Rethinking Energy Data Management: Trends and Challenges in Today's Transforming Markets

Robert Ulbricht[2], Ulrike Fischer[1], Wolfgang Lehner[1], Hilko Donker[2]

[1]Dresden University of Technology, Database Technology Group, Germany
[2]Robotron Datenbank-Software GmbH, Dresden, Germany
first name.name@tu-dresden.de
first name.name@robotron.de

Abstract: The energy market domain is subject to a continuous transformation process, mostly driven by governmental regulations. To efficiently handle the large amounts of data and the communication processes between market participants, specialized database applications have been developed. In this paper, we present the energy data management system (EDMS) as a standard software solution, describing its core components and typical system integration aspects. However, current market topics like smart metering, energy saving, forecasting for renewable energy sources, mobile consumption and smart grids lead to new database challenges. We provide an overview of these trends and discuss their impact on existing information systems, focusing on the technical challenges of data integration, data storage, data analytics and scalability. As energy data management has to match those new requirements, promising research opportunities are offered to the database community.

1 Introduction

In the past years we could observe numerous changes on the European energy markets. Starting with the market liberalization in the late 1990s, a wide range of regulations forced energy companies to audit and completely redesign their organizational structures, business processes, and technologies. The unbundling of the former vertically integrated and not uncommonly public utility companies subdivided them into local or regional distribution system operators (DSOs), transmission system operators (TSOs) and energy producers or -suppliers. The political objective was the establishment of a competitive market environment, enforcing higher service quality, more innovative technological concepts and simultaneous cost reductions. As the new market roles' activities had to be restricted to certain operations [JP05], participants were forced to interact intensively. This also created new requirements on existing system landscapes. To efficiently handle the massive amount of data created by the new communication streams, the first generation of a specialized information system sub-type was introduced: the *Energy Data Management System* (EDMS). The EDMS' main tasks are (1) to store and process almost all kinds of master- and transaction data related to energy logistics, and (2) to manage all automated data exchange processes between the different market partners in given time frames according to particular market rules. For example, the DSO has to send daily load curves retrieved from customers equipped with remote load profile meters to the corresponding energy supplier, the supplier is responsible for purchasing the distributed energy from a trader or broker, and finally the trader has to contract the capacities from energy producers to close his open positions. The TSO provides the infrastructure and necessary balancing energy for the trans-regional high-voltage transmission network. His associated balancing coordinator is responsible for monitoring the matching between demand and supply for all energy traders operating within the transmission grid (see Figure 1).

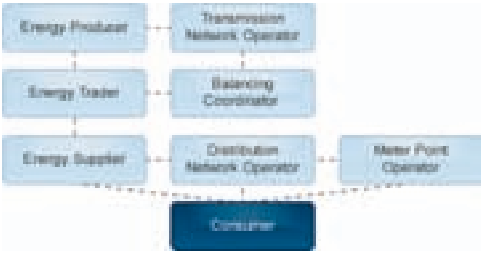

Figure 1: Market roles and relations after complete liberalization

Once established as costly individual implementations, pioneers and early movers soon initialized ambitious standardization efforts turning their solutions into market standard software. Because of the obtained synergies, the EDMS soon became an interesting and affordable option even for smaller, non-unbundled utility companies or big consumers like energy-intensive industries.

With market rules frequently being changed driven by diverse governmental regulation efforts (e.g. climate saving propositions), and new technologies becoming available, the EDMS is subject to a continuous and evolutionary adaptation process. So we notice the strong influence of constantly increasing capacities of renewable energy sources due to excessive funding policies (e.g., [Eur11a], [Ren12]) and industrial promotion, making conventional power plants becoming less attractive to private investors. Renewable energy is characterized by a decentralized allocation and fluctuating output, thus making it difficult to maintain stability in power networks where energy supply and demand must be balanced carefully. Amongst others, this is one security-related aspect of the energy supply system transformation (or transition) studied in [NPS$^+$12]. To prevent collapsing grids in the near future, and to battle the negative influences of the growing economies' ever increasing energy consumption on greenhouse gas emissions, new technical concepts are introduced. This includes e.g. intelligent and networked measurement devices, improved forecasting methods, efficient storage systems or the integration of mobile consumption points.

In this paper, we will address the technical challenges for energy data management systems against the background of contemporary market policies. In detail, we make the following contributions:

- First, we will show a brief market review of existing standard products and some ongoing related research projects (Section 2).

- Second, we will give a general description of commercial EDMS, their core components, the tasks involving them within a standard architecture and the position of the EDMS in a typical system landscape (Section 3).
- Further we will provide an overview of current topics related to energy market transition and outline their impact on requirements for information system architectures (Section 4).

Finally, we conclude our observations and finish by giving some additional directions for future developments (Section 5).

2 Market Relevance

Energy data management systems are already available as standardized commercial products provided by software developing companies. Leading systems with numerous installations in different countries are *Oracle Lodestar* or the *SAP IS-U-EDM*, but due to many country-specific requirements on local markets, there are no all-dominating multinational players. Most of the market participants are still acting locally or regionally, and therefore often prefer to choose national service providers instead of adapting foreign system philosophies. Take a look at Germany for example, Europe's biggest energy market, where well-known and widely spread solutions especially among large utility companies are the *robotron*ecount* product family and *Soptim's* energy logistics solution, or the *BelVis* EDMS primarily focusing at customers of small and very small size. There are a couple of other very active but smaller players, appearing on the complete listing recently published in [Sto12]. Less frequently and shrinking in importance, EDMSs can also still be found as proprietary solutions implemented by the utility companies.

Beside commercial implementation efforts, some interesting local and supranational projects and research initiatives shall be mentioned. Each of them addresses the field of modern energy data management or related topics. In Europe currently very popular are all type of smart grid projects, almost 200 are listed on [GGFS11]. Amongst them, the most notable examples are: the *MIRABEL* project [BDD+12], trying to balance energy supply and demand using micro-requests and flex offers; the *EU DEEP*[1] project, dealing with the integration of distributed energy sources in today's electricity systems, and the *ADDRESS* [PBB+09] project led by Italy, focusing on the active participation of small and commercial consumers in a large-scale demand respond concept implementation. The *e-Highway2050*[2] is a recently approved research project on the pan-European transmission network in order to facilitate cross-border energy exchange. The Danish *EDISON*[3] project focuses on intelligent system integration to manage charges of distributed electric vehicles plugged into a grid. Also in the United States, smart grid and energy storage projects receive massive governmental support [US 12], and even Asia's emerging countries like

[1] http://www.eu-deep.com
[2] http://www.entsoe.eu/system-development/2050-electricity-highways/
[3] http://www.edison-net.dk

China and India are dealing with the implementation of smart grid technology in order to compensate their relatively week network infrastructures [HHM11].

As shown in this section, there are plenty of commercial and scientific activities involving energy data management topics. Next up we will present the EDMS as part of the underlying information technology in more detail.

3 EDMS architecture

Although the standard functionality of today's available EDMS strongly varies depending on national market rules and specific business requirements, some similar design concepts can be observed. In this section we introduce the EDMS giving a brief description of its architecture (compare Section 3.1), some supported standard processes (3.2) and common interfaces (3.3).

3.1 Core Components

In the system architecture of EDMS, we can identify seven generic core components shown in Figure 2: First, a database is used to store *Time Series* and includes an application programming interface (API), another database contains the *Master Data* objects. Then, realized either as functional modules within the database or as external applications, there are a *Task Scheduler*, a *Communication Gateway*, a *Calculation Module* and an *Administration Module*. They represent the common base for any additional specialized advancement, for example energy balancing or forecasting, modeled as *Business Logic Modules*. Therefore, they are considered as the obligatory part of an EDMS and referred as the energy data application framework. Below, we describe this framework looking closer at its elements, focusing on embedded functionality and relations within the system architecture.

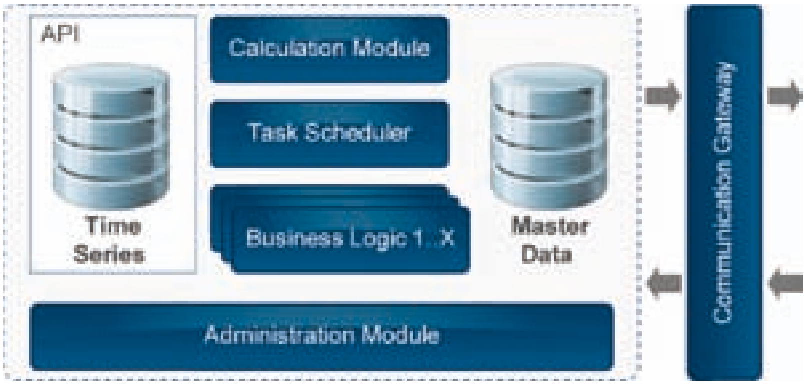

Figure 2: EDMS core components

Time Series Management

Doubtless the most important requirement of a modern EDMS is the ability to handle time series data, which can be defined as sequences of numerical values in successive order collected over a period of time. Time series are common data structures for energy logistics, used to handle production or consumption data, temperatures, prices or any other related numerical information. They occur either with equidistant (=equally spaced) or non-equidistant intervals and can have additional attributes like quality status information.

Because of the large amount of time series exchanged on the market every day, all values and attributes have to be stored in a space saving and therefore compressed form but must be easy to extract, thus guaranteeing user access at any given time. This means that compression and decompression has to be done directly within the database. In Figure 3 we demonstrate an exemplary way to reduce numerical raw data combining it with separately stored context information (extracted from master data database): Every equidistant time series is stored as one single row in a table. Now time stamps are redundant, as any value can be addressed looping through the string record having the time series starting point and its period. Second, energy consumption values can carry additional status information characters extracted from the meter (e.g. T = true value, S = substitution value, E = error). Run-length encoding is another good strategy to compress them, as this type of time series is supposed to contain many true values mixed with very few disturbed ones. Like anywhere else, choosing the best compression method is a cost-optimization problem, balancing reduced hard disk space against additional CPU capacity.

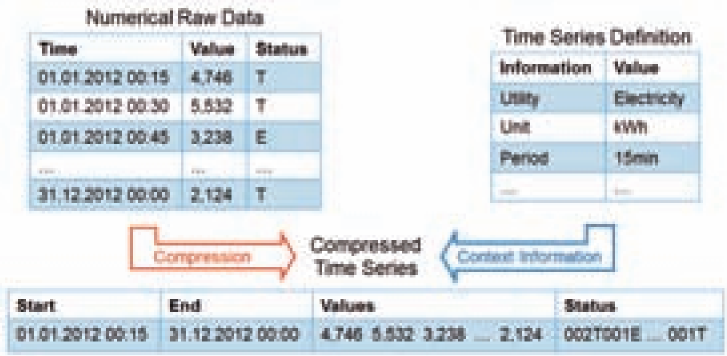

Figure 3: Structural transformation of equidistant time series

To avoid integrity problems within compressed structures, direct data manipulation cannot be allowed. Any reading or writing operation must strictly be done via an adequate application programming interface, called the time series API (Figure 2). Time zones and daylight saving time shifts have to be taken into consideration too and must be included. Further, due to the financial aspect of billing-relevant time series, all changes made on original raw data must be logged and archived. Transaction logging enables the system to automatically recover a historical situation whenever it should be requested by a market partner.

Master Data Storage

Any pure numerical data has to be related to master data objects in order to access its context information and for further processing. Master data can be extensive and modeled in many different ways. However, we observe an often repeated fundamental design concept: market partner and meter point are represented as independent entities, being temporally connected by the contract entity as shown in Figure 4. For instance, the relationship between an energy supplier and a customer (both market partners) is represented as one or more supply contracts between them, having one meter point assigned to each contract. This allows the system to handle all relationship changes (e.g. supplier change, customer moving in or out) modifying only the contract object, leaving everything else untouched. Also, contracts are not stringently required to have a real-world counterpart. Being virtual objects, they can be used to model any connection between market partners in the system.

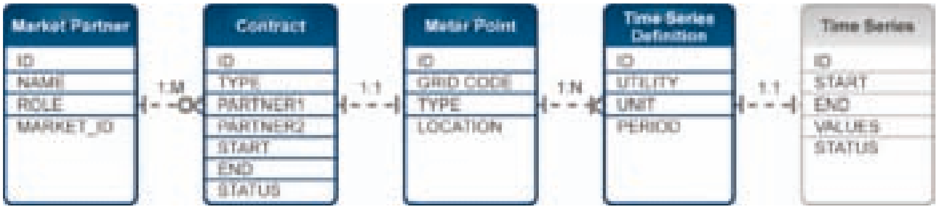

Figure 4: Relational master data objects

Load curves (time series) or numerical consumption values are assigned to the meter point object via the time series definition, which contains the additional context information shown in Figure 3. The meter point is identified by its official and unique market ID, the so-called metering- or grid code. The application must contain an audit trail for the key master data objects. The audit trail contains the prior and the new state of the data, the modifying user and date. Some of them are relevant for triggering events in order to communicate updates on master data to external market partners or neighboring systems.

Calculation Module

The calculation module offers all necessary mathematical and logical functions to manipulate existing time series or to calculate new ones based on input data (e.g. balancing sums). To offer the highest performance possible, it is directly based on the time series API. Physical units and periods of source data are automatically converted to the corresponding targets definition. Calculations can be arranged in a tree-like hierarchy according to existing master data structures. This is a common way to model different levels of aggregation. Once a hierarchical calculation is defined and activated, results are materialized and stored. Optimized maintenance strategies are implemented to force an automatic recalculation of all dependent nodes in the tree with every single update on underlying source data.

Communication Gateway

The communication gateway represents the connection between the EDMS operator and the energy market. The framework element is responsible for sending and receiving data using standardized electronic message formats like Edifact subsets or customized file formats (e.g. CSV or XML structures). Therefore, in opposite to traditional database applications, automatic file format conversion is a major issue. Today's market communication is mostly done by exchanging automated emails between market partners, so protocols like SMTP, IMAP or FTP must be supported. Manual interactions by people via phone or fax are reduced to an absolute minimum, speeding up business processes through the elimination of typical human errors and restrictions. All received messages must be confirmed to the sender, indicating their syntactical and semantical correctness. Validation is done outside the system to keep incorrect files away from database transaction processes. Any transmitted data has to be treated as confidential information. That is why the module's architecture must offer a possibility to include state-of-the-art encryption mechanisms. The communication gateway must be directly connected to the master data and time series databases in order to import or export data.

Task Scheduler

As most of daily work consists of handling mass data, an EDMS is designed to perform all tasks at a highly automatized level. This is usually done by using time or event based task scheduling. Realizing time based scheduling using database core functionality is a simple and standard automatization approach. All tasks are programmed as scheduler jobs and can either be executed immediately or stored in a queue and processed step by step by assigning pre-defined priority levels. If results are not needed immediately, large report queries or complex hierarchical calculations can be scheduled in low-workload time frames, thus avoiding interferences with daily work. Job results are protocoled and result quality can be classified. Failed jobs generate new tasks, for example by mailing the error log file to an administrator. Compared to time based scheduling, tracking down possible events in order to start certain scheduler jobs is the more complex way by using common pushing or polling techniques. As for pushing, triggers must be implemented and integrity checks must be used on all relevant tables. However, triggers are hard to debug and can slow down a system's performance notably as they are active on every transaction or can cause other triggers to fire. To avoid all this polling can be used instead, but depending on the data model, the query intervals and the number of simultaneous tasks this can lead to expensive operations in relational databases.

Administration Module

Like any other standard software of comparable complexity, the EDMS needs a vast number of configuration options to offer maximum flexibility. These activities are bundled in the administration module. Parameters can be set at different levels: global, thus affecting operations throughout the whole system, like setting the servers time zone, and client- or user-specific (e.g. language settings, output file format, etc.). Additionally, local settings like user accounts and user groups are administrated here. The application has to implement a robust user security model. We distinguish between role-based restrictions on certain database functions and all or only partial-level data access restrictions with different granularities for reading and writing operations.

Business Logic Modules

Besides basic functionality, the framework includes additional elements offering support for specialized business purposes. This includes pricing, forecasting, a workflow-engine with pre-defined business process schemes, contract management and balancing-, trading- or procurement modules. There are different points of view on what type of processes should be included in the EDMS or which should rather be handled in an external stand-alone application. These are influenced fore-mostly by three factors: (1) the market role of the system operator, (2) the evolutionary level of development and (3) the software company's product strategy. Some important functions needed by network operators are dispensable for energy retailers and vice versa. State-of-the-art systems tend to offer much more complexity than older products. For our work, we abstain from a further particularization at this point and refer to those elements as business logic modules in general.

3.2 Typical Energy Data Management Processes

Now we demonstrate how the recently introduced energy data application frame-work components work together and can involve additional functionality. Therefore we describe two typical business processes as exemplaric use cases.

Load profile processing

Load profile processing is a fundamental process which has to be handled by every network operator and is always following the same pattern (Figure 5): The DSO receives load curves from an automated meter reading system (AMRS) or from external market partners in daily intervals. The data is directly imported into the EDMS using the communication module (compare Section 3.1) and converted into time series (*raw data import*).

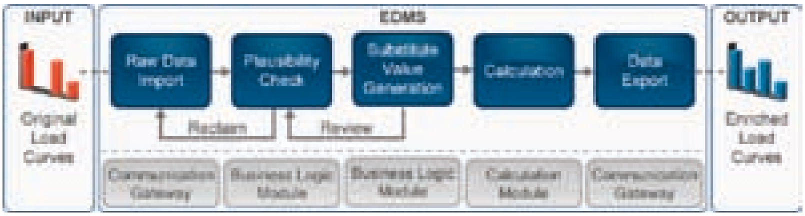

Figure 5: Load profile processing

Second, the DSO runs a *data plausibility* check. Using pre-defined comparison parameters, the quality of received load curves is verified. Any missing, incorrect or mismatching values are identified, marked for further treatment and excluded from processes like calculations or data exports. To correct these problems, *substitute values* are generated using common mechanisms like linear interpolation, single value distribution, copying historic values from comparable days or, if available, from a control measurement. These three process steps are iterative, so if all correction efforts fail the missing data must be claimed again from the sender. Once having a corrected load curve, the DSO can start with the data processing, normally consisting of calculating aggregated load curves for customers with multiple meter points, balancing sums for external suppliers operating within the DSO's network or providing billing-relevant consumption values (*calculation*).

Finally, the data export step extracts and ships the enriched, corrected or completed data using the communication module again. Because of the high amount of raw data and intraday handling times, the whole process chain is executed almost automatically by the task scheduler, reducing manual interaction to the indispensable cases only.

Energy sales and distribution

Besides the DSO point of view, another good example for the practical use of the EDMS is the main business process conducted by any energy supplying company: Selling and distributing energy to consumers (Figure 6).

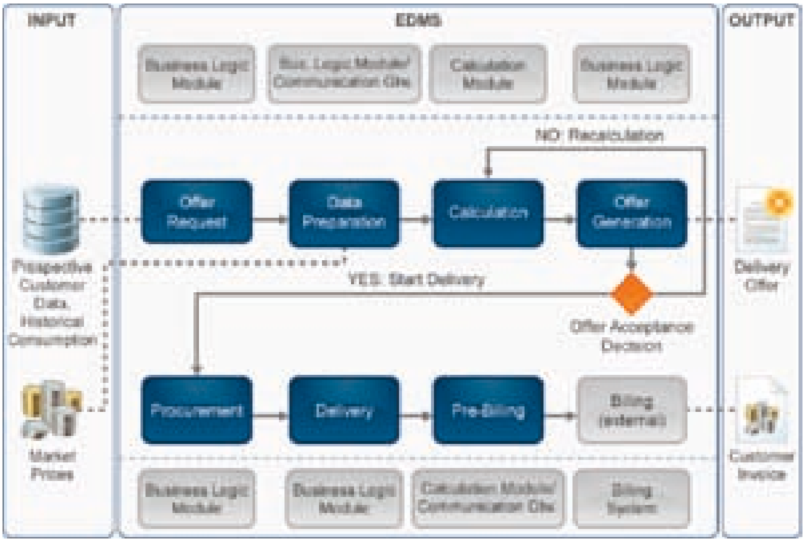

Figure 6: Standard energy sales and distribution process

The process chain starts with a prospective customer contacting the sales department requesting an offer (*offer request*). The next step is the *data preparation*. This means that the master data structures are created, the future energy consumption during the whole requested period is forecasted using historical values (if available) or based upon experience made with similar demand profiles. Furthermore, the market price information, received either from an external information broker (e.g. stock exchange services) or derived from own investigations, must be added. Now *calculation* can start by using the predicted load curve and the corresponding market price curves. Having all source data available, the process moves on to *offer generation*, where the calculated prices are approved by superiors, printed in a standardized form and handed over to the potential customer. At this point the further course depends on customer's decision: A refused offer can be reworked (e.g. including discount) or closed, thus terminating the whole process; an accepted offer initiates the *procurement* phase. A new supply contract is created and added as open position to the supplier's portfolio in order to procure the offered amount of energy from the market. After that, the *delivery* can start. The new contract is communicated to the corresponding network operator via the supplier change process. Only now the customer starts receiving energy from his/her new supplier. At the end of the billing period (e.g. one month after first delivery), the consumption data has to be collected, pre-processed and handed over which is done in the *pre-billing* phase. Finally, *billing* can start. However, this activity has rather to be realized in a proper billing system.

As shown in this sub-section, numerous modules are involved in different energy data management processes, each module offering dedicated functionality. This underlines that an EDMS is a complex application, which requires a deep level of component integration in order to avoid unsolicited breakpoints in highly automated processes.

3.3 System Integration

In practice, we experience different approaches to integrate the EDMS into a company's system landscape. Generally, we consider the EDMS to be the central data processing unit within a utility company due to the following reasons: (1) energy data management is a centralized task involving various independent corporate units, (2) most of the data needed for daily operations and reporting is available in the system and (3) it's powerful communication module is the primary gateway to external market partners or can be well used as internal interface to neighboring systems within the corporation.

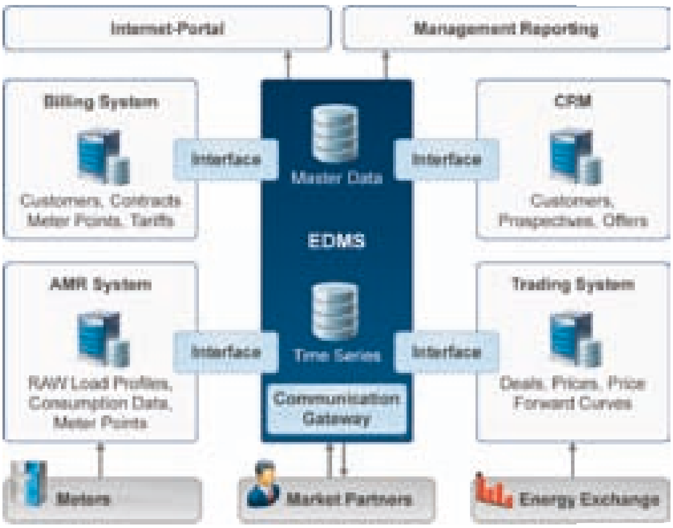

Figure 7: Example for an EDMS integration into a system landscape

In Figure 7 we demonstrate a typical system landscape in an integrated view for both DSO and energy supplier: The central positioned EDMS receives master data either from the billing system or the customer relationship management system (CRM), as they usually possess master data superiority. Meters send their load profiles to an AMRS, where raw data is accumulated, converted into time series and forwarded to the EDMS for further processing operations. A trading system provides real-time or forward price curves needed for retail offer generation (see 3.2) - deals and portfolios are managed here too. The EDMS can be integrated into a web portal, allowing customers to access their accounted consumption values online. Furthermore, the database is used to generate all kind of management reports on relevant information.

Usual interface designs include database links, either directly between the relevant systems or using a connecting middleware, web-based services frequently being used in SOA architectures or even simple file transfer based on plain text formats or XML structures. A second way to realize data exchange between dedicated systems is to use the official standard market formats (e.g. Edifact), assuming that applications should be able to handle them accordingly. This makes implementation projects cheaper and faster, but has certain disadvantages: the information transfer is restricted to the format's syntactical capability and with every change on external file formats the interface needs maintenance. Finding the optimal technology basically depends on transfer frequency, data volume and -complexity and reasonable maintenance expenditures.

4 Challenges

New technological trends and political changes constantly create challenges and additional requirements for existing system landscapes. In this section we address some of the market's frequently discussed topics and analyze their possible impact on energy data management. As for the market challenges, smart metering (4.1), energy forecasting (4.2), energy saving (4.3), the integration of mobile consumption devices (4.4) and the smart grid technology (4.5) are the most relevant drivers. They result directly in typical challenges for data management technology and processes, especially as far as data integration, data analytics, data storage and scalability (including flanking aspects like multi-tenancy and parallelized processing) are concerned. The relationship between the database and market challenges is demonstrated in Figure 8 and will be explained more in detail in the following subsections.

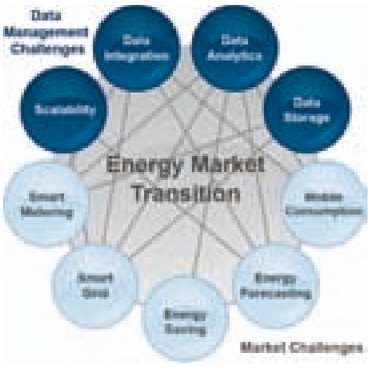

Figure 8: Relations between Market- and Data Management Challenges

4.1 Smart Metering

Many of the meters installed in households and small businesses still do not have a communication module. Load curves are not available and consumption values have to be retrieved manually in several periods, usually once or twice a year. Automated meter reading installations based on one-way communication are able to automatically transmit consumption data to a centralized database system. Making use of two-way data communication protocols enables additional features, like networked meters or remote control applications like meter status verification or on-demand data reading. This is meant by Smart Meter solutions and is widely seen as the technological key element for future smart grids (see Section 4.5).

A topic not limited to liberalized European markets, smart meters are being introduced in many of the world's developed countries. Europe, driven to a large extent by regulations, had an early start in Italy in the 2000s, soon followed by full-scale rollouts in Scandinavia and other countries [Ryb12]. North America has actually the world's highest penetration of automatic meter reading, while the Asia-Pacific region is still situated in an early stage of the adoption process with Japan and South Korea moving ahead.

Data Management Challenges

The major issue with smart metering technology is the handling of massive data amounts flooding the central data management systems on a daily basis. Instead of one single value a year, automated meters protocol load curves with a measurement rate of 5 to 60 minutes (corresponds 8760 to 105K values/year) depending on national regulations. The data is transmitted once a day or even with an intraday frequency, thus creating a challenge even for advanced *data storage* capabilities and compression algorithms.

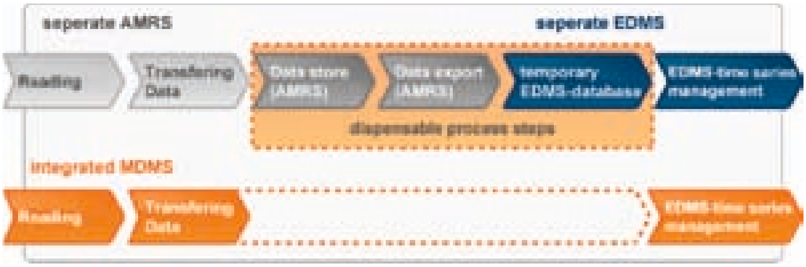

Figure 9: Data transfer process optimization using integrated MDMS

As one attempt in order to reduce data transaction time and typical *data integration* problems, we observe the consolidation of automated meter reading systems and EDMS to *Meter Data Management Systems* (MDMS), integrating both meter reading and energy data management functionality. The goal is to shorten the transaction process chain introduced as load profile processing in Section 3.2 (compare Figure 5). This is done by eliminating interim steps caused by data transfer between independent systems, thereby at the same time sparing storage space for redundant raw data archiving as demonstrated in Figure 9: Using integrated MDMS, data can be transfered directly to the time series database. These consolidations have two possible development directions: (1) Adding a meter reading module directly to the EDMS framework or (2) implementing EDMS components in the AMRS, each eliminating the importance of the respective opposing system.

Apparently, smart meters are cost-intensive compared to conventional metering technologies. Functions like automatic firmware updates or remote meter locking, such as a comfortable fraud prevention measures, are useful for the meter operators. However, the use of recorded data is still limited to energy delivery and billing purposes, thus not providing any advantage for private end-supply customers. That is why innovative concepts are required to make advanced meter technologies more attractive. Using advanced *data analytics* methods to support the creation of generic, consumption-specific customer profiles allows the offering of new products, for example flexible tariff contracts based on real-time energy prices. Providing value-adding features like networking smart meters to create completely integrated smart home solutions, as recommended in [GV12] and demonstrated in the form of a prototypal implementation by [JJP$^+$10], can also make a difference.

Another aspect is growing legal discussions concerning data security in a smart meter gateway, where some important rules have to be kept in mind. Consumption data must be protected using secure software architectures and must be processed in anonymized or clustered form to prevent any possible conclusions of a single customer's consumption behavior. Furthermore, data encryption might be applied to protect the confidentiality of the customer requiring *efficient query processing* techniques on such data (e.g., [HILM02]).

4.2 Energy Forecasting

Even the smartest meter technology can only provide a recorded view on past situations, but the key to balance an energy distribution network successfully is to plan actions foresightful by predicting as many of the most influencing (correlated) parameters for operations as possible. For both demand and supply forecasts the same generalized model [Has07] can be applied (Figure 10).

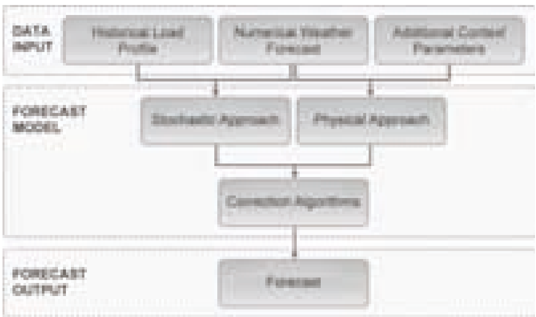

Figure 10: General energy forecast approaches

On the input side, either *historical load profiles* or *context parameters* like a supply installation's total capacity and transformation effectiveness must be provided. In case of weather-aware forecast models, the parameters must be combined with *numerical weather forecasts*, usually obtained from contracted metrological services.

After input data preparation, a forecast model must be selected and applied. *Physical Prediction* models rest upon an installations physical and technical character. Not depending on the availability of measured load curves, these approaches can quickly be implemented, making them indispensable especially for new installations, where no historical data is available. However, their practical usefulness is limited, as DSOs normally do not have access to all required technical details. In contrast, *Stochastic Prediction* models (e.g. regression, exponential smoothing) are based on time series analyses. The advantage of their employment is the direct availability of historical load profiles from the time series database and the maintenance of data assured by regular updating processes (compare Section 3.2).

Finally, different data *Correction Algorithms* like online validation of model output might be applied and the forecasted data is computed. The described forecasting elements can be combined and used in an iterative way.

Data Management Challenges

Demand and supply forecasts with high accuracy are crucial to reduce the overall energy mismatch and penalties paid for any kind of imbalances, requiring sophisticated *data analytics* approaches.

On the demand side, very good results can already be achieved based on historical data, user experience and the application of simple mathematical functions [TE08]. Domestic and industrial consumption behavior usually follows certain, often repeated patterns, and dependence on exogenous influences is limited to outside temperatures, affecting only heating or cooling installations.

In contrast, the supply side is much more challenging and becoming increasingly important. In 2010, the total share of renewables in the global electricity production had already reached 19.4%. However, as many countries still do not accomplish their final capacity targets [Ren12], this number is expected to increase quickly; mostly driven by additional wind and solar power installations. Both of them are numerous, decentralized and heavily depending on weather conditions like wind speed and -direction, global heating or cloud coverage. Especially the last circumstance adds a new dimension to energy forecasting and is the main reason why in the past years a lot of research has been conducted to improve prediction quality. Advanced approaches like the appliance of ensemble models to increase robustness and accuracy are developed.

The need for fast response times to react to new market situations as well as continuous streams of new demand and supply measurements leads to additional *data integration* and *scalability* challenges. A tight coupling and the integration of energy forecasting within the databases avoids data transfer and enables the usage of existing database optimization structures and techniques [FRL12]. The creation of complex and multiple energy models requires optimizations to handle real-time mass prediction processes. Pre-calculation and materialization of stochastic models allows for fast output generation but also requires the development of efficient maintenance strategies as new data becomes available [DBLH11]. Finally, the granularity of the forecast models (e.g., single wind installations vs. complete regions) needs to be carefully selected as it strongly influences the quality as well as the efficiency of the output.

4.3 Energy Saving

Besides the integration of decentralized and/or renewable energy sources, the efficient use of energy is considered to be the second key component of a sustainable energy policy. All technical and organizational efforts to reduce primary energy consumption with the motivation of reducing costs or emissions can be summarized by this term. These goals are mainly achieved by adopting more efficient industrial production technologies or processes [Die07], affecting all stages of the energy chain: generation, transformation, distribution and final consumption. Most of the introduced measures focus on the public transport and building sectors, where the highest potential for savings is expected [Eur11b].

In many markets, particularly in those with low primary energy prices, customers are still unaware of existing energy saving propositions. The smart meter roll-out (compare Section 4.1) is expected to encourage consumers to manage their energy use more efficiently. As for the EDMS, we will concentrate on providing the informational base and analytical equipment needed to (1) identify any possible saving potentials, to (2) measure the impact of adopted improvements on energy consumption and (3) for permanent process monitoring. These are the key activities required in order to establish active private or corporate energy controlling and waste prevention policies.

Data Management Challenges

A typical *data analytics* use case can be found in retailing companies like supermarket chains. Due to centralized management and procurement, all retailers are supposed to have equal technical equipment and to follow the same daily operational processes. To eliminate most of the external influences, store clusters have to be defined based on context information like geographical regions and selling areas in use or average yearly revenues. Now in theory, all markets in a cluster should have identical load curves but in practice they will not, as there are always variances. Storage managers can compare their energy consumption patterns within a cluster to identify process problems and energy saving opportunities.

Consider the real-world example in Figure 11, where the values of four different load curves in one cluster are visualized, representing the best, average, worst energy consumption in the concerning cluster. Additionally, the load curve of a single store (me) is shown. Energy saving potentials are marked whenever the energy consumption of the store strongly differs from the reference curves. Data mining approaches like clustering or time series similarity, e.g. as demonstrated in the work of Misiti et al [MMOP10], help to automatically identify such saving potentials even on single appliance level.

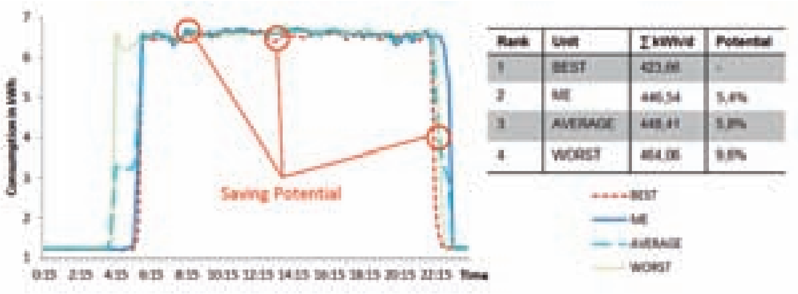

Figure 11: Exemplaric daily load curve cluster analysis

4.4 Mobile Consumption Devices

There are many research and development activities going on in the area of mobile consumption of electrical energy, also called *eMobility*. These efforts mainly rely on ecological aspects. Governments are trying to implement cleaner private and public transportation systems by using the prospective surplus production of renewable energy instead of climate-damaging, dwindling fossil fuel resources.

Data Management Challenges

Generally, information and communication technology is seen as the key enabler for eMobility, offering a multitude of basic and advanced services. The integration of these services will allow the usability for the end user without regional limitations, for example the adaption of roaming processes already known from the telecom business or recharging location services. Before any large-scale eMobility rollouts can take place, trans-regional standards have to be established for suitable infrastructures, particularly concerning regulatory frameworks, physical charging equipment and above all, the underlying information technology. The entire concept is dependent on a connected, interoperable, flexible back-end application and infrastructure system that is simultaneously used for data recording, as well as a platform for the future implementation of value-added services. Case studies like [IT12] show that a robust, open architecture is essential for the success of an eMobility project, concerning both consumers and service providers. Users must be automatically identified and data from charging stations has to be collected in order to assess the driving pattern and charging behavior of each participant. With the corresponding *data analytics* methods, the information can be used by the utility company for (1) billing purposes, (2) to create demand forecasts for mobile energy (Section 4.2) and (3) to develop attractive pricing models like time-of-use tariffs or monthly flat fees.

On-Board Metering is one approach to reduce these immense infrastructure spendings. Instead of the expensive installation and maintenance of meters and communication units in dedicated public charging stations, the vehicles will be equipped with all required measuring devices [Lan12]. This enables vehicle-to-grid communication and initiates the transformation of today's static meter point installations into mobile ones, which are no longer assigned to a specific distribution network and therefore harder to handle by the DSOs who follow conventional energy distribution processes. The *data integration* challenge here is that the contract entity can not longer be statically related to an explicit meter point, and the meter point's current location will be unknown. Although representing a new problem in the energy domain, this is an issue already being addressed by the research topic of *moving objects databases*, resulting for example in dedicated applications as such for cellphone providers or public transport information systems (e.g., [WSX^{+}99]), wherefrom related concepts and ideas might be adopted.

4.5 Smart Grids

Finally, we turn our focus on the *Smart Grid*, one of the energy sector's latest fashionable terms, frequently used by scientists, industrial leaders and politicians. The general vision is to optimize the allocation of limited energy resources by using modern information and communication technology.

Figure 12: Combined networking elements in a smart distribution grid

A smart grid can briefly be described as a technically implemented demand response concept, achieved by connecting and monitoring conventional and renewable energy producers, consumers and flexible storage units as demonstrated in Figure 12. This is done via an intelligent interactive network, principally based on smart meters (Section 4.1) and internet technology. The idea behind this is to match energy supply and demand within the grid, by using directed, short-term interventions coordinated by a data management system which is serving as the centralized brain or steering unit. This is the point where we consider the EDMS with an integrated AMR module as a possible option to take part in.

Data Management Challenges

One challenge is to scale down its level of operations from complete TSO or DSO networks to smart grids or even further, to single smart homes forming a combined energy supply micro-system. At first appearance, this might not be an appropriate use for the complex EDMS design idea, but using *multi-tenancy* optimization to intelligent share and allocate the system resources allows the simultaneous management of many parallel or connected grids within the same database system.

Another important element of the smart grid we have not discussed yet is the energy storage unit. With the increasing share of fluctuating renewable energy sources, the resilience of distribution networks will be tested. The ability to store and allocate produced energy accordingly is an important matter to maintain network stability in peak situations and to enable possible load shifting options, but still remains difficult due to the physical characteristics of electrical power. Either classical pump storage stations or modern air compression storages are restricted to accordingly conditioned geographical regions; today's common batteries are slow-charging and low-capacity units and therefore not really able to create a significant impact on the daily load balance.

By using the smart grid communication network, all of these usually isolated micro-units can be connected in a cloud. Combined with small and flexible cogeneration units, they form a virtual power plant which can be remotely controlled by the central data management unit [PRS07]. In order to ensure the successful coordination of hundreds or thousands of such small supply and storage units, all data streams must be real-time transmitted and analyzed instantly (*data analytics*). In first place this allows an on-line validation of pre-generated load curve forecasts, followed by immediate reactions in the form of executing corresponding events like additional supply orders.

Second, the *integrity* of processed master and transaction data must be monitored at an absolutely high level to keep track of every single grid element's status. This implies plenty of work for the communication gateway, which has to manage the synchronization between the central system and all meters and data concentrators in the grid, so the modules scalability combined with parallelized data processing by the central steering and control unit is a critical factor.

5 Conclusions

In the above sections we have shown that currently available EDMS are highly specialized, standardized and powerful database applications, capable to handle mass data exchange processes and thereby satisfying most of today's energy market's requirements. However, along with the ongoing energy market's transformation, new technical challenges remain to be solved. These are basically bundled in the terms of data analytics, data and/or system integration using adequate communication technology, advanced data storage and compression techniques, and scalability optimization approaches.

For future developments, we think that additional aspects, such as data security and robustness of the chosen architectures need to be addressed. Furthermore, we consider transaction process optimization and the integration of new standard core components into the centralized EDMS framework to be the most promising implementation directions, thereby creating a close link between data and functionality. These will enable the system to handle operations more efficiently which is critical for future success, like discussed above against the background of automatic meter reading and forecasting functionality. Besides, as exemplarily described in association with mobile energy consumption, there exists the possibility of adopting existing solutions obtained while treating similar problem constellations in different market domains. After all, energy data management is a promising field for database researchers and developers, offering many opportunities for future improvements in form of interesting challenges.

Acknowledgment

Part of the work presented in this paper was funded by the European Regional Development Fund (EFRE) under co-financing by the Free State of Saxony and Robotron Datenbank-Software GmbH.

References

[BDD+12] M. Böhm, L. Dannecker, A. Doms, E. Dovgan, B. Filipic, U. Fischer, W. Lehner, T. B. Pedersen, Y. Pitarch, L. Siksnys, and T. Tusar. Data management in the MIRABEL smart grid system. In *EDBT/ICDT Workshops*, pages 95–102, 2012.

[DBLH11] L. Dannecker, M. Böhm, W. Lehner, and G. Hackenbroich. Forcasting evolving time series of energy demand and supply. In *Proceedings of the 15th international conference on Advances in databases and information systems*, ADBIS'11, pages 302–315, Berlin, Heidelberg, 2011. Springer-Verlag.

[Die07] M. Diesendorf. *Greenhouse Solutions With Sustainable Energy*. NewSouth Publishing, 2007.

[Eur11a] European Commission. Energy Roadmap 2050. http://ec.europa.eu/energy/energy-2020/roadmap/index_en.htm, 2011.

[Eur11b] European Commission. Proposal for a Directive on energy efficiency. http://ec.europa.eu/energy/efficiency/eed/eed_en.htm, 2011.

[FRL12] U. Fischer, F. Rosenthal, and W. Lehner. F2DB: The Flash-Forward Database System. In *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*, pages 1245 –1248, april 2012.

[GGFS11] V. Giordano, F. Gangale, G. Fulli, and M. Sánchez. Smart Grid projects in Europe. http://ses.jrc.ec.europa.eu/smart-grid-catalogue, 2011.

[GV12] T. Goette and K. Vortanz. Smart Home als Schlüsselrolle für Smart Metering. *e—m—w Zeitschrift für Energie, Markt, Wettbewerb*, 3:10–13, 2012.

[Has07] B. Hasche. Analyse von Prognosen der Windgeschwindigkeit und Windstromeinspeisung. Technical report, IER Institut für Energiewirtschaft und Rationelle Energieanwendung, Universität Stuttgart, 2007.

[HHM11] M. Hashmi, S. Hanninen, and K. Maki. Survey of smart grid concepts, architectures, and technological demonstrations worldwide. In *Innovative Smart Grid Technologies (ISGT Latin America), 2011 IEEE PES Conference on*, pages 1 –7, oct. 2011.

[HILM02] H. Haciguemues, B. Iyer, C. Li, and S. Mehrotra. Executing SQL over Encrypted Data in Database-service-provider Model. *Proc. of the ACM SIGMOD Conference on Management of Data*, June 2002.

[IT12] C. Isernhagen and D. Tarafdar. Best Practices: e-Mobiliy- Challenges and Opportunities: A Case Study from Singapore., 2012.

[JJP+10] M. Jahn, M. Jentsch, C.R. Prause, F. Pramudianto, A. Al-Akkad, and R. Reiners. The Energy Aware Smart Home. In *Future Information Technology (FutureTech), 2010 5th International Conference on*, pages 1 –8, may 2010.

[JP05] T. Jamasb and M. Pollitt. Electricity Market Reform in the EU: Review of Progress toward Liberalization & Integration. *The Energy Journal*, 26:11–41, 2005.

[Lan12] B. Lange. Unter Spannung: Die IT hinter der Lade-Infrastruktur für E-Autos. *iX Magazin für professionelle Informationstechnik*, 7:87–93, 2012.

[MMOP10] M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi. Optimized clusters for disaggregated elecricity load forecasting. *REVSTAT*, 8:105–124, 2010.

[NPS+12] J. Nitsch, T. Pregger, Y. Scholz, T. Naegler, D. Heide, D. Luca de Tena, F. Trieb, K. Nienhaus, N. Gerhardt, T. Trost, A. von Oehsen, R. Schwinn, C. Pape, H. Hahn, M. Wickert, M. Sterner, and B. Wenzel. Long-term scenarios and strategies for the deployment of renewable energies in Germany in view of European and global developments. http://www.dlr.de/tt/Portaldata/41/Resources/-dokumente/institut/system/publications/leitstudie2011_kurz_en_bf.pdf, 2012.

[PBB+09] E. Peeters, R. Belhomme, C. Batlle, F. Bouffard, S. Karkkainen, D. Six, and M. Hommelberg. ADDRESS: Scenarios and architecture for Active Demand development in the smart grids of the future. In *Electricity Distribution - Part 1, 2009. CIRED 2009. 20th International Conference and Exhibition on*, pages 1 –4, june 2009.

[PRS07] D. Pudjianto, C. Ramsay, and G. Strbac. Virtual power plant and system integration of distributed energy resources. *Renewable Power Generation, IET*, 1(1):10 –16, march 2007.

[Ren12] Renewable Energy Policy Network for the 21st Century. Renewables 2011 - Global Status Report. http://www.ren21.net/REN21Activities/Publications/GlobalStatusReport/-GSR2011/tabid/56142/Default.aspx, 2012.

[Ryb12] T. Ryberg. Smart Metering in Western Europe, Seventh Edition. Technical report, Berg Insight, 2012.

[Sto12] D. Stolarski. Marktüberblick Energiedatenmanagement. *e—m—w Zeitschrift für Energie, Markt, Wettbewerb*, 3:99–109, 2012.

[TE08] J. Taylor and A. Espasa. Energy forecasting. *International Journal of Forecasting*, 24(4):561–565, 2008.

[US 12] US Department of Energy. Smart Grid Investment Grant Program, Progress Report. http://www.smartgrid.gov/sites/default/files/doc/files/sgig-progressreport-final-submitted-07-16-12.pdf, july 2012.

[WSX+99] Ouri Wolfson, Prasad Sistla, Bo Xu, Jutai Zhou, and Sam Chamberlain. DOMINO: databases fOr MovINg Objects tracking. In *Proceedings of the 1999 ACM SIGMOD international conference on Management of data*, SIGMOD '99, pages 547–549, New York, NY, USA, 1999.