Design of a Management Infrastructure for Smart Grid Pilot Data Processing and Analysis

Matthias Strobbe, Tom Verschueren, Stijn Melis, Dieter Verslype, Kevin Mets, Filip De Turck, Chris Develder
Ghent University - iMinds
Department of Information Technology
Gaston Crommenlaan 8 bus 201, 9050 Gent, Belgium

Contact author: Matthias Strobbe (matthias.strobbe@intec.ugent.be)

Abstract—**Future smart grids will combine power grid technologies with information and communication technologies to enable a more efficient, reliable and sustainable energy production and distribution. To realize such a smart grid, large scale pilot projects are currently implemented and evaluated. Such pilot projects generate an excessive amount of data that needs to be processed: energy measurements, information on available flexibility from smart devices that can be shifted in time, control signals, dynamic prices, environmental data, etc. To validate and analyze the gathered data and adjust the running experiments in real-time, an optimized data management infrastructure is needed as well as comprehensive visualization tools.**

In this paper we present a data management infrastructure optimized for the follow up of the large scale smart grid project called Linear. In this project a pilot in over 200 households is implemented to evaluate several business cases including the balancing of renewable energy supply and the mitigation of voltage and power issues in distribution grids. By decoupling the gathering of the incoming data, the processing and storage of the data, and the data visualization and analysis on different servers, each with their own, optimized database, we obtain an efficient system for validation of the generated data in the pilot, management of the deployed set-ups and follow-up of the ongoing experiments in real-time.

I. INTRODUCTION

The power grid is moving away from the current centralized power generation paradigm. With governments promoting local renewable power generation at residential sites, distributed power generation is gaining in popularity. Environmental concerns and efforts to become less dependent on fossil fuels are the driving force for the replacement of traditional energy sources by green alternatives. The EU 20-20-20 targets aim for a reduction in EU greenhouse gas emissions of at least 20%, 20% of EU energy consumption to come from renewable energy sources, and a reduction in energy consumption of 20% by 2020 [1].

Renewable energy sources, such as solar or wind, offer a greener solution compared to more traditional energy sources such as fossil fuels (i.e. coal). However, their intermittent nature makes it difficult to balance demand and supply, which is essential for the correct operation of the power grid. Additionaly, electricity demand is rising, e.g. as a result of the ongoing electrification of the vehicle fleet and the adoption of heat pumps.

A future smart grid will enable a more efficient (both economical and energetic), reliable and sustainable energy production and distribution by combining power grid technologies with information and communication technologies .

Smart grids support bidirectional real-time exchange of both energy and information. As a result the end users have access to more accurate and timely information regarding their energy consumption and to several options regarding different tariff structures, which enables demand side integration and an improvement of the energy-efficiency. To implement the new structure for sustainable energy supply on a large scale in Flanders by 2020 (and beyond), a transition is necessary with short term action points, that are however based on a mid and long term strategy. The breakthrough project Linear (Local Intelligent Networks and Energy Active Regions) is a first crucial step in this transition towards Smart Grids. The project focuses on the realization of a technological and implementation breakthrough by innovative technological research and a large scale pilot in a residential area. All this in close collaboration with industrial partners and associated Flemish innovation platforms [2].

The aim of the project is to activate households in order to boost retail energy markets and increase efficiency and sustainability in an economically feasible and societal acceptable way. The research and implementation activities consider technical innovations (both on electrical network technology as the controlling ICT infrastructure, architectures and protocols), but also incorporate business modelling and user acceptance analysis.

The Linear project consists of two parts: one focusing on research and development - which aims at technological breakthroughs - and one primarily focusing on an implementation breakthrough by the preparation and set-up of a residential pilot in over 200 households to evaluate several business cases concerning portfolio management, wind balancing, reduction of peak loads on transformers and the mitigation of voltage issues on feeder lines.

In this paper we focus on the set-up of this residential pilot and especially on the data management aspects as a large amount of data will be generated during the pilot: energy measurements from smart and regular devices, status and metadata information of the smart devices (e.g. configuration time, deadline and actual start time for a remotely controllable washing machine, but also the used washing program and corresponding power train), dynamic prices and control signals

978-3-901882-50-0

to start and stop the smart devices, temperature and current measurements from transformers, voltages measurements at the connection point between house and feeder, forecasted and actual weather data, etc.

To analyze the gathered data and to be able to adjust the running experiments in real-time, an optimized data management infrastructure is needed as well as good visualization and analysis tools. In the following section we first discuss in more detail the set-up and objectives of the pilot and motivate why an optimized data management infrastructure is needed. In sections III and IV we discuss in detail the designed and implemented data management architecture and data visualization and analysis tool. In section V we discuss related work and we state our conclusions in section VI.

II. PILOT DETAILS

The goal of the residential pilot of the Linear project is to test and evaluate active demand concepts on a large scale and in a realistic setting. Therefore a set-up consisting of smart and sub meters, smart household devices, home gateway and touch display is rolled out in more than 200 households. Figure 1 shows an example of a such a set-up. A smart meter measures the total consumption and (if applicable) solar production of the household. Energy measurements of smart and regular devices are provided via smart plugs which use Zigbee [3] for their communication. Every household contains one or more smart devices (washing machine, tumble dryer, dish washer, domestic hot water buffer (DHW)) which can be controlled remotely. Inside the home, Zigbee and powerline communication (PLC) is used for transferring control signals between the home gateway and the controllable devices. Users are asked to provide as much flexibility as possible when configuring their whitegood devices, for example by configuring these devices before going to work with a deadline at the time they return home. The users also receive an in-home display showing dynamic prices or the accumulated amount of provided flexibility, depending on the specific experiment they are involved in.

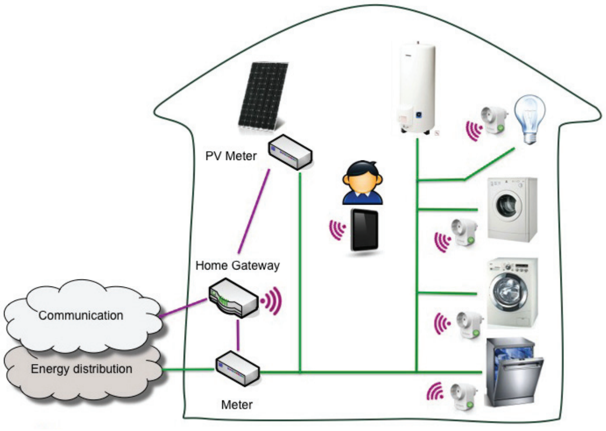

Fig. 1. Household set-up for the Linear pilot.

The pilot is implemented in 3 phases with rising numbers of users and devices, and an evaluation of increasingly complex business models. Phase 1 and 2 are geographically spread over the region of Flanders. In these phases business cases on portfolio management and wind balancing will be evaluated. The goal of the portfolio management case is to shift energy consumption as much as possible towards off-peak hours and periods with an abundant amount of renewable energy. The wind balancing business case is important for energy suppliers that need to balance their energy production with the anticipated consumption of energy. Wind farms in their portfolio make this balancing exercise more difficult as not only the consumption might be different than anticipated but also the production. Instead of paying imbalance costs to the TSO or adjust gas production plants, adjusting the demand might be a cheaper alternative.

Phase 3 is implemented in a few selected neighbourhoods with several participating households connected to the same feeder and transformer. This allows to test location specific business cases such as the reduction of peak loads on transformers and the mitigation of voltage violations on a feeder line, for example caused by the injection of a high amount of solar energy on sunny summer days. These business cases are especially important for distribution network operators as the lifetime of their assets is increased and investments in the distribution grid might be deferred or reduced.

The set-ups for phase 1 and 2 are rolled out in 2011 and begin of 2012. Roll-out of phase 3 started in the autumn of 2012. The actual experiments will start in the summer of 2013, but all currently installed set-ups already generate reference energy measurements which will be used to assess the effect of the experiments on the consumption patterns of the involved households.

Two types of control strategies will be used for evaluating the four business cases:

- Time of Use (ToU) tariffs: dynamic tariffs (applying to fixed time periods of the day (night, morning, noon, afternoon, evening, late evening), but different from day to day) are sent one day ahead to the households and shown on a display to motivate the users to shift their consumption towards cheap periods (manual ToU) or used by an intelligent gateway to control the smart devices (whitegoods, DHW) in the house in an automatic way (automatic ToU). This control strategy will be used for evaluating the portfolio management business case.
- Automatic active demand: maximum production and consumption limits are sent in real-time to individual households or clusters of multiple households. The latter allows a more optimal control of the offered flexibility as all devices in a cluster can be controlled individually. This control strategy will be used for the evaluation of the three other business cases.

The pilot is coordinated via a central server responsible for gathering all generated data and metadata, and sending control signals and dynamic price information to the individual households and devices. This is illustrated in figure 2. All gateways and plugs are managed by commercial home gateway providers, each using their own set of devices and protocols inside the homes. Every home gateway provider is responsible for aggregating all metering and device metadata for the homes

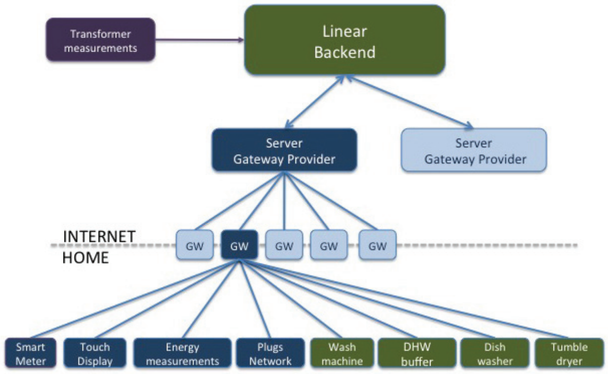

Fig. 2. Overal pilot set-up.

TABLE I. OVERVIEW OF GENERATED DATA IN THE LINEAR PILOT.

Device	Data	Frequency
Smart meter	Overall consumption and production measurements, voltage measurements at connection point between house and feeder	Every 15 minutes
Smart plug	Consumption measurements and device status info	Every 15 minutes
Gateway	Device status info	Every 15 minutes
Whitegood device	Configuration time, deadline, estimated power train, selected program, device status, actual start time, start type	On every configuration
DHW buffer	Status, mode, state of charge, temperature, energy and power parameters	On state change
Linear server	Dynamic prices	Once a day
Linear server	Control signals	Experiment dependent
Transformer	Voltage, current, cos phi and temperature data	Every 15 minutes

they control and tranferring this information via web services to the central Linear backend.

Table I gives an overview of the different types of data that will be exchanged between the households and the Linear Backend during the pilot. This amount of data is very large: an average household reports 10 to 15 energy measurements on a quarterly basis. The total duration of the pilot (reference period included) is 2 years, so more than 150 million energy measurements will be gathered over this period. Furthermore a lot of flexibility and device status data is generated which is essential to validate and evaluate the conducted experiments.

To analyze this amount of data in real-time and adjust running experiments, an optimized data management infrastructure is needed as well as good visualization tools. These will be discussed in the following sections.

III. DATA MANAGEMENT ARCHITECTURE

A. Overview

Important requirements for the central linear backend are a guaranteed availability so that no data is lost nor an experiment is interrupted, a secure storage and access to the acquired data and tools for an efficient analysis and validation of the data in real-time.

To meet these requirements an architecture consisting of several stages was designed as outlined in figure 3. The servers of the gateway providers communicate with the *Linear Live Server*. This server contains a number of web services for storing metering and metadata and for sending control signals and dynamic prices. New users and appliances can be registered and updated and reports on consumption patterns and offered flexibility can be retrieved to show on the home displays of the participants.

Manipulating the entire set of received data and calculating the control signals on the same server might cause performance problems for the web services and thus running experiments could be affected. Therefore the calculation of the control signals for a specific business case is performed on the *Linear Control Server* and all data is replicated, in real-time, on the *Linear Data Server*. This server contains all data generated over the whole duration of the pilot, while the *Linear Live Server* only contains the received data for the last 2 weeks in its database. The main functions of the *Linear Data Server* are the possibility to conduct SQL queries on the complete dataset, for example for detailed analyses of the conducted experiments, and processing and interpreting the received data by means of warehousing scripts. These scripts extract power usage patterns, aggregated reports, etc. for monitoring and analysis purposes. The results of these warehousing scripts are forwarded to a third server, the *Linear Development & Analysis Server* which runs visualization and analysis tools (see section IV). The phase 3 users already have an official smart meter in their house from a joint project of the main Flemish distribution network operators. Historical reference measurements of these users are directly imported in the *Linear Data Server*.

B. Availability

Both the *Linear Live Server* and *Linear Development & Analysis Server* are optimized for availability. These dedicated servers are hosted in a professional datacenter with a very high guaranteed uptime assuring that no data from the pilot is lost.

The *Development & Analysis Server* also acts as a complete mirror of the *Live Server* providing the same web services and a replication of its database using the same real-time synchronization process as between the *Live Server* and *Data Server*. This allows a flawless fallback in case of a problem on the *Live Server*.

Furthermore, the *Linear Development & Analysis Server* is also used as development server to implement and test the needed interfaces and control logic for conducting the actual expirements (starting in the summer of 2013), while reference measurements are already supplied to the backend at this moment.

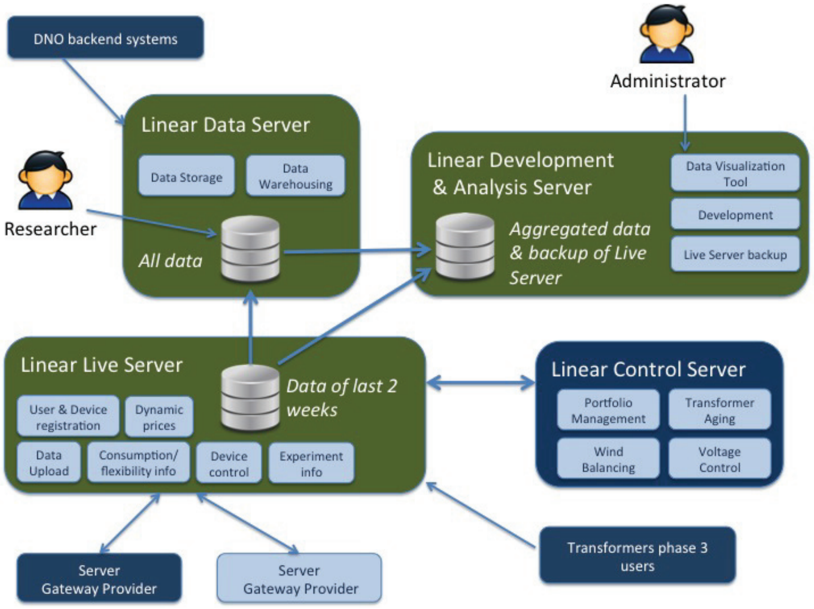

Fig. 3. Overview of the data management architecture.

C. Security

To provide secure communication between the data reporting clients of the gateway providers and the *Linear Live Server*, we rely on the Web Services Security (WS-Security) specification [4]. Clients are required to encrypt all traffic using certificates issued by our Public Key Infrastructure (PKI). This way a system is created with authenticated hosts, exchanging confidential, non-reproducible messages.

All servers are protected by a firewall and the data analysis tool can only be used by a few researchers and administrators with an account. The connection between the data server and a researcher who wants to perform a query is secured using secure shell (SSH). Data replication and synchronization between the different servers is done over SSH tunnelled connections.

IV. DATA ANALYSIS TOOL

The data analysis tool is part of the *Linear Development & Analysis Server* and provides a frontend to the aggregated data provided by the scripts of the *Linear Data Server*.

The functionality of the data analysis tool is constantly extended with at this moment support for visualizing:

- Household energy measurements.
- Aggregated consumption and production patterns.
- Metadata of the smart devices in the homes.
- Status information on the received energy measurements.

The visualization of detailed energy measurements per day or per week for a specific household allows to detect problems with the installed plugs or smart devices and gives insight in the energy consumption patterns of the users. This is illustrated in figure 4 showing a screenshot of the tool with the detailed energy measurements for one household for one day. A day is divided in 96 quarters and for each quarter the measurements of the selected plugs are shown. Hovering over a specific measurement gives more details on time and reported energy value and by selecting a part of the lower graph one can zoom in on a specific part of the day.

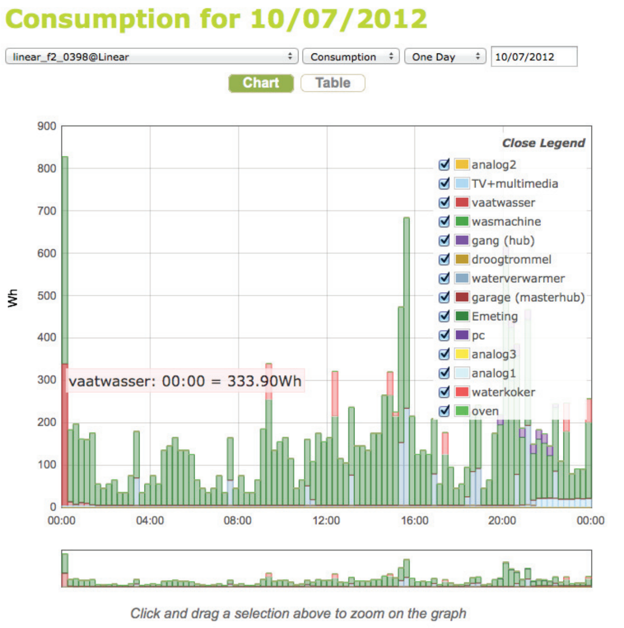

Fig. 4. Detailed consumption measurements for a single household for one day

Besides consumption and production data also detailed

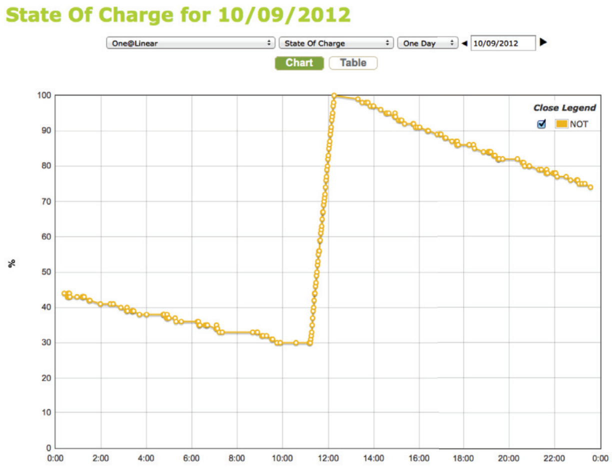

Fig. 5. Evolution of the state of charge of a boiler over 24 hours.

status information of the smart devices can be visualized such as config times, deadlines, effective start times, selected program, boiler temperature, enery and power settings, . . . As an illustration figure 5 shows the evolution of the state of charge of a boiler over the period of one day.

The tool shows aggregated consumption and production patterns per appliance type (e.g. consumption of all dishwashers or production of all PV panels) and for specific household groups. Grouping can be realized through various parameters, such as geographical location, installed appliance types, availability of local renewable energy sources, participation in a certain experiment, pilot phase, etc.

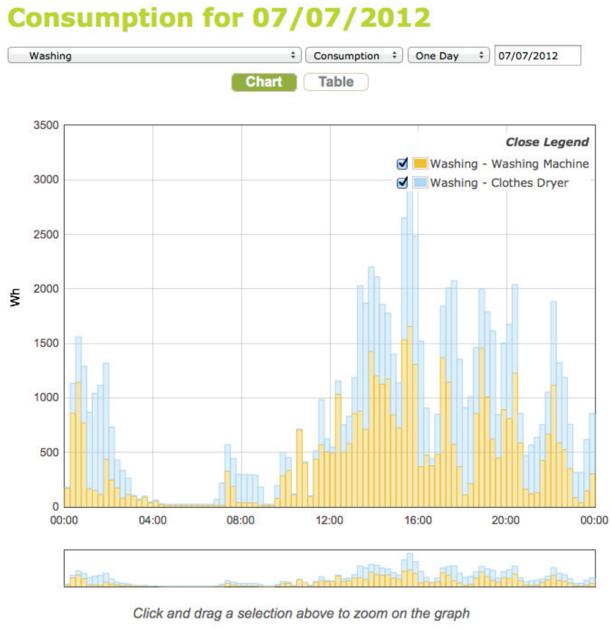

Fig. 6. Aggregation of energy consumption of all washing machines and clothes dryers for a single day

Figure 6 shows an example of the aggregated energy consumption of all registered washing machines and clothes dryers on a single day. This graph gives already some insight in the energy flexibility that can be expected during the actual experiments.

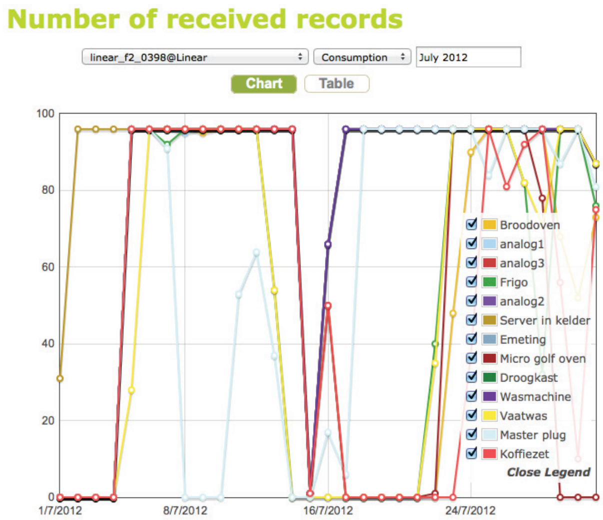

Fig. 7. Number of received energy measurements per smart plug for a single household for one month

The tool provides status information on the received information, generating warnings when less data than expected was reported. Figure 7 gives an example: the number of received energy measurements per smart plug are shown for one month, for a single household. As measurements are reported on a 15 minute basis, 96 values are expected each day. As can be seen from the graph, there were some problems with a number of plugs for this user, especially around July 16th.

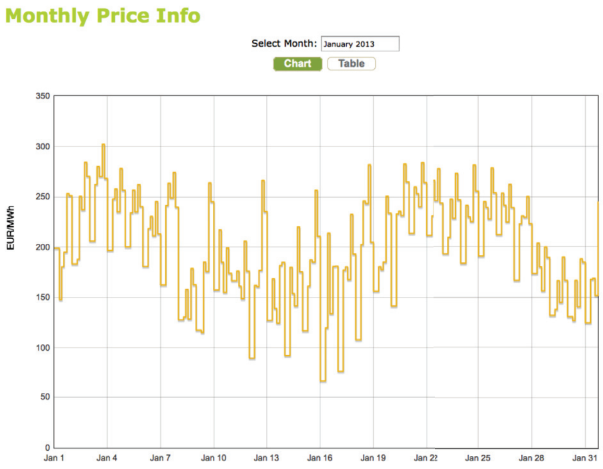

Fig. 8. Dynamic tariffs used for the experiments for a period of one month.

To follow up the current and past experiments also input data can be visualized such as the used dynamic tariffs. This is illustrated in figure 8.

Most data is reported every 15 minutes. The warehousing scripts on the *Linear Data Server* are executed at the same frequency to generate the data for the data analysis tool. As this tool employs its own database, containing only the generated data, it is very responsive and allows a real-time follow-up of the running experiments. For more specific and detailed analyses, e.g. of finished experiments, researchers can log in on the *Linear Data Server* by SSH to directly execute SQL queries on the complete dataset.

Extra functionality that will be added in the near future is the visualization of the executed control actions, and an overview of the past, running and future experiments with their corresponding parameters.

V. RELATED WORK

The research presented in this paper builds on the work presented in [5]. In that paper a first version of a data management infrastructure for the Linear pilot is presented focusing on data provided by several partners for initial analyses. This data consisted of metering and questionnaire information but was not provided in real-time and the amount of data was also considerably smaller than the data that will be generated in the actual pilot. Therefore a simpler infrastructure was used where all data was directly inserted into one large database.

A commercial party, for example a utility, will typically use commercial enterprise data management platforms from companies as IBM or Oracle to store and analyze all the gathered data from smart meters, network sensors and other devices. But these solutions are too expensive for a research project. Therefore we used the open source MySQL database [6] as basis for our data management infrastructure, as MySQL is the most used relational database with a lot of tools and documentation available.

When the setup of the pilot would be extended towards for example a whole country, NoSQL solutions such as MongoDB [7], Redis [8], Cassandra [9], etc. might be necessary as the amount of generated data will extremely huge then, but for this pilot the amount of data is not large enough to justify these technologies which still lack maturity, standard APIs and query language.

In a roll-out scenario on a large scale also other communication platforms will be needed for an efficient exchange of the data. [10] discusses a promising, cloud-based approach leveraging data-centric communication, publisher/subscriber and topic based group communication to make demand response secure, scalable and reliable. Furthermore also high performance computing tools and techniques will be needed to enable near-to-real-time state estimation across large-scale distribution networks whilst concurrently supporting on the same computational infrastructure condition monitoring of network assets and advanced network restoration solutions [11].

VI. CONCLUSIONS

In this paper we presented the Linear project which strives for a technological and implementation breakthrough of smart grids in Flanders. Therefore a large scale pilot is set-up in over 200 households in collaboration with industrial partners, research institutes and universities to implement and evaluate active demand concepts.

At this moment most of the participating households are equipped with smart meters and submeters, and smart household devices. These meters already generate reference measurements which are stored in the Linear backend. Control logic is implemented and tested at this moment and will be used from june 2013 on to evaluate several business cases: portfolio management to maximize the consumption of renewable energy, wind balancing, reduction of transformer peak loads and the mitigation of voltage violations on feeder lines.

These reference measurements already represent a considerable amount of data that is transferred every 15 minutes to the Linear backend and this will only increase in the future with the addition of device configuration parameters and flexibility information. Therefore the Linear backend should contain an optimized data management infrastructure. By decoupling the gathering of the incoming data, the calculation of control signals, the processing and storage of the data, and the data visualization and analysis on different servers, each with their own, optimized database, we obtain an efficient system for validation of the generated data in the pilot, management of the deployed set-ups and follow-up of the ongoing experiments, in real-time.

In the following months the data analysis tool will be used for extensive validation of the obtained data so far, to solve remaining issues with the current installations and follow up of the actual experiments.

ACKNOWLEDGMENT

This work is supported by the Ministry of Science (Minister I. Lieten) via the project Linear organized by the Institute for Science and Technology (IWT). K. Mets would like to thank the Agency for Innovation by Science and Technology in Flanders (IWT) for financial support through his Ph.D. grant. C. Develder is partially supported by the Research Foundation - Flanders (FWO - Vlaanderen) as a post-doctoral fellow.

REFERENCES

[1] M. Webb, *SMART 2020: enabling the low carbon economy in the information age, a report by The Climate Group on behalf of the Global eSustainability Initiative (GeSI)*, 2008.

[2] E. Peeters, C. Develder, J. Das, J. Driesen, and R. Belmans, “Linear: towards a breakthrough of smart grids in flanders,” in *Proc. 2nd Int. Conf. Innovation for Sustainable Production (i-SUP 2010)*, vol. 3, Bruges, Belgium, 18–21 Apr. 2010, pp. 3–6.

[3] Zigbee Alliance, “Zigbee smart energy,” http://www.zigbee.org/Standards/ZigBeeSmartEnergy/Overview.aspx.

[4] OASIS - Advancing open standards for the information society, “Web services security,” https://www.oasis-open.org/standardswssv1.1.1.

[5] W. Labeeuw, S. Claessens, K. Mets, C. Develder, and G. Deconinck, “Infrastructure for collaborating data-researchers in a smart grid pilot,” in *Accepted for the 3rd IEEE PES Innovative Smart Grid Technologies (ISGT) Europe Conference*, 14–17 Oct. 2012.

[6] MySQL, “Mysql, the world’s most popular open source database,” http://www.mysql.com/.

[7] 10Gen, Inc., “Mongodb,” http://www.mongodb.org/.

[8] Citrusbyte, “Redis,” http://redis.io/.

[9] The Apache Software Foundation, “Cassandra,” http://cassandra.apache.org/.

[10] H. Kim, Y.-J. Kim, K. Yang, and M. Thottan, “Cloud-based demand response for smart grid: Architecture and distributed algorithms,” in *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*, oct. 2011, pp. 398 –403.

[11] G. Taylor, D. Wallom, S. Grenard, A. Yunta Huete, and C. Axon, "Recent developments towards novel high performance computing and communications solutions for smart distribution network operation," in *Innovative Smart Grid Technologies (ISGT Europe), 2011 2nd IEEE PES International Conference and Exhibition on*, dec. 2011, pp. 1 –8.