- Normal: Obtaining registered data without filters or validations. The use of this option is not recommended, as it will consume the most resources on the technology platform. The rules described in sections 3.3 and 3.4 apply.
- Discrimination: Obtaining data, eliminating invalid values. The measurement values consistency is verified according to the rules described in section 3.2.
- Missing rows: Sometimes, some causes can induce, in a period, specific measurement equipment cannot be requested for data gathering and therefore the historical database lacks information. This function reports the missing samples in the period consulted for each selected equipment. It includes two options: Only missing rows and Raw data with missing rows.

4.2 Statistical data

The following values can be calculated for any time range, at any of the data groupings by frequency, for any recorded steady-state variable. The rules described in section 3.2 apply.

- Average: Arithmetic average of the requested values.
- Maximum: The highest value of the requested values. It is used to identify an extreme value that could be an outlier or a data entry error.
- Minimum: The lowest value of the requested values. It is used to identify an extreme value that could be an outlier or a data entry error.
- Sum: The result of summing all the requested values. It is used to integrate values in a region, for example, the real power of substation circuits or a substations group.
- Standard deviation: It is the square root of the variance of the requested values. This measure of dispersion is the most characteristics.

4.3 Data grouping by frequency

The following grouping strategies of the calculated values of section 4.2 allow optimizing the queries, generating and transporting only the data that is really useful to the end-user, depending on the function in which it will be used.

- Hourly: Returns a single data for each hour requested. It allows generating of daily profiles of EPS electrical behavior.
- Daily: Returns a single data for each day requested. It allows generating of weekly or monthly profiles of EPS electrical behavior.
- Weekly: Returns a single data for each week requested. It allows generating of monthly profiles of EPS electrical behavior.
- Monthly: Returns a single data for each month requested. It allows generating of annual profiles of EPS electrical behavior.
- Annual: Returns a single data for each year requested. It allows comparing the annual EPS electrical behavior and annual trends.
- Period: Returns a single data for the entire time range requested. This function is used to compare the EPS electrical behavior in specific periods of interest.

4.4 Power Quality events

The following values can be requested for any time range, but metering devices must have power quality functions.

- Interruptions: It is an instantaneous change of frequency from the steady-state of the current, the voltage, or both. It has a unidirectional polarity and is characterized mainly by its rise and fall times and its maximum value.
    - Momentary. Obtains the values with a voltage percent less than or equal to 10% and duration less than or equal to 3,000 [ms].
    - Temporary. Obtains the values with a voltage percent less than or equal to 10% and duration greater than or equal to 3,000 [ms] but less than or equal to 60,000 [ms].
    - Sustained. Obtains the values with a voltage percent of 0% and duration greater than or equal to 60,000 [ms].
    - All. Gets all interrupts when the current flow stops for any reason in the selected time range.
- SAGS: Decrease in the effective voltage value between 0.9 and 0.1 P.U. and duration from 16 [ms] up to a few seconds.
    - Instant. Values with a voltage percent greater than or equal to 10% but less than or equal to 90%, and duration greater than or equal to 16 [ms] and less than or equal to 500 [ms].
    - Momentary. Values with voltage percent greater than or equal to 10% but less than or equal to 90%, and duration greater than 500 [ms] and less than or equal to 3,000 [ms].
    - Temporary. Values with a voltage percent greater than or equal to 10% but less than or equal to 90%, and duration greater than 3,000 [ms] and less than or equal to 60,000 [ms].

  - All. Gets all SAGS records stored for the selected time range.
- SWELL: Increase in the effective voltage value between 1.1 and 1.8 P.U. and duration from 16 [ms] up to a few seconds.
  - Instant. Values with voltage percent greater than or equal to 110% but less than or equal to 180%, and duration greater than or equal to 16 [ms] and less than or equal to 500 [ms].
  - Momentary. Values with voltage percent greater than or equal to 110% but less than or equal to 140%, and duration greater than 500 [ms] and less than or equal to 3,000 [ms].
  - Temporary. Values with voltage percent greater than or equal to 110% but less than or equal to 120%, and duration greater than 3,000 [ms] and less than or equal to 60,000 [ms].
  - All. Gets all SWELLS records stored in the selected time range.

4.5 Calculated data

- SCADA Equivalent Value: It is used when the real-time data is not available, normally from a SCADA, or it is necessary to compare the actual SCADA value with an estimated value based on historical data. It is obtained with the following sequence:
  - Calculate the equivalent previous date, as the day of the previous week that is similar to the current day, or the day of the previous month that is similar to the current day.
  - Take the current time without minutes.
  - Request the average of historical values for the current time on the equivalent previous date.
- Last Stored Value: For any variable, it is requested which was the last value that was entered in the historical record and the corresponding timestamp. It allows to carry out validations of the historical record operational status to estimate the quality of the stored data.

5 Conclusion

The value that the **Optimal Extractor** has added to the processes of support the operational decisions in a Smart Grid context has been very relevant, and every time specialist users have found new ways to take advantage of this advantage of being able to visualize the EPS behavior over time.

For example, the **Optimal Extractor** allows you to graph or export to Excel the hourly average measurements by Phase for each Substation Circuits in one year; this query will generate approximately 8,760 values per electrical parameter (365 * 24) regardless of the equipment sampling frequency; a normal extraction of raw data with a sampling frequency of 10-minutes implies transfer approximately 52,560 values for each parameter (365 * 24 * 6) and double if the sampling frequency is 5-minutes (365 * 24 * 12), in addition to the time and hardware resources on the client-side, required to process all the data obtained.

The integrated technology, the optimal data extraction strategies, and the adopted standards have made it possible to perform high-level functions with extraordinary performance, reducing even 95% of the waiting time to user response. For instance:

- An EPS operator can get the graph of the hourly real and reactive power profile for the last 24 hours in approximately 5 seconds.
- The graph or table of the hourly maximum values for voltage or real power measurements of a circuit for a year is generated in approximately 50 seconds.
- If the above query is made per day, the response time is less than 20 seconds.

A very representative and valuable query of the **Optimal Extractor** as support to the operational decisions for an EPS operator, during a failure and reestablishment event, is to have the ability to calculate on the fly (OLAP), the maximum hourly demand profile for the circuits involved, the circuit faulted and those that can support the restoration; the integrated function allows having the graph on screen in less than 3 seconds for each circuit.

Another integrated function that provides greater value for the users in charge of EPS operational analysis, is the computing of the Coincident Peak Demand (CPD) for all the circuits in a geographical region in a year [4]. Manually, this analysis for a geographical region with at least 500 circuits, can take from 2 to 3 months; the **Optimal Extractor** calculates the value in approximately 30 seconds and 2 minutes if the data quality algorithms are applied and generating the calculation memory required to support the results and operational decisions.

Table 1 shows results obtained comparing the three architecures performance:

- ARQ1: Traditional components architecture.
- ARQ2: Proposed components architecture for syntactic interoperability.
- ARQ3: Proposed components architecture for semantic interoperability.

In all the Test Cases, the ARQ2 obtains the better time response for the final user, and the improvements compared to the ARQ1 are very notable. Regarding ARQ3, when the data amount is relatively low, the time response for the final user can be greater than ARQ1, this is due to the metadata required by implementing the CIM Instances, in general, this effect is not presented when the data amount is increased, and the final user perception is not affected due the total time added is less than 2 [s].

The architecture and strategies proposed for the **Optimal Extractor** allow progress in the implementation of functions for the Smart Grid in the context of the operation of the EPS.

Table 1. Comparison results using the three architectures described.

Test Case			ARQ1		ARQ2			AQR3		
Case	Grouped by frequency	Period	Float values* transferred	Total time for user [s]	Float values* transferred	Total time for user [s]	Time % (ARQ2/ARQ1)	Float values* transferred	Total time for user [s]	Time % (ARQ3/ARQ1)
Statistical (AVG)	Hourly	1 Hour	420	0.15	35	0.09	60.0%	35	0.28	186.7%
		1 Day	10,080	0.47	840	0.16	34.0%	840	1.52	323.4%
		1 Week	70,560	0.80	5,880	0.14	17.5%	5,880	2.42	302.5%
		1 Month	302,400	6.73	25,200	0.86	12.8%	25,200	5.57	82.8%
		1 Year	3,679,200	71.36	306,600	6.92	9.7%	306,600	15.89	22.3%
	Daily	1 Day	10,080	0.37	35	0.13	35.1%	35	0.32	86.5%
		1 Week	70,560	0.70	245	0.29	41.4%	245	1.33	190.0%
		1 Month	302,400	5.93	1,050	1.36	22.9%	1,050	2.73	46.0%
		1 Year	3,679,200	70.86	12,775	2.62	3.7%	12,775	6.88	9.7%
	Monthly	1 Month	302,400	6.23	35	0.27	4.3%	35	0.43	6.9%
		1 Year	3,679,200	70.06	426	1.92	2.7%	426	3.06	4.4%
	Anualy	1 Year	3,679,200	69.26	35	1.74	2.5%	35	1.89	2.7%
CPD	Anualy	1 Year	18,396,000	^ 86729.55	4	3.90	0.004%	4	4.09	0.005%

* 32 Electrical measurements + Timestamp (date - time) + Circuit ID
^ Includes 86,400 [s] for CPD processing in the client side

References:

[1] A. Espinosa-Reza, A. Quintero-Reyes, R. Garcia-Mendoza, J.F. Borjas-Diaz, T.M. Calleros-Torres, B. Sierra-Rodriguez and R. Torres-Abrego, On-Line Simulator for Decision Support in Distribution Control Centers in a Smart Grid Context, *WSEAS Transactions on Systems and Control*, Vol. 5, No. 10, 2010, pp. 814-816.

[2] A. Espinosa-Reza, H.R. Aguilar-Valenzuela, M. Molina-Marin, M.L. Torres-Espindola, T.M. Calleros-Torres, E. Granados-Gomez, R. Garcia-Mendoza and C.F. Villatoro-Hernandez, Semantic Interoperability for Operational Planning for the Electric Power Distribution System, *WSEAS Transactions on Power Systems*, Vol. 11, 2016, pp. 289-298.

[3] M. Molina-Marin, E. Granados-Gomez, A. Espinosa-Reza and H. R. Aguilar-Valenzuela, CIM-Based System for Implementing a Dynamic Dashboard and Analysis Tool for Losses Reduction in the Distribution Power Systems in México, *WSEAS Transactions on Computers*, Vol. 15, 2016, pp. 24-33.

[4] A. Espinosa, S. Gonzalez and H.R. Aguilar, Results of applying a semantic interoperability strategy in Smart Grid applications for DSO in Mexico, *CIGRE Session 2016, General Meeting, Paris, France*, 2016.

[5] A. Espinosa-Reza, H. R. Aguilar-Valenzuela, M. Molina-Marin, M. L. Torres-Espindola, T. M. Calleros-Torres, E. Granados-Gomez, R. Garcia-Mendoza and C.F. Villatoro-Hernandez, Implementation of a CIM-Based Semantic Interoperability Strategy for Smart Grid in Mexico, *International Journal of Power Systems*, Vol 1, 2016, pp. 33-40.

[6] A. Espinosa-Reza, M.L. Torres-Espindola, M. Molina-Marin, E. Granados-Gomez and H.R. Aguilar-Valenzuela, Semantic Interoperability for Historical and Real Time Data Using CIM and OPC-UA for the Smart Grid in Mexico, *WSEAS Transactions on Cumputers*, Vol. 15, 2016, pp. 1-11.

[7] Jose Alfredo Sánchez-López, Eduardo Islas-Pérez, Alfredo Espinosa-Reza and Agustín Quintero-Reyes, Deploying SCADA Data to Web Services for Interoperability Purposes, *IEEE 2015 Global Information Infrastructure and Networking Symposium GIIS 2015*, Mexico, 2015.

[8] R. Rhodes, *Common Information Model Primer, Fourth Edition*, EPRI Technical Report, 2018.

Contribution of individual authors

Marxa Torres-Espindola, carried out the development of the components in C#.NET and implementation in many information systems.

Alfredo Espinosa-Reza, was responsible for the design of the architecture proposed, CIM Profile validation, and tester of the final products.