IMPLEMENTATION OF AN RPA TECHNOLOGY MAKING USE OF ARTIFICIAL INTELLIGENCE WITHIN THE STG WEB FLOW CHART

Master Thesis in Smart Grids in the University of Comillas and University of Strathclyde

Ángel Burgos Rastoll

Author of the project
Master student in the University of
Comillas and University of Strathclyde
Madrid (Spain)
Glasgow (Scotland)

***Abstract*—This report aims to provide a full picture of STG web in Iberdrola in order to provide enough information to the RPA provider so that a bot that will automatize the whole system can have enough input and output arguments to perform correctly.**

Keywords— RPA, automatization, STG web, Intelligent Automation, clustering.

I. INTRODUCTION

Iberdrola is the leader of the energetic sector in Spain and one of the top companies in the world in this market. It has been almost a decade since they implemented STG (Sistema de Telegestión) to control the incidences that take place through the grid. STG is the remote management system that is in charge of detecting manual and automatic incidences in the grid and manage their control and supervision.

First of all, incidents are generated in the system due to the malfunction of an element and then a diagnosis is assigned to it. With this diagnosis, a series of actuations and actions are promoted to solve the incident and, if everything has worked well, its status is changed to solved and it will be closed so that it does not appear in the system as something to work with. However, this process of actions can be improved by optimizing and automating the process in some points where human intervention does not contribute. This is a point of improvement and economic profitability for the company.

For these reasons, Iberdrola is trying to adopt a new methodology which enables them to increase their efficiency and to be more competitive in their market. This methodology must be achieved by implementing some type of technology that helps moving the company towards that direction. After a long process of research, some ideas were proposed and, at the end, one of them was chosen as the most suitable for the system. This new solution would be the one that should take over the process to achieve an optimization of the whole system. Therefore, once the decision was made, it was a matter of time that one of these technologies was selected and the process of renovation started out.

.

II. DESCRIPTION OF TECHNOLOGIES

As technology continue to grow, and the benefits they provide to companies become more tangible, automation technologies providers intend to offer a wider spectrum of products to their clients so each of them can adjust better to the certain needs of every customer. In this chapter, it will be analyzed the different solutions that the market is offering right now and the cases of use that are already implemented in Iberdrola.

A. RPA technologies

Nowadays, digital processes are flooded with tons of data that multiply themselves over the years. RPA (Robotic Process Automation) refers to a bot or software that executes itself virtually. It aims to automatize repetitive tasks that do not require too much complexity to solve, like collecting and transfer of data, file management, scanning or gathering information [1]. It is a great tool for big scale processes where there is a huge volume of easy and standard tasks that can be optimized with this technology.

Back-office departments are some of the most beneficiated by RPA since human resources, finances, accounting or client service among others need to deal with tons of information that can be standardized and managed as a robotic process taking away load of work from the worker. This could be seen as a threat by employees, but it is actually a tool that they can take advantage from so they can focus on other tasks that really need the supervision of a human. This way, a better use of time and optimization is achieved only by removing simple and repetitive tasks from the requirements of workers.

This technology is being already applied worldwide and is expanding its range of market. Its software is not invasive, performs fast and does not need much supervision once it its installed and working. It can be easily integrated in IT environments and complements the rest of processes involved in the system of the company enhancing a great collaboration between them. Its configuration is designed to be treated transversally so it does not stick to certain kind of departments, and it can be adapted to methods that were not originally supposed to be used.

The robot is designed to work in the user interface (UI) as if it was a human worker, so it would move the mouse or use the keyboard like any other user would [2]. Before it gets implemented in the system it must be trained so that it gets used to the configuration and environment where it will perform. This training is considered as an education of the robot that is fundamental in structured systems with specified rules that it must follow and must adapt its methodologies to them. It is important to state that these bots are not trained to learn from experience and to act wisely when they find an exception in the system, then it is the human expert the one who must solve the issue reducing his area of influence to really needed tasks.

B. *Intelligent Automation*

Intelligent Automation is a concept that covers a set of automation tools based on RPA technologies combined with Machine Learning (ML) and Artificial Intelligence (AI) [3]. It is defined as any kind of tool that can improve their own processes and results, optimizing IT resources and efficiency through data analytics.

Intelligent automation could be considered as a combination of different technologies that conform a higher-level concept with a bigger scope. In the next figure, it can be appreciated the structured data processes, based on rules, combined with non-structure data processes, based on judgements:

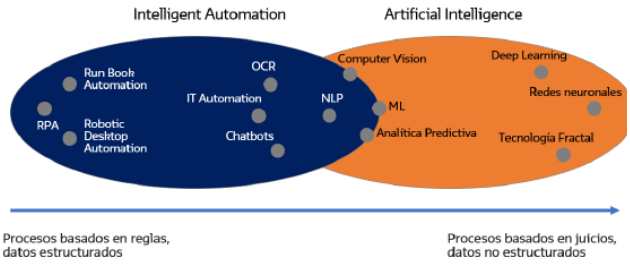

Figure 1. Intelligent automation as a combination of AI and RPA technologies

III. Definition of the Work

A. *Justification*

RPA technologies have become a great tool for the adaptation of established systems to the new ages. Nowadays, companies are trying to update their models of work so that they can be competitive against the rest. Iberdrola wants to follow this path and get a better efficiency in their departments. These technologies are already implemented in some departments but, in this report, we will focus on the adaptation of RPAs to STG (Sistema de Telegestión), and the preparation of intel of the system and different cases of use so once the bot is implemented it can be properly trained and work in the direction we need.

The later implementation of the bot requires a previous work that provides enough information to design the automatization of the system properly so an optimization in the whole grid is achieved. This automatization will be performed attending to specific criteria that will be proposed according to the needs and the potential improvement areas in the process. Once the system has been totally reviewed and conclusions have been made, it can be approached by the robot to improve the system and avoid waste of resources that could be used in other segments of the system.

B. *Objectives*

a) Division of data in specific groups to be studied separately: There are tons of data in the STG that make it really hard to understand what is going on in every moment and the reasons why error occur. One of our goals is to determine what is really interesting in the system and which data can give us useful information that can be utilized to design the robot. A division of data in groups makes it much easier to discriminate important and interesting information from the rest and to regroup those cases with characteristics in common to create a bigger picture of the system and allows the user to focus on the specific data that will help advance the project.

b) Automatization of repetitive tasks: There are multiple tasks in STG that are performed again and again and are not complex at all. These are the ones that RPAs can help manage, because by automating them the system gains efficiency and saves time to focus on more relevant issues. Therefore, the localization of these tasks and the calculation of their time, delays and costs is one of the main objectives of this report.

c) Optimization of the whole process: This is the main objective of this project; it is the reason why this study is being performed and why Iberdrola wants to adapt robotic automation in STG. The system of incidences is far from being perfect and a lot of processes and configuration can be modified so that work is truly effective and there is no waste of time. There is a need to identify those processes and the reason behind their low performance and, if they are correctly selected and analyzed, suitable modifications can be implemented that will make the system more efficient and will increase the degree of optimization of the whole grid.

d) Introducing improvements in the system to embrace new technologies in the future: The process of automatization does not only cover the short-term goals that we are trying to reach in this report, data will also continue to grow, there will be more items and issues involved in the system in the future, some of them we cannot even think now about. For this reason, it is mandatory to prepare the system for future innovations and technologies that will be able to get included more easily if we already take that chance into account.

C. *Methodology*

a) Collection of intel about the system: Before attempting to filter data and to analyze the steps in the process

to be optimized, we need a deep understanding of the system and how the grid and its infrastructure is organized. In order to do so, there must be an exhaustive research on the documentation about STG web and the electrical and telecommunications grid that conform it. The collection of documents and files that explain this system should provide tools and items to be used at the time of designing the robot and will help implementing it by a fluent integration. The following steps in the methodology are based on this research so sources must be official, and the information must be proved and useful.

b) Data filtering (Clustering): At the time of dealing with so much information, it is unavoidable the use of data analytics in order to filter and to gather different elements into groups that will be organized through some characteristics that will be defined according to our preferences. Clustering is a great tool that gathers data in groups *[4]* according to some predefined criteria that must be established according to the research of information that was performed in the previous step. When specific criteria are selected, the construction of the analytical model must be developed. This model might suffer from changes depending on the direction that may take the project in certain moments but, at the end, it will be flexible enough to cover all the possible cases that are interesting to analyze for the bot design. After the model is completed, all the results must be noted so the next step can be performed right.

c) Study of the most relevant cases: Once the information is divided into groups, there is to decide which of them will be more interesting for our interest. Results from the clustering methodology will offer such a variety of potential cases of study that a analysis of their interest for the project and their difficulty to be executed later becomes truly essential. Some cases could turn out to be not too useful at the end but even that confirmation might be necessary to complete the project. Before choosing the right criteria to pick some cases or others, a conversation with expert on the topic in Iberdrola is always advisable not to make mistakes in such an important stage of the process.

d) Analyze the possible changes to be introduced: With all the collected data, it is time to study how the most relevant cases can be modified so the process is optimized and an increase of the efficiency can be achieved. Analyzing the most relevant cases will answer the questions about their performances and their potentials, if they provide enough information, a scheduling of improvements can be established and modifications that adapt to these cases can be designed properly.

e) Combining AI methodology with RPA technologies: When applying clustering methodology, it is necessary to make use of data analytics techniques and other artificial intelligence methods. However, the intention of our project is to include all this previous work in the implementation of the bot from the RPA provider, that means that all the work must be done to be complementary with RPA technologies. It is important to combine languages and tools so every step of the process can be managed from both sides and there can be a fluent communication between them.

IV. DESCRIPTION OF THE MODEL

At the time of developing the model, there must be a clear path to follow, so the steps are clear and are easy to follow until the obtention of the final result. The first stage of the process is the analysis of the system, where all the intel from it is properly organized for its study. STG web is the object of study of this report but a look on the system where it is included is essential to understand its functioning. This system is called TitaniumSTG, that is in charge of connecting STGweb and other systems to the database so there is a fluent exchange of information between the general data of the company and the other systems.

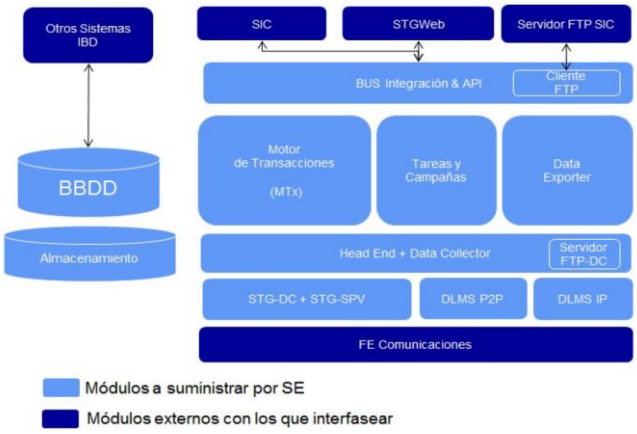

Figure 2. General vision of TitaniumSTG system

Inside TitaniumSTG, STG web focuses on the electrical grid, specifically the low-level elements that conform it and their communication to the system and between them. The process of generation, detection and actuations over the incidences in the system is the one that is aimed to optimize with the development of this project. Incidences can be divided into manual or automatic depending on how they get generated, if an agent observes a problem and wants it to be in the system a manual incidence will appear after its inclusion or if the device itself detects a malfunction in the system it is generated without the intervention of any worker. In this case, only automatic incidences will be object of study because they are the right ones to be automatized by the robot once its design is completed.

During the process of an incidence in the system since they are generated and they are concluded, there are several steps that it can follow depending on how their status changes and the actuation promoted on them.

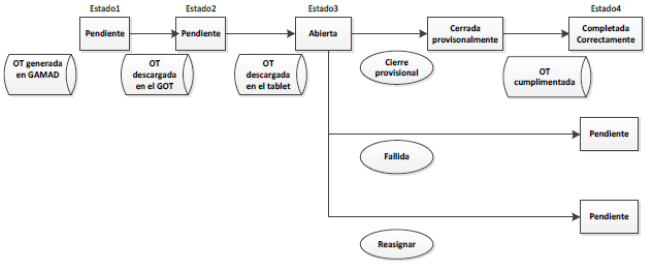

Figure 3. Flow chart of changes of status in STG web

The design of the model is performed attending to the collected information of the whole system and will be carried out making use of data analytics techniques, because of their capacity to work with huge loads of data. In this case, clustering is the selected methodology and it will allow a division in groups of the data according to their characteristics and the interest of the objectives of the project.

The different cases of study that will be performed are the most common diagnoses, aiming to tackle the worst and most repetitive type of incidences in the system; the incidences with a higher number of actuations, because they are those that need a longer path before they get completed and solved; the distribution of incidences per year, to see a full picture of the evolution of the number of incidences since the creation of STG web; the distribution of incidences per month, in order to appreciate possible differences in the share of incidences per time of the year; and distribution per zones, where there might be a pattern that shows that more incidences are generated in some territories and the distribution of them depending if it is summer or the rest of the year.

At last, the implementation of the clustering methodology will be carried out with Matlab 2020 and will include all the proposed cases of study.

V. Analysis of Results

The executed methodology for the implementation of the model provides really useful information that can be subjected to a deep analysis so that the future implementation of the bot in the system can be successfully done.

A. *Most common diagnoses*

At first, the most common diagnoses were detected and analyzed to find the main reasons behind these issues. After the collection of results, it can be stated that most common diagnoses are related to devices that do not answer back to communications. This might happen for a lot of reasons, but the most frequent one is the malfunction of the communication system. This system is continuously under review aiming to be perfectioned while more agents enter the system and new devices are integrated there. If we want to reduce these kinds of diagnoses, it could be done an increase on the margin that these devices have to create the incidence, the more flexibility you give the system, the less incidences are generated.

B. *Incidences with more actuations*

In this case, we were trying to see why some incidences were redirected too many times until they got solved. That is the reason why we searched for those that had a bigger number of actuations. According to the predefined criterion, we would need between 10 and 15 samples to make remarkable results so we searched for those that had more than 18 actuations.

When looking at the incidences that have more actuations, there is a clear pattern through them that is repeated in most of the cases. The main reason behind them is the lack of understanding between the different agents or the dispute for having the best solution in every case. However, there is another reason that covers many of them that is the mistakes in solving an issue in the way that the specialist of the device has said it should be. This can be due to the tons of incidences happening every day that can cause misunderstandings and confusion to the different agents.

C. *Distribution of incidences per years*

A full picture of STG web performing can be achieved by observing the number of incidences that have occurred since it was implemented in Iberdrola. For this reason, an algorithm to calculate them has been developed and the results are:

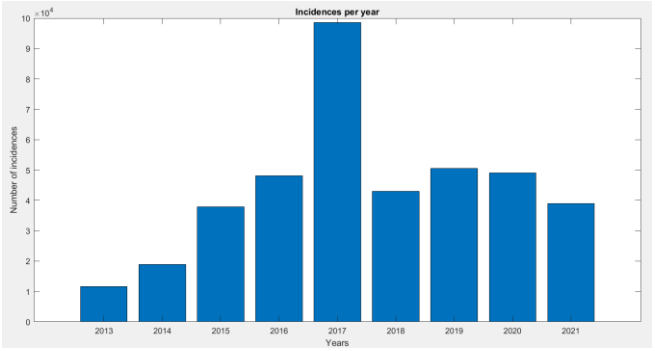

Figure 4. Number of incidences per year

As it can be appreciated, incidences show a curve that grows from 2013 to 2017 and then decreases until today. There are multiple reasons to analyze this effect, one of them is the inclusion on the grid of devices and their entry to STG web. In 2013, it was a new system with only not many smart devices so less incidences were created, but then lots of them were being included during the years and that obviously caused an increase in this aspect. However, the adoption of 3G technology instead of 2G has been also a progressive process that along with the inclusion of new technologies and tools has enhanced a decrease in the number of incidences as it can be appreciated since 2017. Since that year, a stability has been reached because of the balance of these new implemented technologies and the addition to the system of new devices that generate more incidences.

D. *Distribution of incidences per months*

Just like it was done with years, the same analysis is performed with months in order to see if there is some pattern in the incidences during a year. In the next figure, it will be shown the results of the number of incidences of every month since the implementation of STG web.

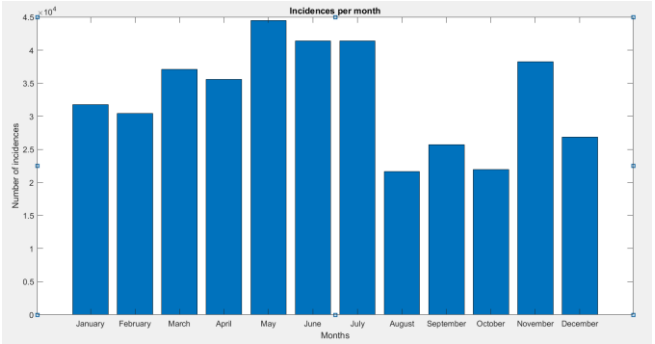

Figure 5. Number of incidences per month

At first sight, it can be observed that it is more or less linear during the year except for the months of August, September and October. However, after matching these results with the graphic of years, we found out that this is more similar to year

2017 and, as this is the year with more incidences, it increases the results on the rest of the months. Knowing that, there is still a little decrease on those months but apparently, after some research, there is not a clear reason behind it, so this case is not too relevant for our purpose of training the bot.

E. Distribution of incidences per zones

At last, we wanted to observe if there was a tendency in the territories to have more or less incidences and if there was a big difference between this proportion in summer and the rest of the year. According to these results, it could be stated that there is a balance between the number of incidences in all the territories except for the East zone, that has more than the double of them than the others. However, after a deeper research, it was found out that East zone was much bigger than the rest and, because of that, there was a higher number of installations in that territory. Therefore, the more installations there are, the more incidences there will be in that zone. To tackle this issue and have a more accurate vision of the system, it was calculated the number of installations in every territory and then, the number of incidences per installation.

.

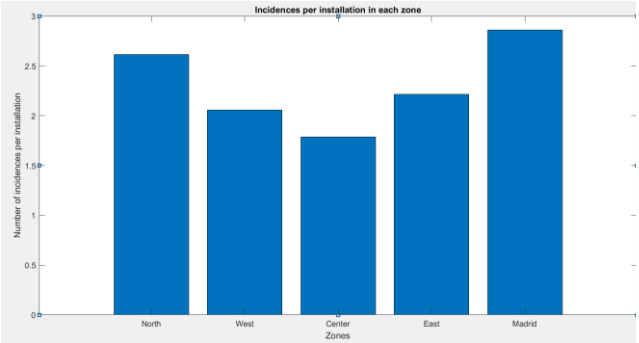

Figure 6. Number of incidences per installation in each zone

Now, it can be seen a more real picture of the situation of incidences in Spain. The area of Madrid is actually the one with more incidences per installation and the Center zone is in the opposite side. The explanation for this is the concentration of devices in a smaller zone, as it happens in urban areas like Madrid, instead of huge territories with a lower density of installation, as it happens in Center zone. A lot of incidences occur because of bad communication of devices that are harder to access when they are too concentrated, so that is the main reason why this is distribution has that shape.

After this analysis, it would be interesting to watch if there are any changes if we filter data between summer and the rest of the year.

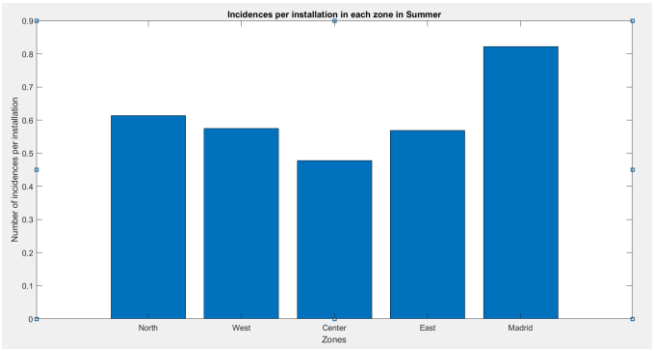

Figure 7. Number of incidences per installation in each zone in Summer

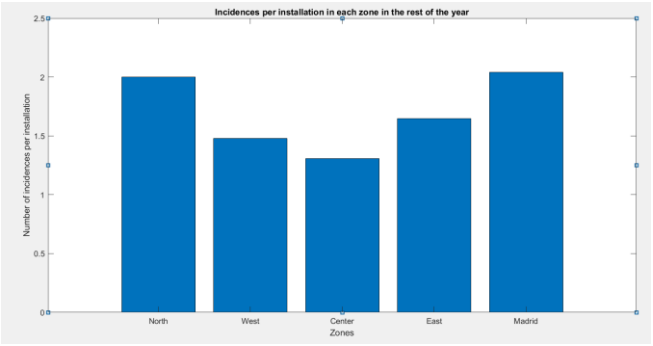

Figure 8. Number of incidences per installation in each zone in the rest of the year

There are no big variations in both situations except for a couple of cases. The proportion only varies because of a little decrease in Madrid and West zone and a bigger increase in North zone. This increase of incidences in the North is due to floods in the rivers of the zone that usually cause malfunctions in the devices nearby. Also, those floods make it difficult to repair the installations and usually the incidences need more actuations until they get completed.

F. Evolution of resolution time

At the time of analyzing the resolution time for the incidences in the system, a "boxplot" graphic was developed in each of the last three year to observe if there are interesting changes and the impact that they had on the grid. With this tool, it is shown the 75 percentile (top blue line), the 25 percentile (bottom blue line) and the average (red line) of the samples, apart from several cases between the 5% and the 25% of appearance.

Evolution of resolution time between 2019 and 2020

Studying the evolution from 2019 and 2020, it can be appreciated the decrease of the resolution time from one year to another in general terms. The average does not change too much due to the high number of incidences in the system, but the percentiles show this reduction of time clearly.

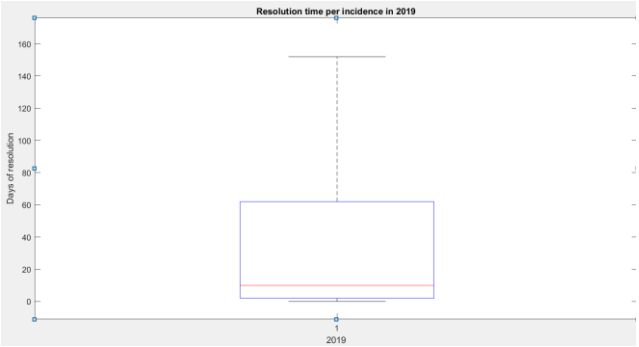

Figure 9. Resolution time per incidence in 2019

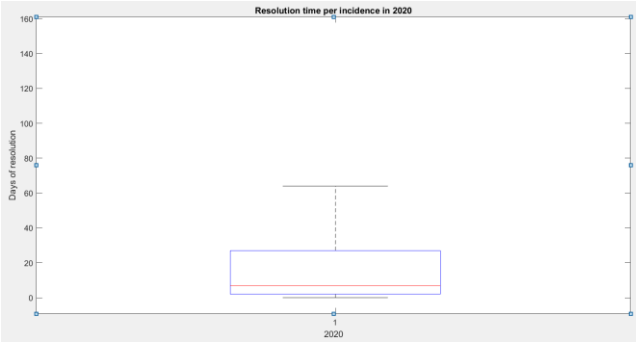

Figure 10. Resolution time per incidence in 2020 (1)

The main reason behind this is the improvement on the resolution processes with better reports, processes and equipment. Also, the experience of operators and digital coordinators is continuously growing so less mistakes and a higher efficiency is achieved.

Evolution of resolution time between 2020 and 2021

It is also interesting to study this evolution of the resolution time in the last year, so it can be analyzed the current situation of the system. The modifications of the processes that have been implemented during this period will be quite similar to those promoted in the next year along with the process of automatization so it can provide interesting information.

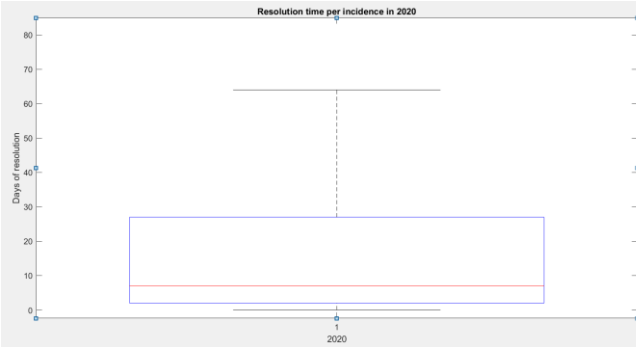

Figure 11. Resolution time per incidence in 2020 (2)

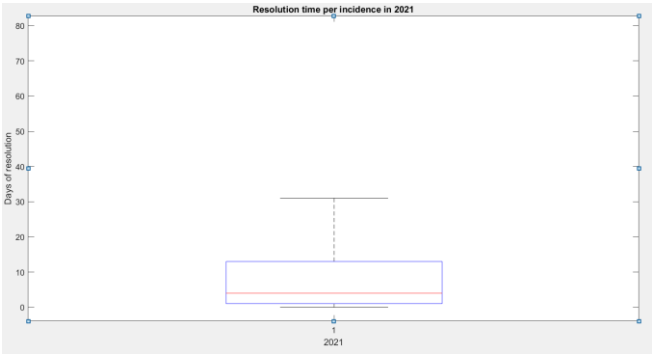

Figure 12. Resolution time per incidence in 2021

There is again a decrease on the resolution time per incidence, which means that the improvements implemented on the different processes have been working correctly. The reasons that explain this evolution are related to modifications and adjustments of the process of generation and discard of incidences. It has been improved so the generation values and the incidences per equipment become more accurate and more realistic compared to the actual situation. The discard process used to create many more incidences than it should because sometimes those discarded incidences showed up again automatically due to a slow update of their status. The system has been optimized so they do not generate again if they are already solved.

VI. Conclusions

After the collection of results and their analysis, there are some conclusions that can be stated regarding the influence of the present report in the later introduction of the RPA technology in STG web. At the time of introducing an automatization methodology in the system, it is crucial to start implementing it in the situations and places where it will be easier to adjust the parts of the system and the results are more favorable. In this case, the project aimed to do a deeper research on the whole STG web to obtain great improvement areas and work methodologies that did not reach the expected performance.

According to the obtained results, there are interesting conclusions that can be assumed. On the one hand, there is a clear pattern in the diagnoses that appear the most in STG web, bad communication between the elements of the grid is the most repeated issue in the system. However, there are multiple cases where these issues are immediately solved and do not require an actuation on them. This is a step in the process that could be optimized by the robot by automatizing that kind of incidences and saving time and effort for the agents, also by adjusting the flexibility of the devices before they call for an incidence, that may be modified to only appear when it is really necessary.

Another interesting aspect that can be applied in the design of the robot is the optimization of the process of actuations of the incidences. This process has the biggest potential to be optimized, but it needs to be analyzed from different perspectives. Attending at those incidences that have more actuations until they are solved, it can be observed that sometimes there is a lack of understanding between the different organisms involved in the process. Disputes between

them at the time of choosing the right criterion in every case and misunderstanding caused by a wrong communication make the process quite longer than it should be resulting in incidences that take months to be solved and multiple efforts and people involved that could be somewhere else doing other tasks.

STG web involves lots of different organisms and agents who theoretically would need to try to find the best solution for the common welfare. There is a clear scheme in STG web that shows the steps to be followed every time an incidence pop ups, however, when humans enter the equation, there can be discussions and differences of criterion that lead to a waste of resources. For this reason, the bot will have to follow a predefined path that, unlike people in the process, will always go on the same direction, without space for misunderstandings. Designing the robot taking into account every possible scenario will lead to a huge optimization of time and cost for STG web.

Urban areas are also another potential field where the system can be improved by the RPA technology. High concentrations of devices usually result in more incidences and warnings, so the robot must help relieving these areas of work for the agents. After the developed study of the system, it can be concluded that there are a lot of similarities between the incidences that are generated in theses areas, they usually include actuations of reviews and modifications that are not efficient until they get to the conclusion that there are too many devices concentrated there and there is the need of applying another measure. If the bot is given this information, it will be able to identify those incidences that are similar between them and create a standard automatized methodology for all of them.

Technology grows exponentially and the number of elements in the grid is going to be increased in the following years, for this reason, the robot must be designed to predict this kind of changes in STG web and to be capable of adapting to the new elements or methodologies that could be implemented in the future. As it could be deduced from the performed analysis in the report, there must be a balance between the growth of the present technologies and elements in the grid and the tools to manage them because if it is not the case, it might lead to a huge increase of incidences in the system that might provoke a collapse in the system.

The implementation of the proposed improvements in the system will not be appreciated until the end of the year 2021 at least. After the delivery of this report, the project will continue and the next step is using all the information and requirements provided here to design and implement the robot in the system. Due to the deadlines of this project, it was not possible to see the changes in the system and to observe if the work is actually well performed, however, according to the feedback of the company it can be stated that the project is truly valuable for Iberdrola and they will keep it for future implementations of automatizations in the system.

REFERENCES

[1] P. Urreta, "Ricoh," September 2019. [Online]. Available: https://digital.ricoh.es/que-es-rpa-automatizar-negocio/.

[2] Deloitte, "Deloitte," 2020. [Online]. Available: https://www2.deloitte.com/es/es/pages/operations/articles/que-es-robotic-process-automation.html.

[3] Penteo, "Intelligent Automation Tools," 2021.

[4] Unir, "Clustering: qué es y cuál es su uso en Big Data," *Unir*, 2021.