A New Approach in System Integration in Smart Grids

J.I. Guerrero, Enrique Personal, and Carlos León
Department of Electronic Technology
University of Seville
Seville, Spain
e-mail: juaguealo@us.es, epersonal@us.es, cleon@us.es

***Abstract*—The emergent technologies related to Smart Grids provide new scenarios with new challenges. Specifically, the deployment of Smart Grid management infrastructures involves very hard scheduling, very high economic investment, and a lot of resources. Moreover, the traditional systems are based on proprietary architectures, which make more difficult the deployment process. The solution proposed in this paper makes easier the integration process of modern and old systems, in two levels, at level of metadata and data with the Heterogeneous Data Source Integration System, and at level of web services with the Web Service Integration System. Both systems are based on advanced analytics techniques, like Web Service Mining, Process Mining, Metadata Mining, Decision Support Systems, etc. Additionally, this paper establishes a test environment for the simulation of the deployment projects, called Simulation Engine. The proposed solution performed a successful integration and increased the efficiency of infrastructure in more than 30%, increasing in each iteration.**

Keywords- heterogeneous data source integration; web service mining; metadata mining; data mining.

I. INTRODUCTION

The emergent technologies related to Smart Grid (SG) are providing a new scope of functionalities and possibilities for management the power grid and increasing the available services for clients and companies. Moreover, SGs are changing the current scenario of energy markets (due to: the renewable penetration, the electricity batteries, the electric vehicles, etc.), in which the companies manage the energy in different ways. Traditionally, the energy was an utility, but, in the SG scenario, energy has turned into goods, with which the companies commercialize it.

Additionally, the SG ecosystem compounds a great quantity of systems with different standard functionalities. Some of these systems are only aggregators of data or services, to make them available for other systems. Other systems have specific functionalities, like: Energy Management System (EMS), Distributed Management System (DMS), Customer Information System (CIS), Meter Data Management System (MDMS), etc. Moreover, these systems could integrate or implement advanced functionalities at different levels of SG, for example: Building Management System (BMS), Vehicle to Grid Management System (V2GMS), Electrical Lighting Management System (ELMS), etc. Each of these systems has different needs, requires different information and services, and have different user roles (with different user interfaces).

Although, there are several technologies that provide the possibility to integrate different systems: MultiSpeak, Enterprise Service Busses (ESBs), etc. and these systems are usually accepted by standards, the main problem of these technologies is: the success of integration process is based on the compliance of several restrictions or adaptations by different systems. Currently, the systems are developed to come to terms with restrictions of these technologies. However, the deployment of the new systems related to SGs, needs several intermediate steps, that compounds the integration of traditional systems with modern systems. This integration should be in service and data contexts.

Additionally, several organizations are working on the description of Smart Grid Architecture Methodology (SGAM). Although several organizations have different models, it is possible to shape each other, providing an interpretation or comparison between them. For example, GridWise Architecture Council (GWAC) defines an interoperability stack; The Open Group's Architecture Framework (TOGAF) provides the Architecture Development Method (ADM); the European Standardization Organizations (ESO) like the European Committee for Standardization (CEN), European Committee for Electrotechnical Standardization (CENELEC), and the European Telecommunications Standards Institute (ETSI) provides an SGAM aligned to M/490 reference architecture, etc. National Institute of Standards and Technology (NIST) provides the equivalences between these different models. These methodologies provide a general vision of SG ecosystem, and it is strongly recommended to take into account in the integration process, in order to identify the different levels or layers.

Thus, in the SG ecosystems may be different types of systems with different information. The development of new systems should consider the information from old systems, in order to take advantage from the combination of an old and new information. But this process is very complex, and requires long developments and deployments. In this sense, the present paper proposes an automatic system to integrate all information and services from different systems, making available their resources to other systems in SG ecosystem.

In Section II, a bibliographic review is included. In Section III, a general architecture of proposed solution is described. In Section IV, the Heterogeneous Data Source

 ISBN: 978-1-61208-576-0

Integration System (HDSIS) is described. In Section V, the Web Service Integration System (WSIS) is proposed, with a description of all its modules. In Section VI, the experimental results are described. Finally, in Section VII, the conclusion section is included.

II. Bibliographic Review

In this case, there are two technologies related with the proposed paper, the integration of heterogeneous data sources and the integration of the system at the service level.

In the HDSIS case, there are a lot of studies and researches related to heterogeneous data integration based on, for instance, XML [1], Lucene and XQuery [2]. In the same way, heterogeneous data integration has been applied on many areas, such as Livestock Products Traceability [3], safety production [4], management information systems [5], medical information [6], and web environments [7].

There are also examples of the application of data mining mixed with Heterogeneous Data Source Integration (HDSI). These types of solutions increase the capability of solution to adapt it to different and heterogeneous data sources. [8]proposes a framework of a self-Adaptive Heterogeneous Data Integration System (AHDIS), based on ontology, semantic similarity, web service and XML techniques, which can be regulated dynamically. [9] uses On-Line Analytical Processing (OLAP) and data mining to illustrate the advantages for the relational algebra of adding the metadata type attribute and the transpose operator.

In the integration of the system at the service level case, there are several technologies based on the definition of different interfaces and standards, but the Web Service Mining (WSM) [10] is used in this paper. There are also several solutions based on process mining [11], pattern usage discovery [12], hypergraph-based matrix representation with a service set mining algorithm [13], constraint satisfaction [14], semantics-based methods [15], customer value analysis [16], frequent composite algorithm [17], Heterogeneous Feature Selection [18], etc.

III. General Architecture

The proposed solution is based on the merging of two previously published solutions:

- A solution for HDSI [19]. This solution provides the integration from different relational database or data source, providing a new data model based on information standards, and providing models for the main parameters identified in data, according to the results of the application of metadata mining and a decision support system.
- A solution for integration of web services in SG ecosystem [20]. This solution based on WSM and Swarm Intelligence, provides a way to automatically integrate the Web Service (WS) interfaces from all authorized systems, creating and configuring new WSs based on the usage of previously existing WSs.

Both solutions have been integrated, and some modules were updated to interconnect each other. Thus, the proposed solution, shown in Figure 1, provides a solution to integrate information and services, which offers the possibility to create new WSs or information based on the integration process.

Although some of the proposed modules have been previously published, several of them have been updated in order to adapt the architecture, providing additional basic and advanced functionalities. The updates and new functionalities are described below.

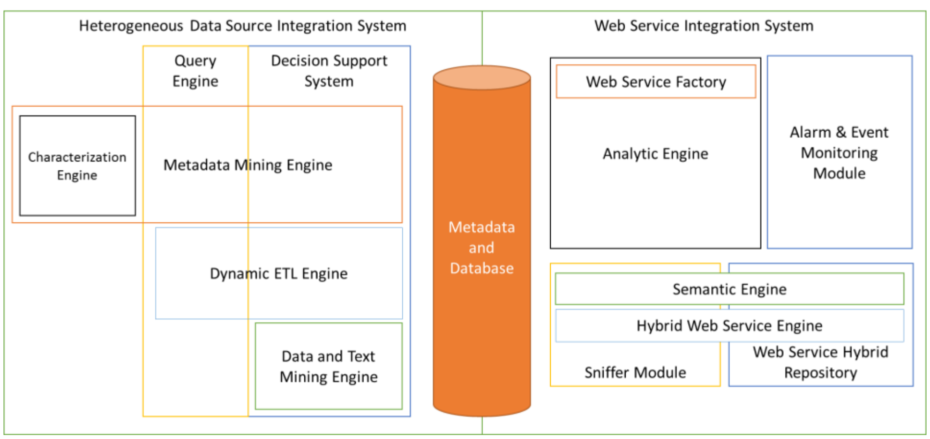

Figure 1. General Architecture.

IV. Heterogeneous Data Source Integration System

The HDSIS has several modules, mainly described in [19]. This system is based on the usage of a Decision Support System (DSS) and a Query Engine (QE), which supports different processes of Metadata Mining, Data Mining, and Text Mining.

QE is an extended library, which provides connection interfaces to different databases or data sources, allowing advanced SQL queries. Decision Support System (DSS) is implemented as a framework, which allows to manage rules, applying them to the different data sources.

 ISBN: 978-1-61208-576-0

Metadata Mining Engine (MDME) has several stages. In a first step, the metadata is extracted from each data source. In the second step, the Characterization Engine characterizes columns, tables, relationships and data sources, calculating several indicators or coefficients. The indicators or coefficients are used to classify the metadata, checking the coherence, the quality, and some other features of data and metadata. Finally, all the generated information is stored in order to integrate all metadata and data.

The Dynamic ETL (Extract, Translate, and Load) Engine (DETLE) implements an ETL with extended functionalities. This module was enhanced since the publication of the other papers. Nevertheless, this module has been modified in order to create specific ETL based on the information from Analytic Engine of WSIS to create specific and theme-oriented data warehouse, with star or snowflake structure. Initially, DETLE was a module which creates a new relational database based on information standard from International Electrotechnical Committee (IEC) or Distributed Management Task Force (DMTF) or, even, data warehouse with star or snowflake structure, according to the information from MDME. Specifically, in this new proposed solution, the DETLE has new functionalities that can generate a new ETL based on information from Analytic Engine (WSIS) and from WSs usage, or can create a new ETL according to a requested model for a specific parameter in any data source integrated in the system. This new functionality has increased the intercommunication with other modules, like Data and Text Mining Engine (DTME).

The DTME has increased their functionalities, too. Currently, this module implements an automatic modelling tool, which can model based on the information from MDME or based on information from WSIS. So, for example, a client can request by using a WS call a creation of a data warehouse for a specific parameter. When this call is fired the system retrieves all information related with this parameter (metadata and data) and tries to release the best data mining or/and text mining model, creating a data warehouse to feed it. The client can modify or update the model, or even, model again.

The HDSIS has provided additional services for the WSIS, which are included in Web Service Hybrid Repository (WSHR).

V. WEB SERVICE INTEGRATION SYSTEM

The WSIS is based on WSM and Process Mining. The first version of this WSIS was previously published in [20]. However, there are new functionalities and some modules have been merged, and others have been updated with new functionalities. In this way, the Sniffer Module integrates functionalities related with monitoring and discovering of new WSs. Traditionally, this task was made by an Ant Colony Optimization (ACO) technique. Currently, this functionality is integrated in Hybrid Web Service Engine (HWSE) and Semantic Engine.

The Analytic Engine analyzes the WS traffic based on WSM, in order to identify the sequences of WSs and their feasible relation to alarm or events in the ecosystem. Using the Web Service Factory (WSF), the Analytic Engine can create new WS with aggregated behavior of sequences of WS. Although the aggregation of WSs can be made by different ways, the proposed solution groups WS according to several features: the semantical interpretation of WS sequences, the number of invocation of the same WS sequence, and the feasible variations in a WS sequence.

The Alarm and Event Monitoring Module provides information about the external events and alarms generated in low level systems. The low level systems usually work in real time, representing from Internet of Things (IoT) devices to Intelligent Electronic Devices (IEDs).

A. *The Web Service Hybrid Repository and the Semantic Engine.*

The WSHR implements a WS repository. Additionally, this module gathers all request or notification WSs in different systems.

The Sniffer Module monitors the channel and performs the task of service discovering.

The HWSE parses all WSs. This engine provided compatibility with different Service Oriented Architectures (SOA). HWSE is a bidirectional module. When is a request WS or a sniffed WS, and it has been generated by the external system, the WS message is gathered by the corresponding module (Sniffer module or WSHR) and analyzed by HWSE and, finally, by Semantic Engine, registering all information in the internal database. If the message is generated by WSIS the message makes the inverse route, it is constructed by Semantic Engine, translated to JSON (JavaScript Object Notation) or XML (eXtended Marked Language), and sent by the corresponding module. The details of both modules are shown in Figure 2.

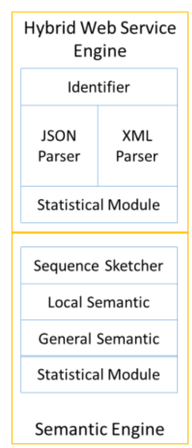

Figure 2. Hybrid Web Service Engine and Semantic Engine details.

The HWSE has different modules (Figure 2). Identifier module identifies the language of the message. According to this identification, the message is parsed by JSON or XML parser. Finally, the parsed message is enriched with additional information about statistics related to other messages or other systems. When this engine treats a message from the WSIS the modules make the inverse function.

The Semantic Engine has several modules (Figure 2), too. When the messages come from external systems, after they are treated by HWSE, they are sent to the Sequence Sketcher. This module establishes a relation in sequence call of different WS based on application of Process Mining techniques. Local Semantic is a module based on fuzzy logic, which extracts the semantic related to different parts of the message. After this step, General Semantic establishes the relation to other messages, and possible future messages. Th module is based on fuzzy and time series. Finally, in Statistical Module, several statistics and results from previous modules are registered in the WSIS database.

When the message come from the internal systems, they are treated in a reverse way. At the end, the Sequence Sketcher sends the message to HWSE. Thus, all messages in the channel are registered by the WSIS.

VI. Experimental Results

The proposed solution was tested in a little cluster of computers, with different features and functionalities. Although, the proposed solution can work in distributed environments, in this case, the proposed solution was integrated in a unique computer.

The cluster was implemented with four servers. The first server has an Intel i7 (3GHz), 16GB RAM and GTX750 (2GB and 640 CUDA cores). The second server has an Intel Xeon E5 (2GHz), 64GB RAM and Quadro K1200 (4GB and 512 CUDA cores). The third and fourth servers have the same configuration: i5 (2.7GHz) and 8 GB RAM. The two first servers are virtualized, and the rest of them are not virtualized, because they have low performance features.

A. General Description of the Test Environment

The proposed solution is integrated in a simulated ecosystem. In this way, several systems are simulated:

- EMS. This system simulates the generation of information related with the demand response and the energy flow management. The internal database follows the IEC standards implemented in a relational database for the Energy Management Systems. The demand response information is randomly generated limited to different intervals. Moreover, when the system detects that the random information is near to the end of proposed intervals generates alarms or events. This system has available several WS related to reporting activities, and several request/response WS that implements commands about consumption and generation.
- Commercial System. This system randomly generates information about consumption. This generation is based on the real consumption curves, from the residential and industrial consumers. The system generates random information for 3400 residential and 50 industrial customers, although these parameters are configurable. The information is stored in a relational database with a non-standard structured, based on real database structure. This system integrates several WS of subscription/notification type, in order to notify the billing process, and request/response services to provide different information for reporting activities.
- Electric Vehicle Charging Infrastructure Management (EVCIM). This system implements several procedures to generate random information about consumption in different points of power grid. The original system is implemented with 6 charging stations, with a maximum power of 50 kW. The information is stored in a relational database without any information standard structure. There is not implemented any simulation system for routes and fleet management, only simulates the impact of information generated to manage the electric vehicles charging.
- DMS. This system stores information about the power grid infrastructure. The system simulates a segment of generic distribution grid. This segment is based on the IEEE 34 Node Test Feeder ([21]). The system stored all information based on International Electrotechnical Committee (IEC) Common Information Model (CIM), extended to allow the acquisition of information from the specific protection modules, which are randomly generated. The main services available in this system are requests/responses retrieving the information from the server and generating alarms and events.
- Photovoltaic Generation Management System (PVGMS). This system implements a simulated photovoltaic generation station, with 1.5 MWh. The data about generation is randomly generated, based on the model generated from real information of photovoltaic farm. This system has several responses/requests WS oriented to retrieve information from the server. Additionally, this system can receive different commands to manage the load on power grid.

The interconnection of these systems are implemented based on an ESB. Two ESB have been tested: Mule and TIBCO. The proposed architecture is shown in Figure 3.

B. Simulation Engine

The Simulation Engine is the external system, which orchestrates the simulation. The Simulation Engine has programmed several simulation sequences related to:

- Normal operation. The request of different WS is related to the normal operation of power grid. In this sense, the DMS, EMS and Commercial System are the systems with a high number of request/response WS, although there are mainly a lot of subscription/notification WS. PVGMS and EVCIM receive commands and generate notifications about the generated information in each system.
- Fault in a specific feeder node. This case is characterized by a very high level of request WS without response, and additionally, the generation of several alarms and events from the simulated external systems.
- Billing period. In this case, the Commercial System increases its activity, with a lot of notification WSs.

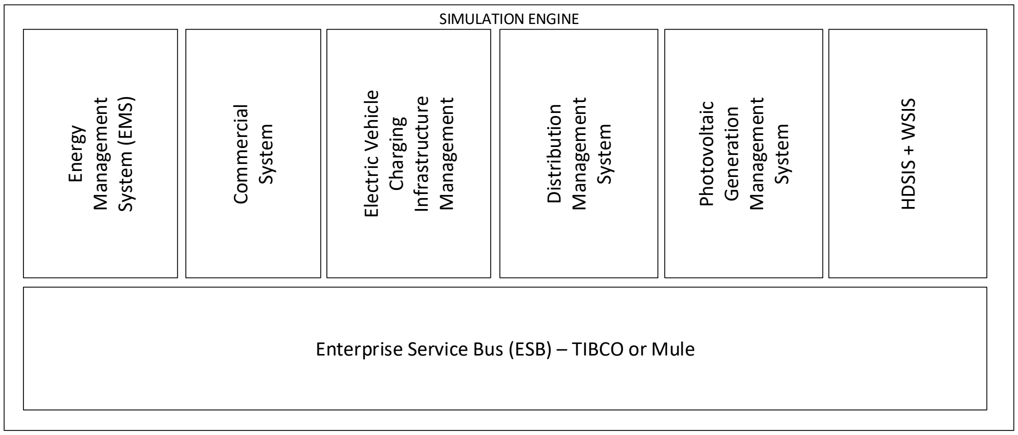

Figure 3. Architecture of simulation infrastructure.

- Massive electric vehicle charging period. In this case, the Commercial System, the PVGMS and the EVCIM generate a lot of notification WSs, and EMS and DMS generate a lot of request WSs and commands.
- Mixed mode. In this case, a combination of two of the previously described cases are combined, generating alternative scenarios, which could provide variations in the detected WS sequences because of the interference of both cases.

The Simulation Engine randomly runs each of these behavior patterns, until 100 patterns are performed. This parameter is configurable. The system converges to the best solution in each iteration, making variations of WS definition.

Although, these sequences are programming according to standards in the Simulation Engine, the proposed solution has not any information about it. The proposed solution only gathers information from ESB.

C. Simulation Results

In the first stage of the system, the HDSIS makes the integration of all available metadata in different systems, providing 47 new WS, which provided access to the old and new information, generated by different models created by DTME. The data access is based on WS call that deals with metadata to address the data target. In this case, due to the lack of disk space, the integration was performed at metadata level. This is a configurable option in HDSIS.

The Simulation Results show that after performing of 10 patterns, the proposed solution has reduced the number of calls in 20.62%, and the response time in 17.6%. The proposed solution increased the efficiency in 32.09%. After the running the 10^{th} pattern, the system can increase the efficiency with a low rate of variation between 0.2% and 8.92% per each 10 performed patterns. After the 10^{th} pattern, the WSHR has increased the available WS in 6 new WS, aggregated from existing WSs.

After several iterations of simulation process, the increase of efficiency rate depends on the regularity of WS usage and the semantic coherence of response. It is very difficult to simulate all contingencies and scenarios of this type of ecosystems. However, the implementation of simulation infrastructure and the simulation processes performed for the project have provided several requirements and modules to establish a basic Simulation Engine for this type of system.

VII. Conclusion

The integration of different systems in a SG ecosystem is one of the most important topics in Smart City, one of the most important way to make a good integration is the standardization of different process related to SG. However, this process is still running and the current distribution system cannot afford them to discard the old system (with all information stored) and start with a new system and with new specifications. The distribution companies have to define hard deployment plans that allow the implementation and integration of the modern architectures with the traditional ones.

The proposed solution is a novel approach to achieve this goal, making easier the deployment processes of old and new systems in a SG ecosystem. The new systems are integrated, providing new information and services, which could be available for other components or systems in the SG. This integration is performed at data level with an HDSIS and WS level with a WSIS, increasing the global efficiency than 30%.

Additionally, a Simulation Engine has been designed and tested, providing a test environment for this type of deployments, which, in the real case, requires a lot of hardware and software resources.

Acknowledgment

The authors are also appreciative of the backing of the SIIAM project (Reference Number: TEC2013-40767-R), which is funded by the Ministry of Economy and Competitiveness of Spain.

References

[1] X. Fengguang, H. Xie, and K. Liqun, “Research and implementation of heterogeneous data integration based on XML,” 9th International Conference on Electronic Measurement Instruments, 2009. ICEMI ’09, Beijing, 2009, pp. 4-711-4-715.

[2] L. Tianyuan, S. Meina, and Z. Xiaoqi, “Research of massive heterogeneous data integration based on Lucene and XQuery,” in 2010 IEEE 2nd Symposium on Web Society (SWS), Beijing, 2010, pp. 648-652.

[3] X. d Chen and J. z Liu, “Research on Heterogeneous Data Integration in the Livestock Products Traceability System,” in International Conference on New Trends in Information and Service Science, 2009. NISS ’09, Beijing, 2009, pp. 969-972.

[4] X. b Han, F. Tian, and F. b Wu, “Research on Heterogeneous Data Integration in the Safety Production and Management of Coal-Mining,” in 2009 First International Workshop on Database Technology and Applications, Wuhan, Hubei, 2009, pp. 87-90.

[5] W. Hailing and H. Yujie, “Research on heterogeneous data integration of management information system,” in 2012 International Conference on Computational Problem-Solving (ICCP), Leshan, 2012, pp. 477-480.

[6] Y. Shi, X. Liu, Y. Xu, and Z. Ji, “Semantic-based data integration model applied to heterogeneous medical information system,” in 2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE), vol. 2, Singapore, 2010, pp. 624-628.

[7] H. Fan and H. Gui, “Study on Heterogeneous Data Integration Issues in Web Environments,” in International Conference on Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007, Shanghai, 2007, pp. 3755-3758.

[8] Y. Cao, Y. Chen, and B. Jiang, “A Study on Self-adaptive Heterogeneous Data Integration Systems,” in Research and Practical Issues of Enterprise Information Systems II, L. D. Xu, A. M. Tjoa, and S. S. Chaudhry, Eds. Springer US, 2007, pp. 65–74.

[9] T. H. Merrett, “Attribute Metadata for Relational OLAP and Data Mining,” in Database Programming Languages, G. Ghelli and G. Grahne, Eds. Springer Berlin Heidelberg, 2001, pp. 97–118.

[10] G. Zheng and A. Bouguettaya, Web Service Mining: Application to Discoveries of Biological Pathways. Boston, MA: Springer Science+Business Media, LLC, 2010. p. 152.

[11] W. v d Aalst, “Service Mining: Using Process Mining to Discover, Check, and Improve Service Behavior,” IEEE Trans. Serv. Comput., vol. 6, no. 4, pp. 525–535, pp. 525–535, Oct. 2013.

[12] Q. A. Liang, J. y Chung, S. Miller, and Y. Ouyang, “Service Pattern Discovery of Web Service Mining in Web Service Registry-Repository,” in 2006 IEEE International Conference on e-Business Engineering (ICEBE’06), Shanghai, 2006, pp. 286-293.

[13] A. Zhao, X. Wang, K. Ren, and Y. Qiu, “Semantic Message Link Based Service Set Mining for Service Composition,” in Fifth International Conference on Semantics, Knowledge and Grid, 2009. SKG 2009, Zhuhai, 2009, pp. 338-341.

[14] Q. A. Liang, S. Miller, and J. Y. Chung, “Service mining for Web service composition,” in IRI -2005 IEEE International Conference on Information Reuse and Integration, Conf. 2005, 2005, pp. 470-475.

[15] H. Luo, L. Liu, and Y. Sun, “Semantics-Based Service Mining Method in Wireless Sensor Networks,” in 2011 Seventh International Conference on Mobile Ad-hoc and Sensor Networks (MSN), Beijing, 2011, pp. 115-121.

[16] W. Hu and Y. Hui, “Service-mining Based on Customer Value Analysis,” in 2007 International Conference on Management Science and Engineering, Harbin, 2007, pp. 109-114.

[17] H. Meng, L. Wu, T. Zhang, G. Chen, and D. Li, “Mining Frequent Composite Service Patterns,” in 2008 Seventh International Conference on Grid and Cooperative Computing, Shenzhen, 2008, pp. 713-718.

[18] L. Chen, Q. Yu, P. S. Yu, and J. Wu, “WS-HFS: A Heterogeneous Feature Selection Framework for Web Services Mining,” in 2015 IEEE International Conference on Web Services (ICWS), New York, NY, 2015, pp. 193-200.

[19] J. I. Guerrero, A. García, E. Personal, J. Luque, and C. León, “Heterogeneous data source integration for smart grid ecosystems based on metadata mining,” Expert Syst. Appl., vol. 79, pp. 254–268, pp. 254–268, Aug. 2017.

[20] J. I. Guerrero, E. Personal, A. Parejo, A. García, and C. León, “Forecasting the Needs of Users and Systems - A New Approach to Web Service Mining,” in The Fifth International Conference on Intelligent Systems and Applications, Barcelona, Spain, 2016, pp. 95-99.

[21] W. H. Kersting, “Radial distribution test feeders,” in 2001 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.01CH37194), vol. 6, no. 3, 1991, pp. 975-985.