International Conference on Renewable Energies and Power Quality (ICREPQ'17)
Malaga (Spain), 4th to 6th April, 2017
Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, No.15 April 2017

A demonstrator tool to provide the network operator with microservices based on big data and semantic web technologies

E. Bionda[1], F. Belloni[1], R. Chiumeo[1], D. Della Giustina[2], D. Pala[1], G. Proserpio[1], S. Pugliese[2], H. Shadmehr[1] and L. Tenti[1]

[1] Ricerca sul Sistema Energetico – RSE S.p.A.
Via Rubattino 54, 20134, Milano, Italy

Phone/Fax number: +00390239921, e-mail: enea.bionda@rse-web.it, federico.belloni@rse-web.it, daniele.pala@rse-web.it, gianluigi.proserpio@rse-web.it, liliana.tenti@rse-web.it

[2] Unareti S.p.A. Reti Elettriche
Via Ponte Nuovo 100, 20128, Milano (Italy)
Phone/Fax number: +00390239921, e-mail: davide.dellagiustina@unareti.eu, salvatore.pugliese@unareti.it

Abstract. The paper presents a demonstrator tool, designed as an architecture of microservices, to visualize and correlate Power Quality (PQ) data with topological, cartographic and meteorological information. Starting from a single web access point, the tool allows a network operator to "cast a glance" to different kind of data, involved in the management of its network or coming from open data sets, to investigate possible correlations among them.
The adopted approach in data management make use of both the ontological and big data paradigma. This choice allows the operator to manage and correlate large volume of data of different kind (structured and unstructured data).
The adopted ontology is the IEC Common Information Model (CIM), which allows the integration of PQ information recorded by the QuEEN monitoring system, at the HV/MV stations of the DSO Unareti S.p.A., with the associated distribution network topology. Services offered regard: (i) the visualization of PQ indices in association with topological and cartographic information; (ii) a classifier, based on a Software Vector Machine (SVM) algorithm, to define the origin of voltage dips; (iii) a web application, developed in the Spark framework, to correlate in space and time QuEEN voltage dips times series with storm cells.

Key words
Power Quality, Common Information Model, Big Data, Semantic Web, Microservices.

1. Introduction

The electric network must be increasingly able to integrate, in a smart way, the actions of different users, generators, consumers and prosumers, to assure an electrical supply sustainable, economical and sure. As a consequence, nowadays, the electrical networks produce a huge volume of data of different origin (data from sensors, relational DB, the internet of things or IoT etc.). Data volumes and variety, together with data processing velocity, are problems to face, possibly with a flexible, expansible and scalable solution in order to help network operators to deal with a great amount of data of different origin, for different purposes, through a single access point.

The paper deals with a demonstrator tool which offers a set of microservices, ranging from PQ and cartographic services, network topology navigation service up to big data facility. Data integration is indeed achieved by a properly mixed approach to network data management based on both ontology and big data paradigms.
The tool is the natural extension of a former web application for the integration of PQ information with the associated system topology both expressed by the CIM standard [1].
PQ data are those recorded at the HV/MV stations of the Distribution operator Unareti S.p.A. by the QuEEN system [2]. Topological and asset information have been provided by Unareti, while meteorological data are provided by the STAF system (Storm, Track, Alert and Forecast system) developed by RSE [3]. Other useful information come from dataset freely available on the web, according to the paradigm provided by the Linked Open Data.
After a synthetic description of the adopted data management approach, the tool design architecture is presented to introduce the different services, which are described in the following in more details. Because of the design solution adopted, characterized by modularity, a simple way to build business logic and scalability, further tool developments will be facilitated.

2. A new approach to network data management

In order to extract useful *information* from such an amount of data it is necessary to identify possible relations among them. Besides, to go from information to "*knowledge*", a model is needed, able to connect all the available data and make forecasts possible. The whole "climbing" process from data to knowledge, described in Fig. 1, could be named the climbing of the "knowledge pyramid".
There are two complementary approaches at disposal "to make the climb" (the pink arrows in the figure).

Knowledge can be achieved by the use of big data techniques together with Machine Learning (ML) algorithms (the path on the right).

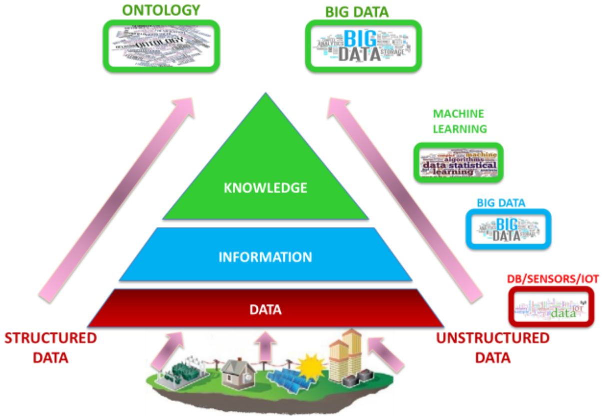

Fig. 1. The "Knowledge pyramid"

This approach is usually associated with large quantities of unstructured information. In this case the relations among data and the final model are not a "a priori" but are somehow "suggested" by the data itself (*data driven approach*). The alternative or *model driven approach* (the path on the left) consists in referring to a common information model that, for the electrical system, is the IEC Common Information Model (CIM).

CIM is a semantic model that describes the physical and logical components of a power system and the relationships between them.

In this work a mixed approach has been used to exploit the potentiality of both the paths, as the offered microservices make use of both structured and unstructured data.

3. The architecture

The tool is characterized by a microservices architecture design, as this solution allows the adoption of the mixed approach above mentioned [4]. Besides, the architecture is based on both message exchanges between the different microservices and a Web Application. This solution allows to integrate data usually managed by different applications. The following figure (Fig. 2) shows the architecture's solution, consisting of four main "logical" parts, described in more details below: i) the ontology (red), ii) the big data (light green), iii) presentation layer (yellow), iv) the service bus and microservices (green), to connect the different services.

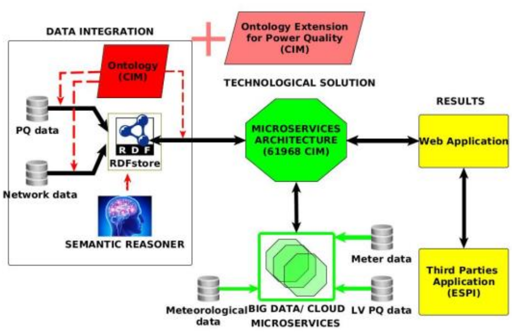

Fig. 2. The logical architecture

A. Ontology

DSO has to manage various kind of information (network topology, assets, geographical information, etc.). To manage these the DSOs' usually use different databases without connections between them. This creates a problem: every information contained in a single database is completely isolated and it is impossible to use it in correlation with data from other databases. This is one of the reason why IEC created a set of standards for electric power transmission and distribution: the Common Information Model (CIM)[5]. Three standards are part of CIM: 61970, 61968 and 62325. The IEC standard 61970 is related to the Energy Management Systems, the IEC 61968 extends this model to Distribution scope supporting data exchanges between components, while the IEC 62325 covers the electricity market and their data exchanges. The CIM standards are based on a semantic model, expressed by the UML language, that describes the physical and logical components of a power system and the relationships between them. CIM defines an XML format for network model exchanges using Resource Description Framework (RDF) [6]. This enables CIM to become a semantic model with the connected advantages. In this context, the full network of Milan and a few Brescia's bus-bars, of the Unareti network, have been transformed from some different "no standard" databases, into a single RDF Store in CIM, by a technique called Model Driven Engineering [7].

B. Service bus and microservices

The architecture is composed of different services (Fig. 3), that exchange messages using the IEC 61968 standard over the Advanced Messaging Queuing Protocol (AMQP) [8]. In this project the protocol AMQP is implemented by the RabbitMQ [9].

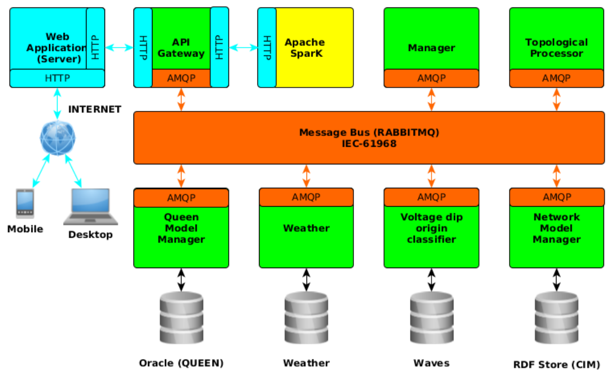

Fig. 3. The Service Bus

In a distributed architecture is fundamental to know what services there are and how can be contacted. To resolve this, the choice was to use a "publish-subscribe" pattern called "topic exchange" [10], where services subscribe a topic or more topics and receive messages about this topic. In this way it is possible to know in advance where you have to send a message about a particular topic, by the relative topic name knowledge.

The topic exchange is one of the pattern to achieve *scalability*, since it decouples message producers from message consumers. As to *reliability* the service bus can be considered a single point of failure but RabbitMQ, configured as a cluster, support highly available queues: every messages and every queues are replicated in every node of the cluster and this permit RabbitMQ to became reliable. As to the services shown in Fig. 3 it is possible to see that there are various services and each of them has its own database. One of the advantages of this "distributed solution", instead of a monolithic one, is the simplicity to develop a service, as well as to use the programming language and the database ideal for that service (the strategy is to use the better database for a particular service). Our examples of services are:

- the *Manager*: it manages the services connected to the infrastructure, some use cases and eventually re-starts a died service;
- the *Topological Processor*: has the function to perform calculations relating to a topological network provided as input in CIM format;
- the *Network Model Manager*: it is responsible of the RDF data store; it stores the CIM information of the networks, and it is responsible of answering at the semantic questions created using SPARQL Protocol and RDF Query Language queries [11];
- the *Queen Model Manager*: it provides PQ data, transforming them from a relational database to the CIM standard by Model Driven Engineering;
- the *Weather*: it provides weather information, mostly historical;
- the *Voltage dip origin classifier*: the main function is to classify a voltage dip on the base of its origin in the HV or MV network. The classification of the voltage dip is semantically linked with the relative busbar;
- the *API Gateway*: it is the service that translate in/to different protocols; in particular it translates HTTP requests in AMQP messages and vice versa. This allows different components, like CIM services, big data services, front-end services, to speak with each other.

C. Big data

This solution manages big data thinking at two different users: the "common user" and the "advanced user" (data scientist). For each of them, there is also another perspective to consider: automatic applications (machine to machine) instead of user application. For these reasons there are two different solutions to manage big data. Each of them is based on the framework Apache Spark [12], an open source framework to manage big data. The first solution consists in the unique use of the software Apache Spark and allows to run automatic applications and supports HTTP requests flowing throw the API Gateway, enabling the common user tasks. The second is a framework on the cloud supplied by Databricks [13], based on the Apache Spark, but it extends this service offering the possibility to increase the processing power, in a simple way, using a cluster, and in addition to providing a graphical interface more suitable for an advanced user like a data scientist.

D. Presentation

The architecture has been studied for separating the presentation layer from the data layer using a Web Application to show and control data, and the REpresentational State Transfer (REST) paradigm to access the data exposed by the microservices through the API Gateway. This separation supports the advantages of creating different applications, such as mobile applications, machine to machine applications and so on, without rebuild the backend.

4. The services

The operator can access the different services offered by the tool by the single access web point shown in Fig. 4.

Fig. 4. Home page

The services offered are briefly presented hereafter with some results.

A. PQ services(QuEEN services)

PQ services implemented offer respectively:

- a synthetic visualization of PQ data in association with other geo-referenced information of interest;
- detailed advanced analysis on PQ events based on Machine Learning supervised techniques.

In both cases the operator has, first of all, to select a specific monitoring location (from a geographical map or a list) and a monitoring period of interest (a week, a month). Hereafter some results obtained.

Visualization of PQ indices in association with topological and cartographic information

This service evaluates key performance indices of monitored PQ continuous phenomena (voltage variations, flicker, THD and unbalance), providing the DSO with a synthetic overview of the PQ performance of its networks together with other useful topological and cartographic information (total lenght and type of lines feeded by the monitored busbar, number of feeders, substations electric diagram, proximity of disturbing plants etc.)[1].

Referring for example to the measurement location A in Fig. 5, it shows a close proximity to steel plants directly connected to the surrounded HV network; its weekly PQ performances are summarized in Fig. 6.

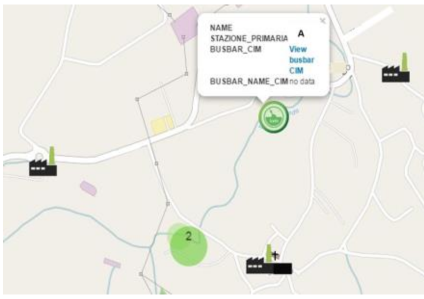

Fig. 5. Measurement locations selection

Fig. 6. Weekly PQ key performance indices

The indices colour used in the table allows the network operator to identify at a glance if there is a PQ problem in the location selected: a green/red indice refers to compliance/violation with EN 50160 standard.

In particular during the working week selected in Fig. 6 (2015 August 1st week) the supply voltage appears affected by a level of flicker higher than that allowed by the standard (EN 50160 requirement: $P_{lt} \leq 1$).

The *heat map*, on the top right of Fig. 7, summarizes the weekly indices performances of both the monitorig site A under analysis (black label) and the other monitoring locations (blue labels) by the following colour code:

- a red square points out a standard violation (*alarm state)*;
- a green square shows compliance with the standard (*normal state*); in particular a light green square means a compliant value that has nevertheless overcome the 80% of the standard threshold (*alert state)*;
- a grey square indicates a measurement unit under maintenance (*unavailable state*).

The heat map analysis shows there are two monitoring location (A and B) which seem affected by flicker problems due to the HV origin of this disturb, network level to which steel plants are connected. Cartographic data confirm that the two location are those nearest to the steel plants (full green balls in Fig. 7) of the Garda lake region.

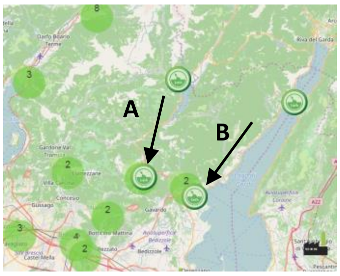

Fig. 7. Flicker affected locations (A , B) are the nearest to steel plants (full green balls)

To confirm this, the heat map referred to August 2nd week (including August holiday) shows compliant level of flicker for A and B too.

As to the other PQ parameters the service can offer other useful information: for istance a continous increasing succession of weekly "unbalance" violations ("unbalance drift") could mean that the measuring units under analysis need maintenance while periodic over/under coming of voltage admissibility threshold could give some information on the connected load.

Voltage dip origin classifier

The service allows the DSO to automatically classify voltage dips, monitored on its network, on the base of their source location (HV and MV). The classifier relies on a Support Vector Machine algorithm (SVM) which works on a set of event features, some of which extracted from the event full waveform by a Kalman filter [14].

The extracted features made available by the service include: i) voltage dip duration and depth, directly registrated by the QuEEN system, ii) voltage dip shape and number of "affected" line to line voltages, extracted by a waveform segmentation and analysis tool named FEXWAVE (Features Extraction from WAVEforms).

The information are then sent to the SVM-classifier to identify HV and MV events. The below figures present how usually would be the aforementioned features associate respectively with a MV event (Fig. 8) and a HV event (Fig. 9).

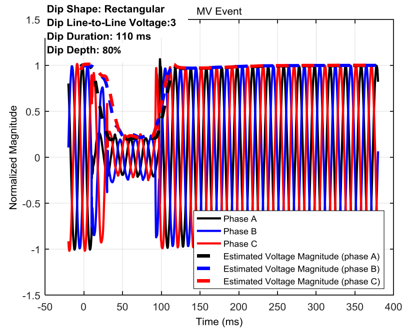

Fig. 8. Typical MV Event

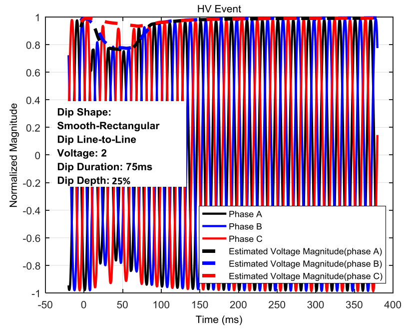

Fig. 9. Typical HV Event

The full process (FEXWAVES tool + SVM classifier) can be activated by the PQ service after the selection of a monitoring location and month of interest. The voltage dip origin is one of the parameter evaluated by the Italian national HV/MV substations monitoring system, recently implemented by the DSOs, on the request of the Italian Regulator [15].

The classifier results are presented both in a synthetic (Fig. 10) or a more detailed way (Fig. 11). The former figure shows HV/MV events monthly statistic while the latter refers to the distribution of the HV/MV events (coloured sphere) along and at the top/under a time line.

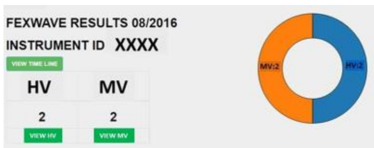

Fig. 10. Monthly statistic of voltage dip monitored at location XXXX: 2 MV origin and 2 HV origin

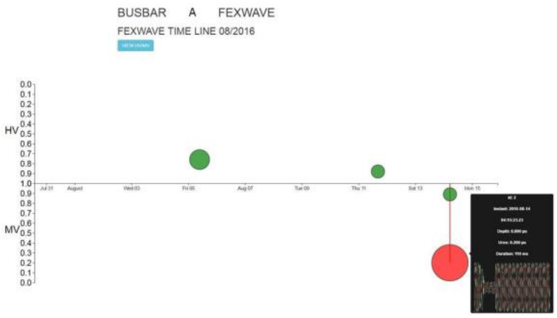

Fig. 11. Classifier results: HV/MV events along the time line with an associated waveform pre-view

The graph makes available “at a glance” some useful information as there is a proportionality between:

- sphere diameter and voltage dip duration;
- sphere vertical position and the residual voltage of the event (in pu).

Besides the sphere green/red colour refers to the event occurrence in a immunity/not immunity zone for class equipments connected to the network, according to IEC 61000-4-34.

The black tooltip at the right of Fig. 11 offers a sort of preview of further information directly achievable by the “sphere selection”, as the 3 line-to-line voltage waveforms of Fig. 12 (in case separately displayable).

A connection to the “Weather service”(§3.B) is available starting from the tooltip to verify the occurence of storms at the same time of events of HV origin.

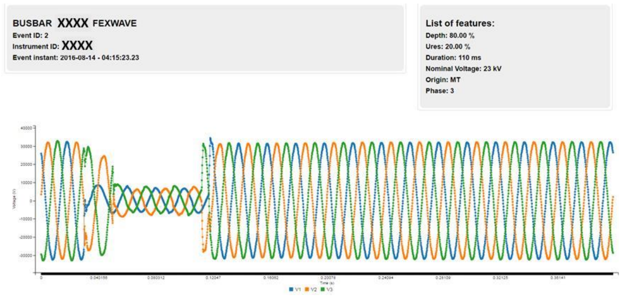

Fig. 12. Event features (top right) and waveform

B. CIM services

The CIM standard lets to manage different aspects of an electric grid that usually a DSO has to manage with different applications. So CIM integrates different information and in our solution, for example, there are integrated data from: (i) network topology; (ii) power quality measurements; (iii) geographical information (GIS); (iv) information about assets such as cables and graphical representations of electric diagrams. With the integration of so many different data it is possible to provide different CIM services, with the advantage to see different point of views in a single web page. In Fig. 13 it is possible to see a geographical visualization of a single feeder (red line): different information about assets can be extracted by simply clicking on it.

In Fig. 14 it is possible to see the geographical information about a feeder together with its CIM links with information about the line and the relative asset.

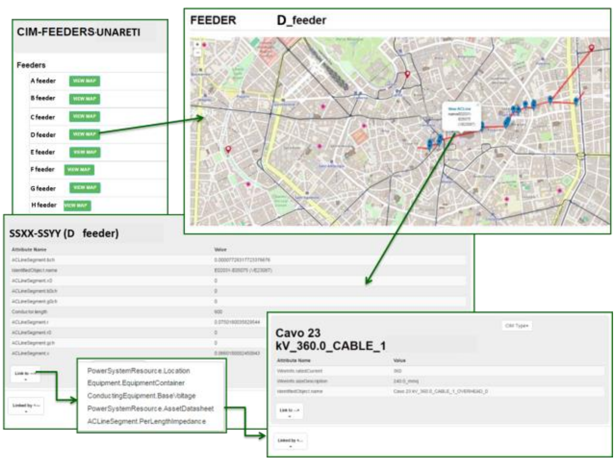

Fig. 13.Geographical representation of a feeder

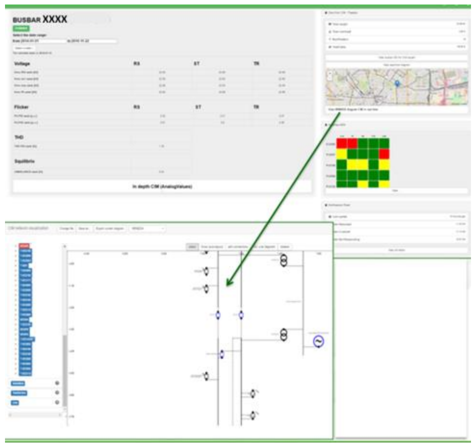

Fig. 14.Information from CIM elements belonging to a feeder

In one single web page the DSO can see the electric diagram of the station, browse the elements directly from it and see the values of the power quality parameters.

C. Big data services

The implemented services allow the identification of possible correlations among “historical” series of Power Quality and meteorological data, provided by different system, as:

- voltage dips monitored by the QuEEN monitoring system;
- radar reflectivity data provided by the Swiss Monte Lema Radar and processed by the STAF system (Storm, Track, Alert and Forecast).

The access to the services is different for “*common*” and “*advanced*” users, as previously described (§3 C).

Common users have the possibility to choose:

- the series of data to correlate;
- the input parameters to be used for the correlation (time interval and distance from the busbar monitored **within which** to find correlations between PQ events and storm cells).

The results appear on a web page as the percentage of voltage dips correlated to storm cells within the chosen time interval and distance from the MV bus-bar monitored. Besides, the user can download the detailed list of correlations.

An *Advanced user* (i.e. data scientist), is provided with a proper framework (Databricks) to write code or proper SQL queries to get more information from the “dataframe”. Some results of the advanced service are shown in Fig. 15÷Fig. 17.

Referring to the number of voltage dips monitored at fourteen Unareti MV busbars (Fig. 15), the measurement unit MU11 shows the highest number of events (63 voltage dips). This number corresponds to about the 38% of the events monitored by all the measurements units in Brescia, in the 6 months monitoring period considered (Fig. 16).

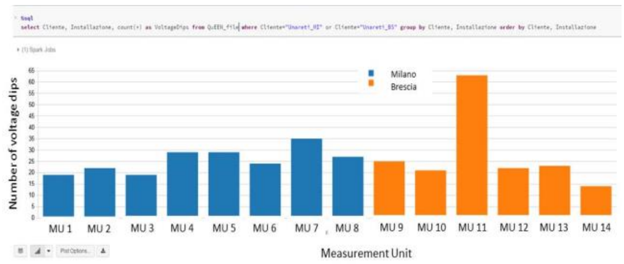

Fig. 15. Voltage dips monitored in Milano (blue bars) and Brescia (orange bars) measuring locations

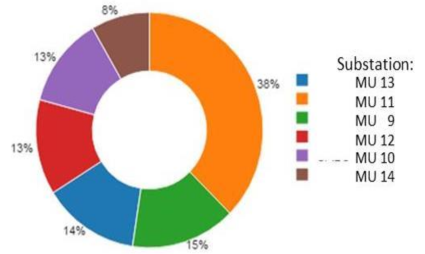

Fig. 16. Voltage dips statistics in Brescia

However most of the events monitored by MU 11 turn out to be correlated to storm cells (38 correlations in Fig. 17), being probably due to metereological cause.

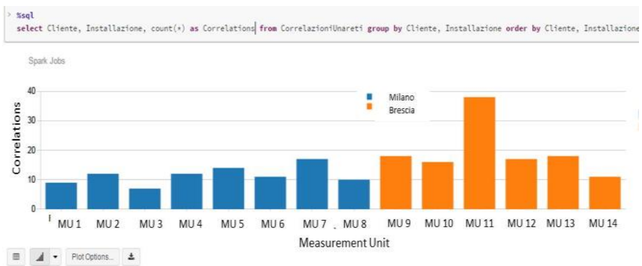

Fig. 17. Correlation between voltage dips and storm cells

5. Conclusion

The demonstrator tool allows the network operator to have access to different services fed by different databases or open data datasets, at a single web access point. The major advantages of the proposed solution relies on its:

- *integration of different paradigms* to create knowledge: semantic and big data approaches;
- *modularity*: you can add services when you need them;
- *a simple way to build business logic*: you join different services like LEGO bricks;
- *scalability*: you can increase the number of the same service (suitable solution for cloud computing);
- *separation between data layer and presentation*: possibility to build different applications such as mobile, machine to machine, without changing the architecture using REST API;
- open source and standards reduce the costs and simplify the interoperability between different services;
- possibility of answering to "semantic questions": the tool can access different databases by proper semantic queries to get integrated information.

From the final user perspective, the tool allows the operator to quickly check its network PQ performances, "at a glance", and make proper assumptions on "what is going on" by crossing PQ data with other information of interest.

Acknowledgement

This work has been financed by the Research Fund for the Italian Electrical System under the Contract Agreement between RSE S.p.A. and the Ministry of Economic Development - General Directorate for Nuclear Energy, Renewable Energy and Energy Efficiency in compliance with the Decree of March 8, 2006.

References

[1]. E. Bionda, D. Della Giustina, D. Pala, G. Proserpio, S. Pugliese and L. Tenti, "Integration of Power Quality information in the framework of a standard semantic model", 26th IEEE Inter. Conf. on Environment and Electrical Engineering, 7-8 June 2016, Florence, paper 449.
[2]. R. Chiumeo, A. Porrino, L. Garbero, L. Tenti, M. de Nigris, 2009, " The Italian Power Quality Monitoring System Of The MV Network: Results Of The Measurements Of Voltage Dips After 3 Years Campaign ", 20th International Conference on Electricity Distribution, CIRED, Prague, paper 0737.
[3]. P. Marcacci, G. Stella, "STAF V.2 - Storm Track Alert and Forecast, Ver. 2", Final Technical Report ERSE 09004789, 31st December 2009 (in Italian).
[4]. Microservices, http://www.martinfowler.com/articles/microservices.html.
[5]. Core IEC Standards, http://www.iec.ch/smartgrid/standards.
[6]. RDF, https://www.w3.org/RDF.
[7]. G. Proserpio, D. Pala, E. Bionda, 2015, Studio ed applicazione di tecnologie ICT per la realizzazione dell'interoperabilità tra i sistemi componenti la smart grid, RSE, RdS, Report 15000384 (in Italian).
[8]. AMQP, https://www.amqp.org.
[9]. RabbitMQ, https://www.rabbitmq.com.
[10]. RabbitMQ-Topics, https://www.rabbitmq.com/tutorials/tutorial-five-python.html.
[11]. SPARQL, http://www.w3.org/TR/rdf-sparql-query.
[12]. Apache Spark, http://spark.apache.org.
[13]. Databricks-Apache Spark, https://databricks.com.
[14]. R. Chiumeo, H. Shadmehr and L. Tenti, "Self-tuning Kalman filter and machine learning algorithms for voltage dips upstream or downstream origin detection", ISSN 2172-038X Renewable Energy & Power Quality Journal, No.14, May 2016, Paper 346, pg. 409-413.
[15]. R. Chiumeo, L. Garbero, F. Malegori, L. Tenti, "Feasible methods to evaluate voltage dips origin", ISN 2172-038X Renewable Energy & Power Quality Journal, No.13, April 2015, Paper 325, 357-361.