# Report on the Third International Workshop on Energy Data Management (EnDM 2014)

Torben Bach Pedersen
Center for Data-intensive Systems (Daisy)
Aalborg University
9220 Aalborg, Denmark
tbp@cs.aau.dk

## 1. INTRODUCTION

The energy sector is in transition–being forced to re-think the current practice and apply data-management based IT solutions to provide a scalable and sustainable supply and distribution of energy. Novel challenges range from renewable energy production over energy distribution and monitoring to controlling and moving energy consumption. Huge amounts of "Big Energy Data," i.e., data from smart meters, new renewable energy sources (RES–such as wind, solar, hydro, thermal, etc), novel distributions mechanisms (Smart Grid), and novel types of consumers and devices, e.g., electric cars, are being collected and must be managed and analyzed to yield their potential. Energy is at the top of the worldwide political agenda. For example, The European Union has stated the "20-20-20 goals" (20% renewable energy, 20% better energy efficiency, and 20% CO2 reduction by 2020). Even more ambitious goals are set for 2030 and 2050. This situation is reflected in research funding schemes such as Horizon 2020 as well as national programs. Increasingly, such programs include joint calls involving both energy and IT partners. Data management is at the heart of this development.

Thus, data management within the energy domain becomes increasingly important. The International Workshop on Energy Data Management (EnDM) focuses on conceptual and system architecture issues related to the management of very large-scale data sets specifically in the context of the energy domain. The overall goal of the EmDM workshop is a) to bridge the gap between domain experts and data management scientists and b) to create awareness of this emerging and very challenging application area. For the workshop's research program, the organizers especially try to attract contributions that push the envelope towards novel schemes for large-scale data processing with special focus on energy data management.

The Third International Workshop on Energy Data Management (EnDM'14)[1] was held in conjunction with EDBT/ICDT 2014 in Athens, Greece, on March 28, 2014. This half-day event brought together researchers and engineers from academia and industry to discuss and exchange ideas related to energy data management and related topics. The workshop featured five research papers, and finished off with a panel/roundtable discussion. The accepted papers spanned a number of exciting topics within energy data management. Three papers concerned modeling of energy data: a pipeline production data model, a semantic web ontology for renewable energy sources, and an ontology for prosumer-oriented smart grids. Two papers were on energy analytics: one on extracting energy consumption profiles from smart meter data and one on a benchmark for renewable energy forecasting systems. The workshop proceedings have been published in a joint volume of all EDBT/ICDT 2014 workshops [1].

## 2. RESEARCH PAPERS

The first paper "Pipeline Production Data Model" by Jitao Yang, Yu Fan, Yinliang Liu, Hui Deng, and Yang Lin proposed a data model for pipeline production that could be used to support planning, scheduling, distribution, metering, energy consumption as well other busines processes within pipeline production. The model was specified in a function-free first-order predicate language. The model could be queries using a corresponding formula language. Two implementations of the model in a pipeline production system were discussed, one with Datalog queries on top of relational storage, and another based on RDF. Only the first had been realized so far: for this system, the functional, system, and software architectures were discussed, along with the hardware used for deployment. The system is running in production, and will be extended to in-

[1] http://endm2014.endm.org

tegrate data from a number of company systems.

The second paper "Renewable Energy Data Sources in the Semantic Web with OpenWatt" by Davide Lamanna and Antonio Maccioni proposed the OpenWatt ontology which is meant to solve a number of problems in the renewable energy sector including data that is only partially available, noisy and inconsistent data, sparse data in heterogeneous sources, unstructured data, and data represented through non-standard and proprietary formats, making the process of using the data ineffective and error-prone. OpenWatt is based on the Linked Open Data paradigm. It proposes a methodology with the steps of data gathering, data cleansing, data modeling, data recognition, ontology definition, data generation, metadata generation, external linking, data and ontology validation, and finally data publication. The OpenWatt ontology captures the types (consumption, production, potential) and categories (wind, biomass,..) of renewable energy sources, their geo-location, organized in a hiearchy, and a description of the associated measures (what quantity was measured, how, and the data source), and allows many kinds of data to be described and linked. The paper discussed the potential impact of OpenWatt and experiences from its implementation.

The third paper "A Generic Ontology for Prosumer-Oriented Smart Grid" by Syed Gillani, Frederique Laforest, and Gauthier Picard, presented a generic and layered ontology for complex prosumer-oriented smart grids. Prosumers are a new type of entity occuring in smart grid that can both produce and consume energy, which is a paradigm shift compared to traditional central energy production, and requires a detailed model of the context. The ontology enabled the integration and management in real-time of many distributed and heterogeneus sources, along with allowing the right granularity level for information. The paper discussed a number of use cases and modeling different types of physical infrastructures, electrical appliances, electrical generation, storage, weather, events, service contracts, and components. Rule-based inductive inference was supported on top of the ontology, with patterns for appliance consumption, alternative energy production, and producer performance.

The fourth paper "Computing Electricity Consumption Profiles from Household Smart Meter Data" by Omid Ardakanian, Negar Koochakzadeh, Rayman Preet Singh, Lukasz Golab, and S Keshav presented a profiling framework for residential consumers that take the variations in electricity consumption at different times of day and at different outside temperatures into account. Concretely, the effect of the outside temperature was isolated and a time-series autoregressive model used for the residual. The profiles can be used for personalized savings recommendations, outlier detection, as well as generaring realistic synthetic data. An experiment on a real-world data set of one thousand homes showed that the approach had better root-mean-squared prediction error than competing approaches.

The final paper "ECAST: A Benchmark Framework for Renewable Energy Forecasting Systems" by Robert Ulbricht, Ulrike Fischer, Lars Kegel, Wolfgang Lehner, and Hilko Donker discussed the need for evaluating and comparing the many new energy supply forecasting techniques developed in recent years. Although more and more data sets are available, it is still difficult and time-consuming to compare techniques with each other. The paper discusses the requirements for benchmarks of energy supply forecasting techniques, followed by presenting the ECAST benchmark framework that simplies the process by automating many tasks. The framework is demonstrated on a real-world scenario comparing different forecasting tools against a naive method. Finally, the paper points to a number of future developments of ECAST.

## 3. ROUNDTABLE/PANEL

The workshop finished off with a panel/roundtable discussion on "Energy Data Management: Where Are We Headed?" The workshop organizer first asked some questions: on what was already done within energy data management, and what is still missing, what the scientific challenges are, what the technical challenges are, and what challenges that necessitate an interdisciplinary approach, and provided his personal opinions on this.

A broad range of open benchmark datasets that can be used to develop robust and effective methods for various energy data management tasks, e.g., detailed device-level datasets for a larger number of households, is missing. Scientific challenges include a) the development of robust and effective methods and techniques for prediction of energy production and consumption down to the device level; b) the development of methods capable of extracting and predicting flexibilities in energy usage; c) the development of scalable etchniques for aggregating, scheduling, and disaggregating micro-level flexibilities, e.g., in individual device consumptions, to large-scale macro-level units suitable for balancing energy supply and demand at the higher levels; On the technical level, there is still a lack of community-wide agreed-upon common definitions of data and information concepts, e.g., standardized ontologies specifying common concepts. Also, standardizing communication protocols, e.g., for available flexibilities, is very important. Interdisciplinary challenges are hard to meet, and include the interplay between computer scientists developing scalable techniques for energy data management, human-computer interaction designers exploring how and at which level of detail to interact with a smart grid system, and economists developing new business and energy taxation schemes.

The roundtable discussion added further perspectives. For data sets, production data is available but consumption data not yet, due to privacy problems. Data collection and generation can be bottom-up, using a home operating system controlling different appliances, but this is problematic for large appliances like central air conditioning. It can also be top-down, but this creates the problem of proper dis-aggregation. It is not feasible to put a sensor in every device, but it is possible to get a long way without dedicated HW. For privacy, consumers generally do not know how critical the data is, e.g., which movie is watched can be determined by analyzing the TV's energy consumption. A solution could be to build a virtual home, an energy data container in the cloud, where people can store their own data, and share it with whom they choose, compare with their neighbors, and give feedback. Getting utilities to release data is difficult, especially to get continued support after the initial data release. The French energy company EDF provides analysis to customers at the household level, but still not at the device level. Scientific challenges include model selection and predicting load of transformers. Technical challenges include that companies make their own proprietary ontologies, there is no standard or reuse. For interdisciplinary challenges, non-cash incentives and social networks are important. An example is the Ontario PeakSaver program, where the set point of the heating system can be changed, and the benefit is distribubed among the participants.

## 4. DISCUSSION AND OUTLOOK

Summing up, if we first look at the topics of the presented papers, we note that they address two main topics, modeling and analytics. In some sense, the three modeling papers all try to tackle the above-mentioned problem of a lack of standardization of common concepts, by proposing generic and extensible ontologies. This is an encouraging trend, which will however, have to be accompanied by a consolidation and integration phase leading up to a formal standardization. The papers on analytics also try to propose generic solutions to common problems, one to the problem of customer segmentation based on the consumption profiles, and the other by providing a common automated framework for benchmarking energy forecasting techniques against each other. Benchmarks are typically a sign that an area is reaching a certain level of maturity, which hopefully is the case for energy data management. Compared to the previous workshops, the topics were more focused. Several papers describe inter-disciplinary collaborations.

Next, when looking at the topics which occurred in the Call for Papers, but not within the accepted papers, we see that about half of the topics are covered in one way or another. Missing are more systems-oriented topics such as data processing architectures and schemes, query processing and optimization, and robustness aspects. We believe this is because energy data management is still new, and thus most systems are still in the development phase. While most papers are based on small case studies, only one paper described a system running in industrial production. We again attribute this to the fact that smart grids are still in development.

Summing up, we conclude that there is a lot of interesting work going on in the area of energy data management, with many remaining challenges to be met. This supports the need for venues that focus on this issue. The EnDM workshop series will continue at EDBT/ICDT 2015 in Bruxelles where the 4thInternational Workshop on Energy Data Management will be held on March 27, 2015[2].

## 5. ACKNOWLEDGEMENTS

The EnDM 2014 Chairs would like to thank all the authors of submitted papers for their support for the workshop and the high paper quality and the distinguished PC members for their careful and dedicated work during the reviewing and discussion phases. Finally, we would like to thank the Organizing Committee of EBDT/ICDT 2014, especially General Chair Vassilis Christophides, Workshop Chair Selcuk Candan, and Proceedings Chair Vincent Leroy. for their support.

## 6. REFERENCES

[1] K. S. Candan, S. Amer-Yahia, N. Schweikardt, V. Christophides, and V. Leroy, editors. *Proceedings of the Workshops of the EDBT/ICDT 2014 Joint Conference (EDBT/ICDT 2014), Athens, Greece, March 28, 2014*, volume 1133 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.

[2] http://endm2015.endm.org