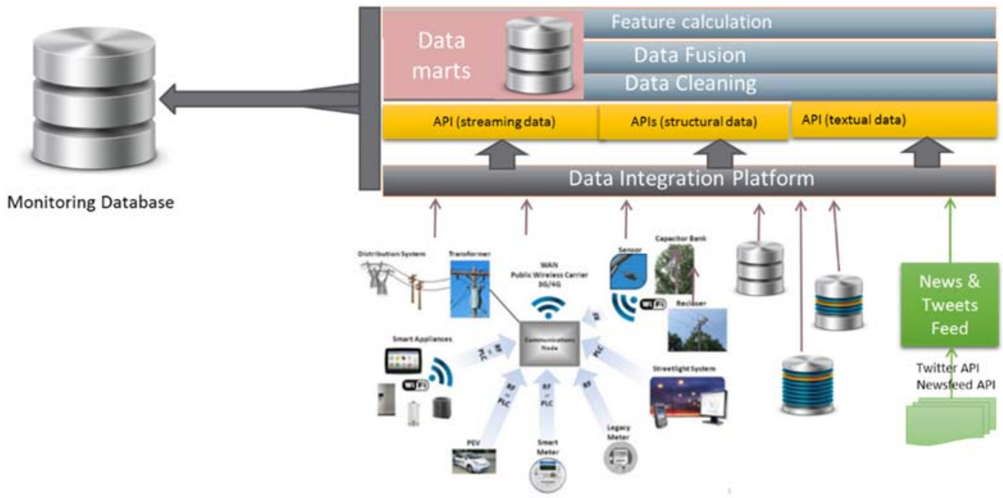

Figure 3: Multimodal data integration and fusion

The final step is data enrichment, which includes feature calculation for selected object measurements. For instance, using timestamp data, we can calculate a number of features: day of the week, month, day of the month, season, hour of the day. Using calendar data for each country, we can add additional information like working or non-working days, holidays. When we have working hours or teaching schedules for schools, we can add information about occupancy, lunch time, etc. At that point, we also calculate various statistical measures, such as trend/averages/min, max for selected history time-frames of consumption measurements, weather features like temperature, humidity, air pressure. These data features are carefully designed to support applications of various prediction methods. On one hand, they are generic enough to be applied on different domains and, on the other hand, these features should also cover specifics of selected pilots. So, the approach we recommend is to calculate a vast number of features andallow the automatic methods to decide which features contain more information. This is done in the data mining modelling cycle, where DM methods select most of the predictive features.

The results of Multimodal data integration and fusion layers' of information enrichment are:

- functionality of reporting and monitoring: which includes:
  - predefined reports (standard reports), which are defined by the end user. This is a set of standard reports that the end user must have each day/hour/week. The added value of these functionalities is that the end user gets important information quickly and in a well structured form.
  - ad hoc reporting – in this scenario , we let an expert user access the monitoring database in order to provide complex queries and test certain hypotheses. The added value for the end users is a flexible reporting environment, which allows full functionalities of reporting over the complex integrated data model. It enables the expert user to test certain hypotheses - for instance – checking seasonal, monthly, weekly trends, influence of working and non-working days, correlation with weather and so on.
  - complex visualizations on data streams: data sources for complex visualization capabilities are specific data structures inside the analytical platform, which is optimized for intensive operations over data streams. It enables expert users to analyse correlations of two or more streams in a single matrix chart. The main added value of this functionality is to enable expert users with deeper insight into relations between data streams and their influences on consumption. It enables correlation down to the level of raw data streams.

Functionalities of reporting and monitoring are described in more details in D4.2, D6.4 and D3.1.

It is important to point out that in the MSDA framework, we are dealing with various levels of complexity in terms of detail of granularity as well as possible space of action measures. There are three main complexity dimensions:

- time (level of granularity varies from almost real-time data (measurements) to 1 month or even seasonal consumption or prediction),
- object (we are dealing with different types of buildings (some divided further – rooms, offices, lamps, campus, electrical cars)
- stages of applied analytics (data modelling, anomaly detection, data fusion, complex event processing, prediction capabilities, profile creation on various levels)

In addition, the proposed MSDA framework and its implementation need to provide functionalities to support two different time horizons:

- information enrichment on lower time granularity – small time window, such as 15 min, 1 hour (like real-time monitoring, triggering alarms for specific events) and
- information enrichment on a higher level of time granularity, like providing various consumption profiles, predictions for longer prediction windows, such as 1 month.

To ensure that all the complexity dimensions are supported in a proper manner, we propose two main data sources: (1) Monitoring database, which deals with longer time-horizons and (2) Data marts in the Analytical platform, where data structures are optimized for stream operations and therefore enable data analysis and insights in more detail.

2.1.2 Information Enrichment layer: Forecasting of consumption, production and price

The Information enrichment layer for the forecasting of different measured quantities provides predictions on selected time-spans for consumption, electricity price on Energy spot market and production. It includes a complex prediction environment, which provides prediction models for various data sources at different prediction windows.

Prediction models provide predictions at the level of a sensor or time-series. In addition, for each prediction window (how far into the future we predict), we need to create a separate model. For better accuracy of results, we include a distributed model approach for the prediction window of 1 day in the MSDA framework. This means that instead of using one prediction model with a prediction window of 24 hours, 24 prediction models are created – for each hour of the day there is a separate model. The set of 24 models is used to "simulate" one prediction model with a prediction window of 24 hours. This distributed approach is described in more detail in D3.1.

Taking into account all of the above, we can encounter that the final selection for one time-series (an example can be seen in

Figure 4), which we want to predict, consists of:

- 1 prediction model with a 15 min prediction window
- 1 prediction model with a 1 hour prediction window
- 24 prediction models with 24 hours of prediction windows
- 1 prediction model for 1 week

In summary, at least 27 prediction models for each time series. Here we also have to take into account the fact that the number of prediction models can be even higher, if we would use a distributed approach for 15 min intervals.

For each of the 27 models, we need to automatically monitor performance and provide appropriate pre-processed data (calculated features). Since, in real-life, we can encounter specific situations where all of the data needed for feature calculation is not available, we need to prepare alternative actions – what set of steps will we perform in order to obtain predictions (it could be some simulation of input data, it could be variations of previous prediction and other more complex solutions).

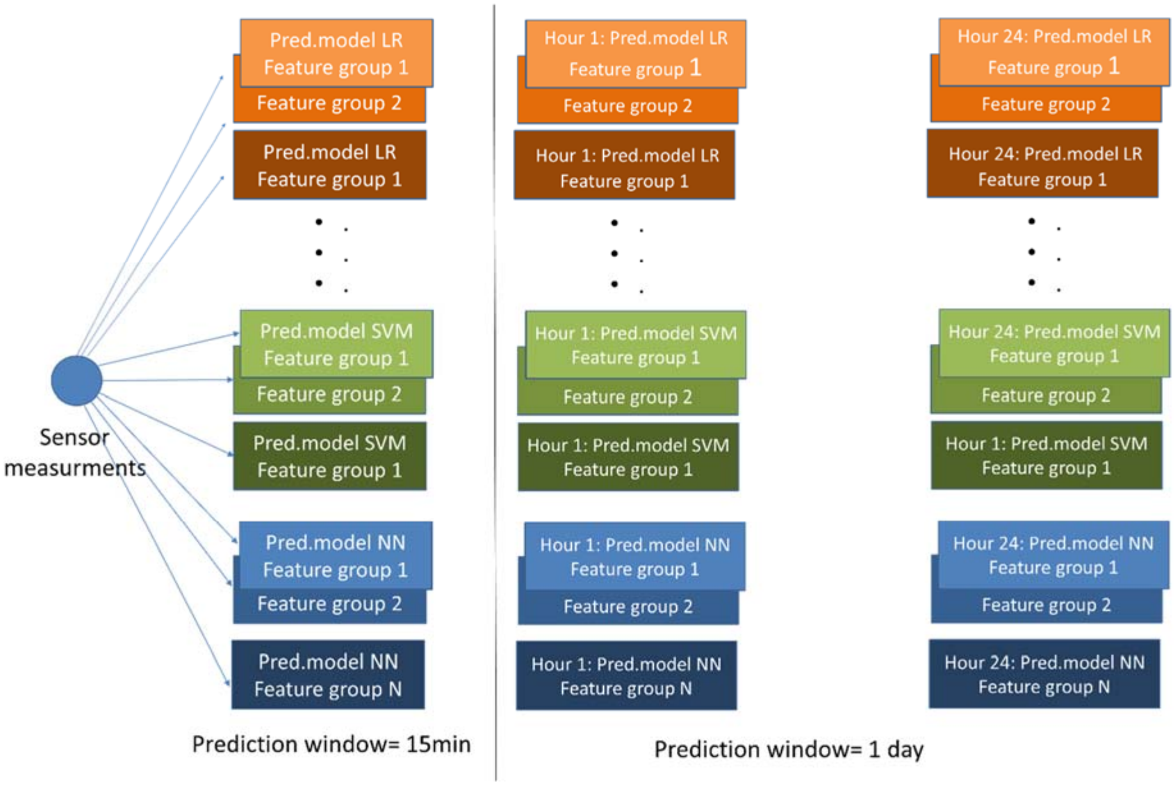

Figure 4: Set of prediction models for single data source (consumption sensor) for two different prediction windows (15 min and 1 day)

During the modelling phase of a single prediction model, we need to analyse various data mining methods and test them on different feature-sets. In the NRG4Cast framework, we opted for various types of linear regressions, different SVM methods, regression tree models and a set of different Neural Networks. Baseline models were also implemented, which include statistical approaches (moving average) or simple last-value models.

In practice, that would mean that for each time series, we have to produce:

- for one specific method (SVM, LR, NN, HT, various statistical methods):
    - different parameter settings, which correspond to different variations of the method – at least 10 different settings should be tested,
    - a model with unique parameter settings needs to be tested on different feature sets – the number of them could be infinite, but for the sake of reality, we are not using more than 15 different features sets (some of them include weather forecast, different statistical measures, measurements from related sensors).

The estimated calculation for each of the tested methods would be 10x15 prediction models learned and tested.

If we multiply 150 models with the number of tested methods, we get a rough estimation of how models needs to be created and tested. The huge number of prediction models needs to be properly stored along with all corresponding meta-data (feature-sets, parameters, test statistics, model ID). Since in the modelling phase, we need to create models, test them and, in the end, as a final phase, decide for a final selection, the computational complexity is huge. Since the basic steps are the same for each method, we propose, within the MSDA framework, an automated modelling environment, which will enable the user to automate the modelling phase and focus on interpreting the results and, therefore, make final model selections.

In the above paragraph, we only focused on the alternating prediction window. However, in addition to this, we can also vary history window (how much historical data model is used in the learning process). With variations of history windows, we can distinguish between:

- stream mining – learning directly on the data stream – very short history window,
- a classical data mining approach - learning off line and therefore using a long history window,

- a hybrid approach - the model is learning with a history window which is longer than the conventional stream mining and shorter than the classical DM.

These three approaches can have different parameters (length of history window), that need to be supported in an automated modelling environment.
There is also another important parameter within the modelling phase: the definition of black period (the time when data cannot be available in real implementation). For instance, in some cases, data measurement using a 15min interval, can be accessible with the delay of 1 day. In that case, black period is set to one day and this needs to be incorporated into the data preparation phase.

In the MSDA framework, we propose an automated environment for the modelling phase, which should be implemented in a way that enables parallel computing. An expert – an analyst that is able to use the results – created the models using corresponding metadata (including test statistics) and therefore focuses more on the model selection phase. When the final model sets will be selected, the corresponding data (modelID, metadata) will be written in Monitoring DB, which will incorporate all of the required information for model maintenance. Model maintenance data includes information about the model itself (expected performance, metadata), actual performance, date of implementation, sensorID, and the active flag. Proper maintenance of prediction models is crucial for successful implementation in practice.

2.1.3 Information Enrichment layer: Complex event processing

Complex event processing (CEP) Information Enrichment layer offers identification, analysis and prediction of complex events. Events are defined as situations, which can be described as a rule or pattern that combines data from multiple sources. These patterns can suggest simple or more complicated circumstances. The goal of complex event processing is to identify meaningful events, which can be opportunities or threats and respond to them as quickly as possible.
Prerequisites for this type of functionality in IT systems is data integration and almost real-time processing of data streams along with access to predictions for measured quantities.
MSDA framework positions CEP functionality on top of the Forecasting Information enrichment layer. The reasons for that are two-fold:

- CEP functionalities in the proposed framework include automatic detection of prediction models,
- enable users to create complex events, including various predictions.

CEP functionalites have access to analytical data marts, which allows access to lower granularity of data. But the main advantage is that the Analytical platform integrates various sensors like consumption points, weather (temperature pressure, humidity,…), weather forecasts and also other consumption points.
MSDA framework provides the following CEP functionalities:

- automatic detection of model performance changes,
- detecting complex events over multiple streams,
- allows the end-users to enter custom patterns.

MSDA framework allows expert users to tailor CEP functionalities to their specific needs. Domain users should define interesting events and corresponding actions. In general events, or in same cases, anomalies can be divided in two major groups: **expected** and **unexpected events**. An example of a simple expected event would be a broken sensor, exceeding temperature, or some other measure. The difference between an expected and unexpected event is that when an anomaly in the data occurs, we know what is the most probable cause and can take appropriate (predefined) action. Anomalies in the data, which were not yet seen or expected, are part of unexpected events. This means that we don't know what was the reason for the anomaly. In that case, a list of such events should be created and sent to the appropriate domain expert to further analyse of the anomalies.

Under expected events, we also encounter model performance, which can be, for instance, worse than expected. When this type of alarm is triggered, appropriate an action would be to update the prediction model. When the prediction model is created, we also get statistical measures of model accuracy. Model accuracy measures bring information about how the model should perform in real life. If the prediction model starts to perform worse than expected, the accuracy measures on real data (automatic validation) would fall under a certain threshold. If such events happen frequently enough, the prediction model needs to be updated.

From a content perspective, alarms and triggers can fall into different categories and therefore require monitoring in a different manner:

- Maintenance – the sensor is malfunctioning, devices are broken, room temperature is increasing (air-conditioning is not working), increased activity of turning on and off devices, warning for deadline for maintenance of devices.
- Behaviour changes: based on office sensors, we can detect inappropriate behaviour – such as open window and the air-conditioning turned on, an unexpected increase of consumption, different patterns of behaviour in longer time periods. This also encounters automatic monitoring of model performance.
- Economic – price of energy dropped or increased significantly.

For maintenance related alerts, a dedicated person needs to be contacted. They should also provide a list of maintenance related events and appropriate actions. Such events can be detected from short term data and usually require immediate actions.
For behaviour patterns and their changes, the contact persons should be energy managers and people who are planning awareness actions. Creating the list of patterns and appropriate actions needs to be done by dedicated people in this area. These types of analyses and alarms are usually done over a longer period of time and don't require immediate response actions, but provide valuable input.
Economic events and related triggers should be within the domain of energy managers and represent valuable insight into the energy market economy. These alarms could be used in short term or long term data.

2.1.4 Information Enrichment layer: Knowledge generation

In the final Information Enrichment layer, we focus on inverted knowledge from enriched information from previous layers. The main results will represent:

- dynamic probabilistic models for multiple-sensor systems, which will enable deeper understanding of a system, measured by multiple sensors. The main idea is that the expert gets a top down perspective on clustered events and the transitions between different clusters.
- behaviour profiles on different levels of consumption (office, one sensor, building). These profiles will enable expert users to discover high level patterns of consumption and its correlations to information from different sources, such as correlation with weather, traffic, economy related factors (EPEX predictions).
- energy dashboard – which, on the one hand, will inform the expert user (energy manager) about the short term and long term trends of energy prices in the energy market, while, on the other hand, will enable the user to track energy related events.
- reasoning over prediction models - since we will have large sets of prediction models, we will apply reasoning techniques in order to detect similarity between them and logical rules.

All of the mentioned results represent knowledge generation at a higher level of information. Behaviour profiles will be generated using various clustering techniques, including hierarchical clustering, metadata about the selected corresponding prediction model, metadata about the corresponding object (type of object, usage, other information from the environment).

The process of creating behaviour profiles demands involvement of the domain user, which will work closely with the analytical expert in order to provide validation of different behaviour profiles from the content perspective. Behaviour profiles will be created on different summarization levels: on the level of each sensor/time-series, group of time-series (for one object), group of objects. The process of creating profiles includes various data integration steps, application of clustering techniques and cyclic content validation of the domain expert and, at the end, a final selection of the behaviour profile. Usually the process of creating profiles, demands the integration of metadata of corresponding prediction models to reveal the most influential prediction features and the additional application of prediction methods, in order to develop behaviour rules.

On top of these results and the original data measurements, we apply clustering methods and extract additional behaviour rules. Those rule sets are then tested and selected based on statistical measures. We push the set of rules with similar accuracy further to domain expert evaluation. As a final result, we get behaviour profiles, which comply with the statistical evaluation and are interesting and valid for domain expert users.

Another similar new knowledge can be generated on the level of a system. A system is defined with several measurement points in the corresponding environment. The Route Cause Analysis approach (described in more detail in D4.3) offers automatic detection of typical (complex) situations in the system over time, along with the typical paths between those situations and corresponding probabilities. An example of initial application of initial RCA results can be seen in Figure 5, where we can clearly identify different heating seasons. With a click on each cluster and its corresponding sub-states, the user can drill down further in order to get more detailed information regarding each state (situation), and what would be possible in the next state.

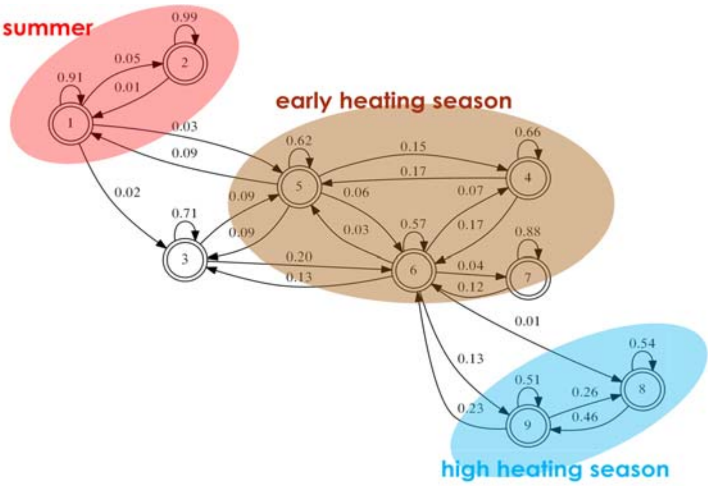

Figure 5: Example of RCA analysis for IREN pilot- thermal plan at Reggio nell' Emilia

Different kinds of new knowledge generation represent reasoning over large sets of prediction models. The main idea is to combine information from the predicted environment at a higher level (as kind of rules), with metadata from prediction models. With application of various reasoning techniques, we will explore similarity between selected prediction models with respect to description (metadata). Similarity will be explained with a set of human readable rules, which will enable analytics to understand the relationship between specific situations and appropriate prediction settings. Reasoning functionalities will combine the concept of situation awareness, context awareness, and case-based reasoning (CBR) along with general domain knowledge to produce suggestions for planning and optimization services. The application of CBR will provide solutions to problems by analyzing similarities to other problems for which known solutions already exist. They use analogical reasoning to infer solutions based on case histories. Results will also include complex situation awareness services based on meta-level reasoning approaches. These functionalities are part of the future work and will be reported in WP5 deliverables.

In years to come, with open energy markets, energy managers will need to be aware of economical situations in the energy market. Different energy markets have different influential factors for predicting energy prices, mainly weather conditions which have the biggest impact on renewable energy production. Besides weather factors, there are also different global business-related events, which can have impact on energy prices. This is exactly what the energy dashboard will provide to users. Users will be able to follow energy prices on energy markets, along with their prediction, which will also include local weather forecast, to collect its influences. Global energy related events will be extracted, tracked and analysed through a textual pipeline (described in more detail in D3.3).

3 Architecture of the NRG4Cast Platform

The MSDA framework represents the logical view of energy management IDSS system. In the NRG4Cast project, we are implementing the MSDA framework on pilot scenarios as the NRG4Cast platform.

The architecture of the platform is a "living" entity, which evolved through the phases of the project. At the end of the 2nd year, the architecture is almost fixed.

Figure 6 depicts the architecture of the NRG4Cast platform as realized and planned after the 2nd year of the project. Each component of the system is annotated with the deliverable number in which its development or improvement has been reported (orange labels represent 1st and 2nd year deliverables, teal labels represent planned 3rd year deliverables). The right-hand side of the figure is dedicated to the text-processing pillar and the other components represent sensor data related components.

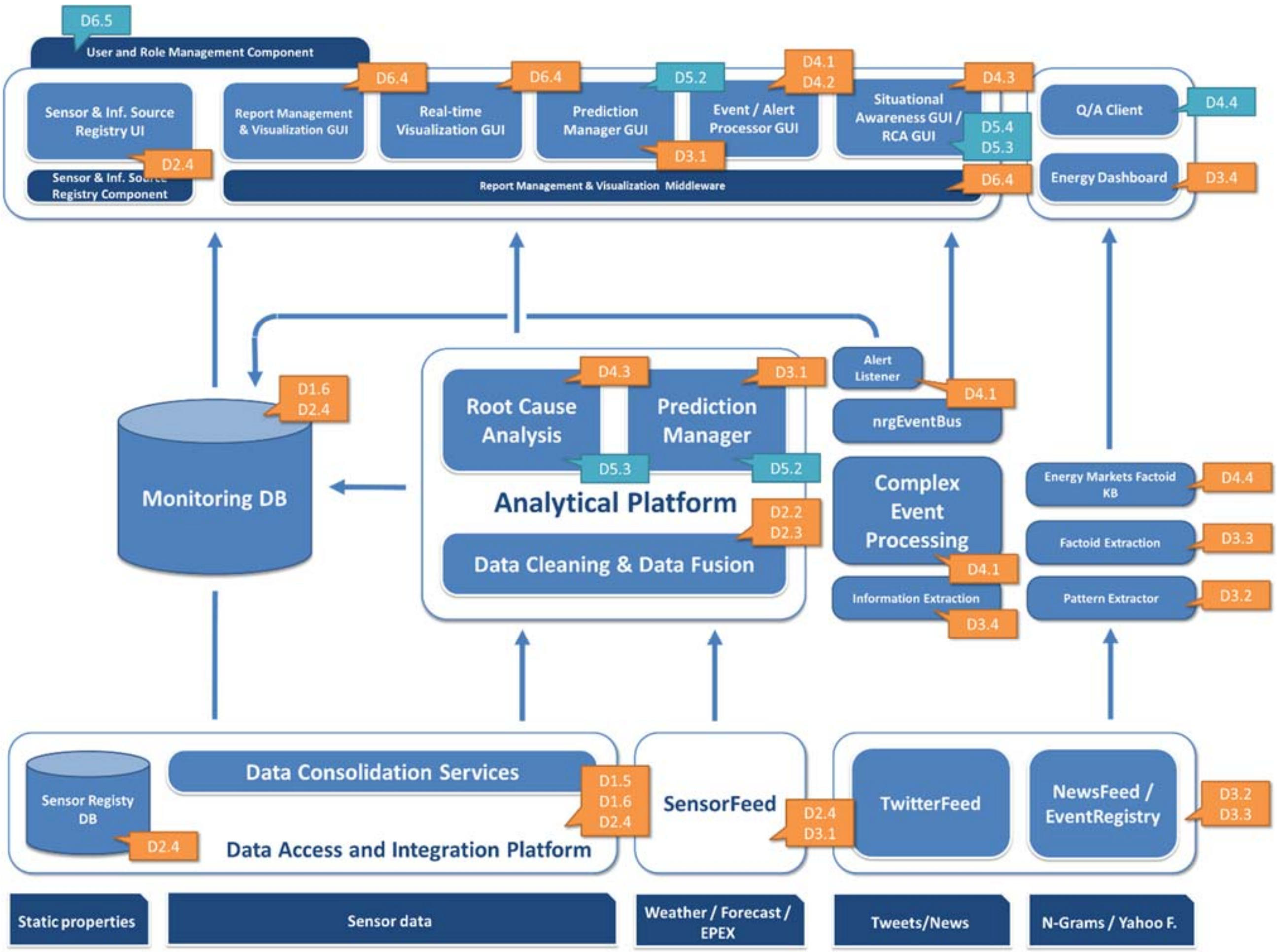

Figure 6: Architecture of the NRG4Cast Platform.

Data Tier

The band of components and data sources at the bottom of the figure represents the data tier of the platform. Most of the work in this part has been done in the 1st year of the project, with two exceptions: textual features have been integrated mostly in the 2nd year and also some additional sources have been included later for modelling needs.

Data Access and Integration (DAI) Platform based on the OGSA-DAI has been implemented in deliverables D1.5, D1.6 (Early and Final Prototype of Data Gathering Infrastructure) and D2.4 (Data Distribution Prototype). The purpose of the platform is the integration of different data sources and exposing this data in a unified way to the upper levels in the platform. The DAI platform does not handle all of the data. Weather data, weather forecasts and energy spot marker parsers operate independently through the **SensorFeed** component, which was also reported in D2.4 (Data Distribution Prototype) and then updated with the weather forecast and energy spot market parsers in D3.1 (Modelling of the Complex Data Space). Textual data is gathered via TwitterFeed (for tweets) and through *NewsFeed* (for news). The latter get aggregated in a component called *EventRegistry*. Both components have been developed outside the NRG4Cast project and are here used as a data source. Their integration has been performed through the work, reported in D3.2 (Semantic Enrichment Prototype) and D3.3 (Metadata Generation Prototype).

Application Tier

The middle band of the figure represents the application tier. An exception in this tier is the Monitoring DB, which should be present in the data tier. It was moved to the application tier for two reasons: 1. Monitoring DB plays a crucial role in the system as it is the central point for all the reporting, and 2. Monitoring DB structure already implements quite some logic and the DB interface itself represents a powerful monitoring/analytical tool.

Monitoring DB has been reported firstly in D2.4 (Data Distribution Prototype) and then updated in D1.6 (Final Prototype of Data Gathering Infrastructure). Modifications to the Monitoring DB were also quite regular throughout the development of the 2nd year prototypes.

The **Analytical platform** that is based on the *QMiner* software represents a stream processing engine. The engine has been prototyped firstly in deliverables D2.2 and D2.3 (Early and Final Data Cleaning and Data Fusion Prototype). Due to major revisions of the *QMiner* software, the platform has been updated in the 2nd year. In the 2nd year, the platform has been extended with multiple specialized instances. The Data Cleaning & Data Fusion instance has been named simply Data Instance in D3.1 (Modelling of the Complex Data Space). The instance is able to save the streaming data and calculates and stores aggregates of the streams. In the same deliverable, the Prediction Instance (Prediction Manager) was reported. Prediction Instance performs as a stream mining component and enables expert users to prototype and deploy different prediction models. Additional complex visualization tools have been developed for the Prediction Instance, which are not represented in the scheme. Prediction Instance will be improved and extended with deliverable D5.2 (Data driven prediction methods environment). Another instance has been developed in D4.3 (Root Cause Analysis service). The instance uses the same infrastructure as the Prediction Instance, but builds hierarchical clustering of multi-dimensional time-series with corresponding transition frequencies. Root Cause Analysis GUI has also been implemented in the same task. The GUI is able to visualize the hierarchically segmented state of the observed system. It is also able to perform some simple Markov chains, computations on top, which enable the user to identify the most likely origin which led the system to the final state. The work will be extended in D5.3 (Knowledge Driven Planning and Optimization Service).

Next to the right, we have a whole pillar dedicated to complex event processing. The main component in this section is the Complex Event Processing**.** The Component is able to detect complex events on top of data streams. Alerts get sent to the nrgEventBus via a HTTP client interface. From there they are broadcasted via WebSockets. The alerts get saved into the Monitoring DB with the help of Alert listener. All of these components, as well as the corresponding GUI's, have been developed within D4.1 (Complex Event Detection Service). GUI for event processing is twofold. On one hand, we have the real-time and historical viewer of the events and, on the other hand, a complex rule editing GUI has been developed, which allows an expert user to enter complex nested queries. Relevant input for the event detection services has been done through the Information Extraction, reported in D3.4 (Information Extraction Prototypes). The far right column in this layer is also dedicated to textual services. The entire idea has been described in detail in D3.2 (Semantic Enrichment Prototype). It contains Pattern Extractor, presented in that same deliverable, Factoid Extractor, presented in D3.3 (Metadata Generation Prototype) and Energy Markets Factoid Knowledge Base, which will be developed in the work of D4.4 (Knowledge Formalization Services).

Presentation Tier

Presentation tier represents GUIs and supporting apps. Sensor & Information Source Registry UI has been developed within D2.4 (Data Distribution Prototype). UI enables the administrator to interact with the

Monitoring DB and change configuration/meta-data of the systems components. Sensor Registry also utilizes a separate middle-ware component. The rest of the elements of the GUI are wrapped into an uniform user interface, which is set on top of Report Management & Visualization Middleware, which was developed especially for the 2nd year prototype in D6.4 (Real Time Monitoring Integration – 2nd Prototype). Of the components that have not been mentioned in description of Application Tier, only 2 are left: Report Management & Visualization GUI and Real-time Visualization GUI. Both have been reported in the previously mentioned deliverable D6.4. Components have been developed based on the work done in the D6.3 (Data Stream Integration – 1st Prototype) from the 1st year of the project.

The far-right components again belong to the textual vertical. They represent Energy Dashboard, which has been developed within D3.4 (Information Extraction Services). In the 3rd year of the project Q/A Client will be developed and reported in D4.4 (Knowledge Formalisation Services).

4 Summary and future work

This report proposes the **MSDA framework (Multimodal Stream Data Analytics)** as a generic IDSS framework for energy management systems for the dynamic prediction of energy demand, energy network failures and the detection of complex events through employing various reasoning approaches.
MSDA incorporates a hybrid approach of combining knowledge discovery in database and advanced analytical technologies, such as Data/Stream Mining and reasoning techniques in an integrated environment for IDSS. Through this hybrid approach, the MSDA framework would emphasize o access to and manipulation of the NRG4CAST models (statistical, DM, financial, optimization) and at the same time suggest and recommend actions specialized in problem solving.
The aim of the proposed framework is to support decision making by recalling past information (history data), inducing “chunks” of domain knowledge from this information (information enrichment) and performing reasoning upon this knowledge in order to reach better-informed conclusions in different business situations.
We described MSDA at different complexity layers of Information enrichment as its corresponding results and used techniques with references to corresponding deliverables for a more detailed description.
With new knowledge, generated from Information enrichment layers, interested energy management stakeholders can extend their set of business actions, which can be used for the introduction of new measures for energy savings or simply for acquiring more precise information on their consumption, which gives them a better negotiating position in the energy market, when signing or extending contracts with energy distributors. NRG4Cast platform will cover a wide range of scenarios, pursued by different types of stakeholders: energy producers, energy distributors, big complex energy consumers (municipality, university campus, smaller private consumers and “prosumers” – electrical cars).

Within the NRG4Cast project, we will implement a MSDA framework as the NRG4Cast platform, which is available at the end of the 2nd year as 2nd year’s Prototype (D6.4). It includes work from WP1, WP2, WP3 and WP4, which provide functional modules, while WP6 is focused on integrating all of the components.

In the 3rd year, we will continue and finish the work on Knowledge formalisation services (D4.4), which will include knowledge formalisation prototypes. We will focus on the development of statistical and machine learning prediction environments, which will incorporate functionalities from the Forecasting Information enrichment layer. We will start with the development of planning and optimization services based upon reasoning techniques which will be part of the New knowledge generation Information Enrichment and will include several optimization approaches based on Machine Learning techniques. An interesting task will be the development of complex situation awareness services based on meta-level reasoning approaches, including CBR.
At end of the 3rd year, NRG4Cast will provide implementation of the MSRA framework with coverage of very different use case scenarios and its corresponding specifics, which will enable NRG4Cast to show its advantages.

References,

[1] EU Energy Efficiency Directive (as of 06/22/2011)

[2] EU Commission Directive 2012/27/EU on energy efficiency

[3] CBArch: A Case-Based Reasoning Framework for Conceptual Design of Commercial Buildings by Andrés Cavieres, Urjit Bhatia, Preetam Joshi, Fei Zhao, Ashwin Ram, AAAI-11 Spring Symposium on Artificial Intelligence and Sustainable Design; www.cc.gatech.edu/faculty/ashwin/papers/er-11-07.pdf

[4] Decision support methodologies on the energy efficiency and energy management in buildings, D. Kolokotsaa, C. Diakakia, E. Grigoroudisb, G. Stavrakakisc & K. Kalaitzakisc , Advances in Building Energy Research Volume 3, Issue 1, 2009, pages 121-146

[5] “Mine the Data”New trends in energy management systemsand benefits for mining applicationsRam Kaushik, Offer Manager, Professional ServicesSchneider ElectricSME Annual Meeting and Exhibition 2013

[6] Reşat Selbaş, Arzu Şencan and Ecir U. Küçüksille (2011). Data Mining Method For Energy System Aplications, Knowledge-Oriented Applications in Data Mining, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-154-1, InTech, DOI: 10.5772/13710. Available from: http://www.intechopen.com/books/knowledge-oriented-applications-in-data-mining/data-mining-method-for-energy-system-aplications

[7] H. Doukas, C. Nychtis, J. Psarras, "Assessing energy-saving measures in buildings through an intelligent decision support model", Building and environment, 2008

[8 V. Singhvi, A. Krause, C. Guestrin, "Intelligent Light Control using Sensor Networks," Proceedings of the 3rd international conference on Embedded networked sensor systems (SenSys 2005), 2005

[9] Barbato, L. Borsani, A. Capone, et al., "Home Energy Saving through a User Profiling System based on Wireless Sensors," Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (BuildSys 2009), 2009.

[10] Y. Gao, E. Tumwesigye, B. Cahill, et al, "Using Data Mining in Optimisation of Building Energy Consumption and Thermal Comfort Management," 2nd International conference on Software Engineering and Data Mining (SEDM), June, 2010.

[11] Farid, F.; Shahrestani, S.; Chun Ruan "Application Based Handover: An Energy-Efficient Approach for Heterogeneous Networks", Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference on, On page(s): 1185 – 1190

[12] D. Xenakis, N. Passas, L. Di Gregorio, and C. Verikoukis, "A Context-Aware Vertical Handover Framework Towards Energy-Efficiency," in Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd, 2011, pp. 1-5.

[13] Juan J. Cárdenas, Luis Romeral, Antonio Garcia, Fabio Andrade, “Load forecasting framework of electricity consumptions for an Intelligent Energy Management System in the user-side”, Expert Systems with Applications, Volume 39, Issue 5, April 2012, Pages 5557-5565

[14] GneraEnergy Group: solution: Energy Management: Monitoring and Forecasting; http://www.gnera.es/ENG/gnera-energy-index.html