NOVA IMS Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

IMPLEMENTATION OF A BUSINESS INTELLIGENCE SYSTEM IN A PUBLIC INSTITUTION

Daniel Leite Teixeira

Project Work presented as partial requirement for obtaining the Master's degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

IMPLEMENTATION OF A BUSINESS INTELLIGENCE SYSTEM IN A PUBLIC INSTITUTION

By

Daniel Leite Teixeira

Project Work presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Business Intelligence & Knowledge Management

Co-Advisors: *Miguel de Castro Neto, Luís Pedro Lopes Batista*

May 2022

ABSTRACT

Technology is capable to support companies to reach their goals by making use data to take advantages from its analysis. Concerning public institutions, it is meaningful to deliver high-quality services and products to society, and technology may lead the organizations to a more efficient service provision.

The purpose of this project is the implementation of a business intelligence system in Consultoria Jurídica da União (CJU) [*Consultancy Office*], the institution responsible for analyzing bidding processes in Brazil. The solution proposed by this work aims to store the business data and provide an analytical tool to display information in dashboards to provide insights to stakeholders, analyzing data trends and tendencies, preventing future unnecessary events, identifying best practices, to finally improve the public tenders to a better application of public funds and provide better services to society.

To reach this project main objective, the technology that surrounds the BI system to be implemented includes the development of a scalable data warehouse to store the organization data and its schema modelling, the extraction-transform-load method to populate the data warehouse tables, and create the analytical tool, named dashboard, to answer the business needs providing the institution information. This business intelligence system intends to improve the legal bidding process in public agencies by making use of technology.

KEYWORDS

Business Analytics; Business Intelligence; Data Warehousing; Data Visualization

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AGU	Advocacia-Geral da União [*Attorney General's Office*]
BI	Business Intelligence
CGU	Consultoria Geral da União [*General Consultancy Office*]
CJU	Consultoria Jurídica da União [*Consultancy Office*]
DW	Data warehouse
ETL	Extract, Transform and Load
IT	Information Technology
PE	Pregão Eletrônico [*Electronic Auction*]
PGF	Procuradoria Geral Federal [*Attorney General for Federal Agencies Office*]
PGU	Procuradoria Geral da União [*Solicitors General's Office*]
SA	Staging Area
SQL	Standard Query Language
SRP	Sistema de Registro de Preços [*Price Registration System*]
SSMS	SQL Server Management Studio
SSIS	SQL Server Integration Services

1. INTRODUCTION

The present work is an implementation of a Business Intelligence (BI) project in Consultoria Jurídica da União (CJU) [*Consultancy Office*], the Brazilian public institution that serves as legal adviser to other agencies when making contracts with third parties through bidding processes. The solution proposed by this work aims to use instruments and techniques learned along the Master's in *Information Management with specialization in Business Intelligence & Knowledge Management* from University of Nova IMS in Lisbon, applying the obtained theoretical knowledge and practical methods to modernize CJU's service and technological structure. The proposed Business Intelligence project aims to provide dashboards to enhance efficiency in bidding processes by providing insights over business analytics, improving the application of public funds, and fostering better service provision to society.

1.1. CONTEXT AND PROBLEM IDENTIFICATION

The World Bank publication (Doing Business, 2020) exposed the bureaucracy while making contracts with the Brazilian government. The study reported that the country takes up to 1.265 days to complete an entire procurement process, since the bid advertisement until the last contract payment, and up to 590 days only for the preparation stage, from the beginning of the process until the contract signing. The data puts the country on the same level as countries like Angola that takes up to 1.263 days to complete the bid process and 573 days for the preparation phase, and higher than Ethiopia that completes the bid processes in 1.036 days and takes up to 348 days for the bid preparation stage. In comparison with other countries in better positions like Portugal (706 total time, 356 preparation phase), Denmark (443 total time, 203 preparation) and Canada (313 total time, 178 preparation) the report displays how far the country is from these indicators and exposes the need of improving the Brazilian government contracting cycle.

Stated by Brazilian Constitution (1998, art. 131), Advocacia Geral da União (AGU) [*Attorney General's Office*] is the Brazilian public institution responsible for the legal defense of the Federal State, representing the Legislative, Executive and Judiciary Branches in court, as well as providing legal guidance to the Executive Branch. It represents the country judicially and extrajudicially, and it is responsible for the activities of legal counsel and advisory services to the Executive Branch.

As the main organism responsible for representing, auditing, controlling and legal protection of Brazil, AGU is composed by three main organizations, that are Procuradoria Geral da União (PGU) [*Solicitor General's Office*], that represents Brazil judicially and extrajudicially, Procuradoria Geral Federal (PGF) [*Attorney General for Federal Agencies Office*], responsible for judicial and extrajudicial representation of autarchies and federal public foundations, and Consultoria Geral da União (CGU) – [*General Consultancy Office*], that for the Federal Government is the agency responsible for the legal consultancy and advisory activities of the Direct Administration. Subordinated to CGU, there are twenty-seven other entities named Consultoria Jurídica da União (CJU) [*Consultancy Offices*], one in each state of the country and another in the Federal District, sharing CGU's same competence.

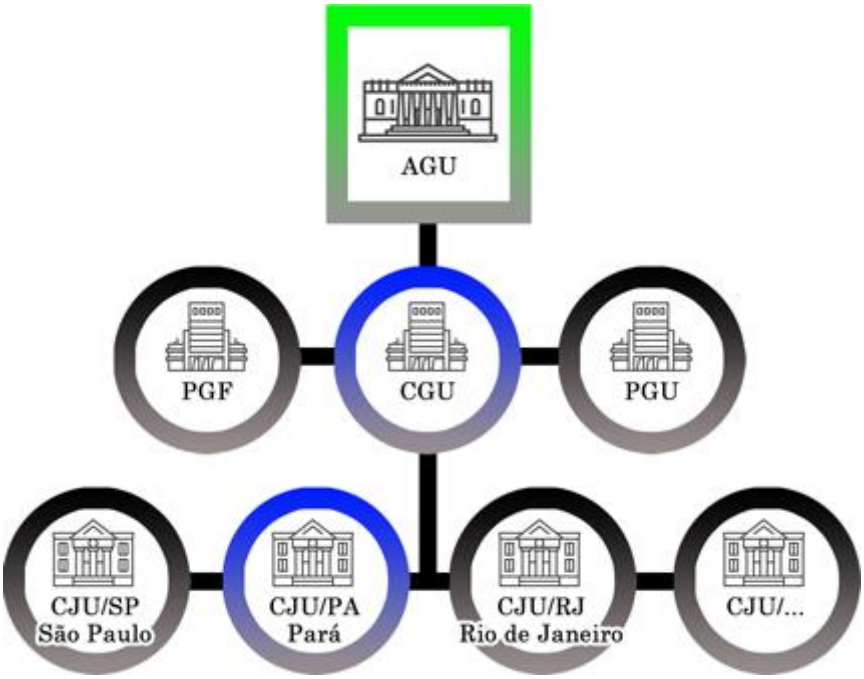


Figure 1-1 - Advocacia Geral da União organizational chart

The Brazilian Complementary Law No. 73 (1993, art. 11) states that among Consultoria Jurídica da União (CJU) [*Consultancy Office*] competences is to examine, previously and conclusively, within the scope of the Ministry, Secretariat and General Staff of the Armed Forces the texts of bid notices, as well as those of the respective contracts or similar instruments, to be published and concluded; the acts that will recognize the unenforceability of or decide to dispense with a bidding process.

The law argues that among AGU duties, it is CJU responsibility to advise the direct administration institutions' bidding processes, agencies responsible for public services directly related to the Executive Branch, namely the Federal Government, its Ministries, Secretaries and Departments. It is mandatory for the direct administration agencies to receive CJU's advice in the preparation phase of public tenders, before publicizing the bidding notes. This project is developed for the Consultancy Office in Pará state – CJU/PA, which is the organization legally responsible for analyzing the bidding processes from public agencies in the state.

CJU first receives their documentation to examine it, then provide them further report with the results of the legal analysis. Depending on this feedback, the bidding process is publicized or not by the requesting institution. If by any reason there is a need of any correction to be made or lack of documents, CJU reports the agency to make it. After it is done, the agency sends back to CJU the documentation with the adjustments, to be reanalyzed. If CJU informs that the bidding is eligible, then the agency follows the rite to publicize the bidding documentation, to receive companies' proposals and judge the better offer to sign the contract. On the other hand, if the documentation is disapproved by the consultants, the agency needs to make a brand-new process to reach the bid purpose.

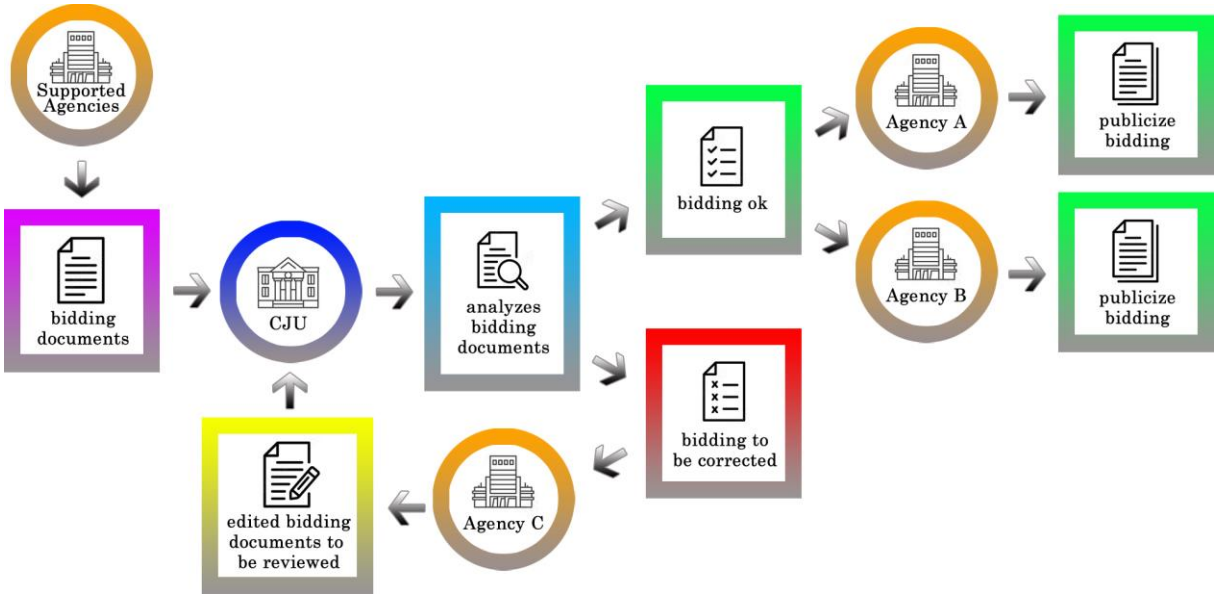


Figure 1-2 - Flowchart of the analysis of Brazilian bidding processes

Bidding processes and contracts of direct administration institutions are regulated by the Law 8.666 (1993, art. 2nd) claiming that buildings, services, including publicity, purchases, disposals, concessions, permissions, and leases of the Public Administration, when contracted with third parties, will necessarily be preceded by a bidding process. The mentioned Law relates five types of bidding, and among them, the most conventional and relevant modalities for CJU routine and due to the workload are Sistema de Registro de Preços (SRP) [*Price Registration System*] and Pregão Eletrônico (PE) [*Electronic Auction*], that are the ones related in this project.

- Sistema de Registro de Preços (SRP) [*Price Registration System*] or Pregão Registro de Preços [*Price Registration Auction*] modality stated by Decree 7.892 (2013, art. II) is adopted when there is a hypothesis of repeated buying or hiring of a good or a service, and when there is a forecast of products deliveries from time to time like buying and receiving some equipment, making possible to previously predict the quantity needed of a good or its delivery frequency.

- Pregão Eletrônico (PE) [*Electronic Auction*] on the other hand is the "*mandatory bidding modality for the acquisition of common goods and services, whose judgment criterion may be the lowest price or the highest discount*" (Law No. 14.133, 2021, cap. 3 XLI), translated from Portuguese language by the author from "*pregão: modalidade de licitação obrigatória para aquisição de bens e serviços comuns, cujo critério de julgamento poderá ser o de menor preço ou o de maior desconto*" (Lei No. 14.133, 2021, cap. 3 XLI). The bidding in the Auction modality, in the electronic form, does not apply to the contracting of engineering works, as well as to real estate leases and disposals in general.

For the public administration biddings, the main difference between the mentioned modalities is that in Pregão Eletrônico (PE) [*Electronic Auction*], there is only one agreement assigned between a public agency and the third party. Whereas making use of Sistema de Registro de Preços (SRP) [*Price Registration System*] or Pregão Registro de Preços [*Price Registration Auction*], it is possible to make several contracts for one the same bidding, to buy and receive an amount of a good or service in different dates but respecting the quantities available for the contract and to its expiration date.

As exposed by the World Bank report, the governmental contracting cycle is too delayed in the country, and within CJU context, while working on documentation analysis the consultants noticed that some of the supported agencies used to repeat the same errors among several different bidding documents, stating that the agencies were not learning from their past mistakes, and it was making the whole process of bidding more delayed and expensive, as the agencies should make corrections recurrently. From the first documentation delivery, receive the report, make the corrections, send it again to CJU review, receive again the documents and publicize it, the process flow was too time-consuming due to this rework needed not only from CJU but also from the requesting agencies. This rework influences the bureaucracy as the analysis made by the legal consultants occurs in the preparation stage, exactly before publicizing the auction notices.

CJU/PA – the Consultancy Office in Pará (PA) state highlighted in Figure 1-1, expressed the need of a technological solution to store the institution information, mainly the bid documents data, to create dashboards that should provide insights about the office's periodic work results, especially about the public tender analysis. To accomplish all the business needs, a business intelligence system was implemented initially to CJU/PA as result of this project, with a projection to be scaled to other CJU Offices countrywide in the future. In the next chapters, when this work states CJU it means CJU/PA.

1.2. PROJECT OBJECTIVE

This work objective is to produce a business intelligence system by making use of tools and techniques learned during this Master at Nova IMS in Lisbon, grounded on the laws that dictate the rite while contracting with the Brazilian government. After identifying the need of a solution for the exposed problem, this project aims to raise the quality of the bidding documents in Brazil, to improve the bureaucracy and promote best practices by sharing knowledge between the agencies supported by CJU.

The processes are stored in data warehouses, loaded after running ETL processes. Later, the project provides dashboards containing the ranking of all agencies and the scores of each legal bidding process, to assist the supported agencies by monitoring its own performance in comparison with the others, and to help decision making by providing descriptive analysis. The dashboards are regularly created when each project cycle ends, and it will make use of data since the last months of 2019.

To reach this project goals and provide an innovative solution, it is strictly necessary to build a consistent business intelligence system using tools available in the organization. The solution proposed by this work records the legal processes information, mainly its errors and achievements, runs an algorithm that provides scores to the processes and the agencies ranking. To provide the analytics as required, the solution needs a data warehouse to store the organization work, especially detailed information of the bid processes as reported by the CJU consultants, allowing fast queries and historical analysis guaranteed by enhanced conformity, quality, and secure data to be scaled whenever needed.

The data repository is fulfilled with data by making use the Extraction-Transform-Load (ETL) processes. The first step called Extraction occurs from transactional data that are lists on the intranet of the institution that are the initial data source as it keeps the information from bid documents and reports. The Transformation takes place to make the adjustments needed in this source to finally Load it into the data warehouse.

The analytical tools, named dashboards, are created to answer the business needs using the data stored in the data warehouse. The dashboards are the project last stage as they should provide the aimed efficiency in the whole bid process by promoting insights over the data gathered regularly. On behalf of governmental transparency (Law No. 12.527, art. 3rd) the dashboards are available to everyone as soon as they are released on AGU's national website, as the bid deals directly with public funds and services provided to citizens.

Part of the institution culture is to promote good practices, and in accordance with this practice, when a year (or "macrocycle") end, other dashboards are publicized with the results from the whole period. The best ranked agencies are prized by the heads of CJU, and they are formally invited to share their knowledge and good practices with all the agencies in the beginning of the next year. As the provided dashboards are the artifacts created as final products, the methodology applied is the Design Science Research.

The main question surrounding the present work is “how to improve the bidding processes in Brazil by making use of technology?”. Based on the question and to reach this project goals, the research objectives are to:

1. Build a scalable data warehouse to store information from bid documents
2. Create dashboards to provide analytics
3. Improve the bidding processes

1.3. STUDY RELEVANCE

Brazilian public institutions only make contracts to acquire good or services through bidding processes. Within the country, CJU is the organization responsible for producing reports on the legal bidding documentation from those institutions, specifically the direct administration ones. Making use of technology to improve the bidding flow means a better application of public funds and a better service provision to society, as it may provide faster contract signing with the third parties, fostering the receiving of the good or services as hired in a shorter period then it was before.

In this context, this project relevance is the improvement of the Brazilian bidding processes by providing analytics on dashboards, based on the reports produced by CJU consultants. Turning data into information, knowledge, and wisdom, driving the stakeholders to make decisions and take actions based in data gathered across time. Raising the public tender’s quality means to raise the quality of public services provided to society when acquiring goods or hiring services. Lack of documents bring slowness to the whole contracts cycle, also mistakes committed by governmental agencies may lead the process to legal action, and the use of technology may lead Brazilian public institutions to a more efficient service provision.

The goal of the analytic tool should be finally to explicitly bring the agencies errors and achievements, committed by themselves or by the others, to induce the supported agencies not only to learn from past mistakes but also to identify good practices, promoting insights over the analytics to enhance their own efficiency and raise the quality of the public service by fostering the institutional learning, particularly while conducting bidding processes.

This work resulted on a project called Programa de Aprendizagem Institucional [*Institutional Learning Project*] initially deployed at CJU in Pará state. Based on the good results obtained, there is an institutional projection of scaling it to other states in the country. The system promoted a significant evolution in the documents quality evidenced by the improvement of the agencies’ performance, aligned with the growth of the modalities’ scores. In two years of system use, Electronic Auction grew up from 82,65 to 108,51 points, and Price Registration System modality from 81,16 to 119,47 points.

This document is structured as follows. The first chapter brings the introduction with an overview of the work done, the context, motivation, objective e relevance. The second chapter is the Literature Review, that presents concepts and practical approaches to develop the proposed BI system. The third chapter is the Methodology where the implementation takes place. The fourth chapter presents the results accomplished, and the fifth is the conclusion. Finally, the six chapter describes the limitations and recommendations for future work.

2. LITERATURE REVIEW

To adapt to a new way of thinking about the business strategy, nowadays it is important to state information technology as an important part of the companies, as the stored data can turn into useful information to help them reach their objectives. BI is an important area of Information System due to the potential to analyze and solve problems of the company (Fourati-Jamoussi & Niamba, 2016). Regarding public agencies, it is important to deliver high-quality products and services, and business intelligence technology may help them to reach it. For Watson (2009) "Business intelligence (BI) is a broad category of applications, technologies, and processes for gathering, storing, accessing, and analyzing data to help business users make better decisions", a BI solution should improve the institution and the agencies by making use of data to support decisions. This work will focus on the most relevant tools and techniques to be applied to reach its objectives.

2.1. DATA WAREHOUSE

Golfarelli and Rizzi (2009) states that "Data warehouses are information repositories specialized in supporting decision making. Since the decisional process typically requires an analysis of historical trends, time and its management acquire a huge importance". Data warehousing is a collection of decision support technologies, aimed at enabling the knowledge worker (executive, manager, analyst) to make better and faster decisions. It serves as a physical implementation of a decision support data model and stores the information on which an enterprise needs to make strategic decisions (Reddy et al., 2010). To provide insights on the dashboards provided by this project, it is necessary to store the data from the reported processes in the data warehouse, as it should be the main repository of information to be analyzed by the stakeholders for a more appropriate and assertive decision-making, based in the data collected and stored.

For most organizations, managing data takes on two predominant forms: 1) operational systems to deal with highly capacious transactional processing using real time data, and 2) data warehouses to facilitate information access by providing a centralized database for all enterprise data organized in a manner specifically for querying (Kimball, 1996). A data warehouse is a "subject-oriented, integrated, time varying, non-volatile collection of data in support of the management's decision-making process" (Inmon, 1996). The data warehouse stores information in a central repository updated periodically with fast access, scalable, standardized, and consistent.

Inmon proposes an enterprise-wide data warehouse that feeds departmental databases, data-driven, considered a top-down architecture as it should provide a complete technological solution with initial high costs. On the other hand, Kimball states about an evolutionary data warehouse bottom-up designed, starting with departmental data marts focused on the end-user since the beginning of the project, process oriented and with lower costs, that may become higher in a future DW evolution as needed.

The data warehouse will serve as the heart of the project, as it will store and provide the organization data that may lead to business analysis. In comparison of the essential features of the two main studies on the topic proposed by Kimball and Inmon. Kimball's is the one that better adapts to this project needs proposing a bottom-up approach, arranged in data marts provided to specific

organization areas or subjects like human resources or inventory, and in this case will be created for CJU consultants. For Yessad & Laboid (2016) Kimball presents a different vision of data warehouses. He considers that the data warehouse can be seen as a set of consistent data marts and based on shared conformed dimensions.

In data warehouse environment "history" and "time" are unavoidable factors to consider at the data warehouse design time. Incremental loading, scheduled on a regular basis, into a target table that cumulates historical data (usually fact table) is typical situation in Data Warehouse environment (Adzic, Fiore & Spelta, 2001). Kimball (1996) describes fact as a dimensional table that stores numerical performance measurements, such as sales amount "We use the term fact to represent a business measure. We can imagine standing in the marketplace watching products being sold and writing down the quantity sold, and dollar sales amount each day for each product in each store. A measurement is taken at the intersection of all the dimensions (day, product, and store)". The fact table should contain the identifiers for the dimension members rather than their descriptions (Larson et al., 2012). The fact tables have two or more foreign keys that are correlated to the dimensions' primary keys, that are assigned as foreign keys in the fact.

In 2009, Watson said that "Dependent data marts often store data in a multidimensional or star schema format. With this data model, fact tables store numerical values and dimension tables store measures that are used to qualify the queries". Fact tables contain metrics, while dimension tables contain attributes of the metrics in the fact tables (Breslin, 2004), and for this purpose the dimensions must be normalized, and a concrete schema must be designed.

- **Star Schema:** If the presentation area is based on a relational database, then these dimensionally modeled tables are referred to as star schemas (Kimball, 1996). The center of the star is formed by the fact table. The fact table has a column for the measure and a column for each dimension containing the foreign key for a member of that dimension. The primary key for this table is created by concatenating all the foreign key fields. The center of the star is formed by the fact table. The fact table has a column for the measure and a column for each dimension containing the foreign key for a member of that dimension (Larson, 2012).

- **Snowflake Schema:** The snowflake model is the result of decomposing one or more of the dimensions, which sometimes have hierarchies themselves. We can define the many-to-one relationships among members within a dimension table as a separate dimension table, forming a hierarchy (Ballard et al., 1998). Star schemas can be refined into snowflake schemas providing support for attribute hierarchies by allowing the dimension tables to have subdimension tables (Lavene & Loizou, 2003).

In a well-designed dimensional model, dimension tables have many columns or attributes. These attributes describe the rows in the dimension table. We strive to include as many meaningful text-like descriptions as possible (Kimball, 1996). If a fact table is described by measurements, the dimensions categorize and describe data to make the data warehouse information understandable as they keep and explain relevant attributes like time, region, products, and customers. The types of fact tables are described by Ross & Kimball (2013) as follows:

- **Transactional Fact Table:** A row in a transaction fact table corresponds to a measurement event at a point in space and time. Atomic transaction grain fact tables are the most dimensional and expressive fact tables; this robust dimensionality enables the maximum slicing and dicing of transaction data. Transaction fact tables may be dense or sparse because rows exist only if measurements take place. These fact tables always contain a foreign key for each associated dimension, and optionally contain precise time stamps and degenerate dimension keys. The measured numeric facts must be consistent with the transaction grain.

- **Periodic Snapshot Fact Table:** A row in a periodic snapshot fact table summarizes many measurement events occurring over a standard period, such as a day, a week, or a month. The grain is the period, not the individual transaction. Periodic snapshot fact tables often contain many facts because any measurement event consistent with the fact table grain is permissible. These fact tables are uniformly dense in their foreign keys because even if no activity takes place during the period, a row is typically inserted in the fact table containing a zero or null for each fact.

- **Accumulating Snapshot Fact Table:** A row in an accumulating snapshot fact table summarizes the measurement events occurring at predictable steps between the beginning and the end of a process. Pipeline or workflow processes, such as order fulfillment or claim processing, that have a defined start point, standard intermediate steps, and defined end point can be modeled with this type of fact table.

- **Factless Fact Table:** Although most measurement events capture numerical results, it is possible that the event merely records a set of dimensional entities coming together at a moment in time. For example, an event of a student attending a class on a given day may not have a recorded numeric fact, but a fact row with foreign keys for calendar day, student, teacher, location, and class is well-defined. Likewise, customer communications are events, but there may be no associated metrics.

The dimension model to be used to build the data warehouse is the star schema by making use of factless fact table, as there are no measurements while recording legal processes data, and it is important to track the events periodically. The dashboards are periodically published from the data recorded in cycles, and a next cycle dashboard bring comparison with the previous (further explanation about the cycles in the next chapter). Once the year is finished, a macrocycle dashboard is published displaying the data from the whole period.

2.2. ETL PROCESSES

After the modelling and before inputting data into the DW it is necessary to run the ETL processes. Demarest states that it is possible that this stage takes up to 80% of the effort time of the whole project (Demarest, 1997). The initial data sources are two Sharepoint lists created for each bid modality previously mentioned that are Sistema de Registro de Preços (SRP) [*Price Registration System*] and Pregão Eletrônico (PE) [*Electronic Auction*] in the CJU's intranet. The bidding documents and reports are recorded by administrative peers in web forms, that serves as inputs in these lists. Whenever a form is sent, one new row is created in the related list, according to the bidding type.

Technologically it flows from MS Forms to MS Sharepoint lists by making use of an instruction in MS PowerAutomate that creates a new line in the list whenever a form is sent. The lists are the data sources from where the Extraction will act, and that will Load the data warehouse later.

To create a data warehouse, one of the most important aspects is the Extraction-Transform-Load, usually named "ETL" processes. This stage aims to gather data from sources, transform the data to be as structured as the data warehouse needs due to standardization and normalization processes, to finally load it into the data warehouse fulfilling its dimensions and facts tables. ETL is essential to the viability of the data warehouse in that it attempts to ensure data integrity within the data warehouse (Breslin, 2004). The source types vary from business and application. For example, Nath, R. P. D., Hose, K., Pedersen, T. B., & Romero, O. (2017) states about non-semantic data sources that it "can be a relational database (RD), a shapefile, an XML file, an object-oriented database, or a CSV. A CSV file is a comma separated values file, which allows data to be saved in a table structured format. Each line of the file is a row of the table, and each row consists of one or more fields, separated by commas".

Inmon (2005) states that ETL processes as it "automates the process of converting, reformatting, and integrating data from multiple legacy operational sources" and it should be executed before the data flowing into the data warehouse. When the Extraction is completed, the Transformation starts in order to normalize/organize/standardize data before finally Loading it into the DW. One of the challenges to be faced is to think about how the data will flow from these actual sources in Sharepoint lists by making the transformation needed to successfully fulfill the data warehouse.

- **Extraction:** Extraction is the first step to acquire the data from the sources as stated by Kakish & Kraft (2012) "In general, the goal of the extraction phase is to convert the data into a single format which is appropriate for transformation processing". Extracting means reading and understanding the source data and copying the data needed for the data warehouse into the staging area for further manipulation (Kimball, 1996). The Staging Area is a storage used to keep the intermediate results of the ETL sub-processes (Nath et al., 2017). The incremental extraction is called changed data capture (CDC) where the ETL processes refresh the DW with the modified and added data in the source systems since the last extraction. This process is periodic according to the refresh cycle and business needs. It also captures only changed data since the last extraction by using many techniques as audit columns, database log, system date, or delta technique (El-Sappagh, S. H. A., Hendawi, A. M. A., & El Bastawissy, A. H., 2011).

- **Transform:** There are many variations of data-level conflicts across sources: duplicated or contradicting records, different value representations (e.g., marital status), different interpretation of the values (e.g., measurement units dollar vs. euro), different aggregation levels (e.g., sales per product vs. sales per product group), or reference to different points in time (e.g., current sales as of yesterday for source 1 vs. Current sales as of last week for source 2). The list is enriched by low-level technical problems such as data type conversions, applying format masks, assigning fields to a sequence number, substituting constants, setting values to NULL or DEFAULT based on a condition, or using simple SQL operators (e.g., UPPER, TRUNC, SUBSTR) (Vassiliadis, P., & Simitsis, A., 2005). After the extraction phase, it is necessary to do some transformation to choose the strictly necessary data, like selection of

certain columns, generating surrogate keys, calculate values, and generate measures. The transform stage applies a series of rules or functions to the extracted data from the source to derive the data for loading into the end target. Some data sources will require very little or even no manipulation of data (Kakish & Kraft, 2012). This process includes data cleaning, transformation, and integration. It defines the granularity of fact tables, the dimension tables, DW schema (stare or snowflake), derived facts, slowly changing dimensions, factless fact tables (El-Sappagh et al., 2011). Once the data is extracted to the staging area, there are numerous potential transformations, such as cleansing the data (correcting misspellings, resolving domain conflicts, dealing with missing elements, or parsing into standard formats), combining data from multiple sources, deduplicating data, and assigning warehouse keys. These transformations are all precursors to loading the data into the data warehouse presentation area (Kimball, 1996).

- **Load:** After completing the transformation process, the data will be loaded as the data warehouse is created. The load process in either case involves placing the data physically. The main concern in this process is appending the newly extracted and transformed data onto the data already in the data warehouse (Breslin, 2004). Regardless of whether we're working with a series of flat files or a normalized data structure in the staging area, the final step of the ETL process is the loading of data. Loading in the data warehouse environment usually takes the form of presenting the quality-assured dimensional tables to the bulk loading facilities of each data mart (Kimball, 1996). Bulk loading is performed (a) during the very first construction of the warehouse and (b) in an incremental way, during its everyday maintenance. During the latter task, a set of new insertions, deletions and updates arrive at the warehouse, after being identified at the extraction phase and subsequently transformed and cleaned (Vassiliadis, 2009). The designer needs to determine the frequency of the ETL process, so that data are fresh, and the overall process fits within the refreshment time window (Vassiliadis, P., Simitsis, A., & Skiadopoulos, S., 2002). Another problem is discriminating between new and existing data at loading time. This problem arises when a set of records has to be classified to (a) the new rows that need to be appended to the warehouse and (b) rows that already exist in the DW but whose value has changed and must be updated (e.g., with an UPDATE command). Modern ETL tools already provide mechanisms towards this problem, mostly through language predicates (Vassiliadis, P., & Simitsis, A., 2005).

2.2.1. Types of ETL

Whatever the information extraction algorithm used, the source data always needs to be pre-processed to get rid of noise, transform it to the representation required by the method, etc. In addition, the output also needs to be post-processed to gather the structured data in the form of attribute-value pairs, which can then be transformed and loaded into the data warehouse. In the case of text, this pipeline of tasks corresponds to what has recently been named textual-ETL (Inmon & Nesavich, 2007). The challenge in textual-ETL consists in identifying how to abstract all the above tasks into operators that can be used to design, optimize, and execute these flows in the same way as for structured data (Dayal, U., Castellanos, M., Simitsis, A., & Wilkinson, K., 2009).

Although the ETL stages have the same previous explanations, there are differences in the implementation about its execution order or frequency, and the type of ETL chosen as it must be the one that better adapts to the business or project.

- **ETL or classical ETL:** After the extraction, the transformed data are kept into a Staging Area. The Staging Area is an intermediate storage where the intermediate results of each process are stored. This can prevent the loss of transformed data in case of the failure of the loading process (Nath et al., 2017).

- **ELT:** In the ELT style, after the extraction phase, data is loaded directly into the data warehouse server and all the transformations are executed there (Dayal et al., 2009). The basic idea is to perform Load process immediately after Extract process and apply Transformation only after getting data stored. ELT, in comparison with ETL, has four following advantages: (1) flexibility in adding new data sources (EL part); (2) aggregation can be applied multiple times on same raw data (T part); (3) Transformation process can be re-adopted, even on legacy data; (4) speed-up process of implementation (Marín-Ortega et al., 2014).

- **ETLT:** In the ETLT case, the idea is to split the transformation phase into two groups of transformations, the first to be executed immediately after the extraction, and the second to be executed after a loading phase. The main argument for these alternatives is that the data warehouse servers are usually scalable, highly parallel machines and in principle could be better at optimizing transformations (Dayal et al., 2009).

2.3. DASHBOARDS AND DATA VISUALIZATION

Once the data warehouse is ready and fully populated, it is possible to obtain information from it and develop the visualization tool, named dashboards, to answer the business needs. Despite providing the answers, one of the challenges faced is to display the information in a clean and easy to interpret manner. The stakeholders need to visualize the whole data gathered from the analyzed processes, its errors and achievements as reported, which means too much information to be shown in graphics, that maybe need several charts to visualize through many pages. For Cairo (2015), while creating a visualization “don’t think just about yourself, the designer, when evaluating your own decisions while creating a graphic. Think of who’s going to decode it.”

Data visualization is concerned with the design, development, and application of computer-generated graphical representation of the data. It provides effective data representation of data originating from different sources. This enables decision makers to see analytics in visual form and makes it easy for them to make sense of the data. It helps them discover patterns, comprehend information, and form an opinion (Sadiku et al., 2016). Dashboards represents the actual state of the organization internal processes or situation and also provide KPI (Key Performance Indicators) fundamental to performance analyzing but is necessary to produce it in high quality to succeed the needs.

Data visualization uses computer graphics to show patterns, trends, and relationship among elements of the data. It can generate pie charts, bar charts, scatter plots, and other types of data

graphs with simple pull-down menus and mouse clicks. Colors are carefully selected for certain types of visualization. When color is used to represent data, we must choose effective colors to differentiate between data elements. In data visualization, data is abstracted and summarized. Spatial variables such as position, size, and shape represent key elements in the data. A visualization system should perform a data reduction, transform, and project the original dataset on a screen (Sadiku et al., 2016).

For Cairo (2013), data visualization is "something that achieves beauty not through the subjective, freely wandering self-expression of the painter or sculptor, but through the careful and restrained tinkering of the engineer." Data visualization is the process of representing data in a graphical or pictorial way in a clear and effective manner. It has emerged as a powerful and widely applicable tool for analyzing and interpreting large and complex data. It has become a quick, easy means of conveying concepts in a universal format. It must communicate complex ideas with clarity, accuracy, and efficiency. These benefits have allowed data visualization to be useful in many fields of study (Sadiku, 2016). Wolfe (2015) states that "data visualization, then, does not just involve representing a given set of numbers; it involves selecting and rethinking the numbers on which the visualization is based".

Graphs and charts let you explore and learn about the structure of the information you collect. Good data visualizations also make it easier to communicate your ideas and findings to other people. Beyond that, producing effective plots from your own data is the best way to develop a good eye for reading and understanding graphs—good and bad—made by others, whether presented in research articles, business slide decks, public policy advocacy, or media reports (Healy, 2018). Stephen Few (2005) says that "the fundamental challenge of dashboard design is to display all the required information on a single screen, clearly and without distraction, in a manner that can be assimilated quickly".

Sample statistics express the characteristics of a sample using a limited set of parameters. They are generally seen as estimates of the corresponding population parameters from which the sample comes from. These characteristics can express the central tendency of the data (arithmetic mean, median, mode), its spread (variance, standard deviation, interquartile range, maximum and minimum value) or some features of its distribution (skewness, kurtosis) (Komorowski, M., Marshall, D. C., Salciccioli, J. D., & Crutain, Y., 2016). For Cairo (2015) most datasets are too complex to be scrutinized directly (try to extract meaning from a table with hundreds of numbers), so we use and manipulate statistics—mean, median, standard deviation, etc.—to analyze them.

The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented. Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graphic itself. Label important events in the data (Tufte, 2001).

Stephen Few (2005) argue that if the user needs to scroll down, it is not a dashboard. And in this case, it is necessary to navigate through several pages to gather the whole information. On the other hand, Kosara & Mackinlay (2013) calls this type of visualization as slideshow, and states that it facilitates telling a good story. The fact is that there is a need to find a balance between both statements to provide all the information needed in a dashboard as clean as possible, and easy to use and to interpret even to someone using the tool for the first time.

3. METHODOLOGY

Regarding the instruments that will be used to achieve this project's objectives, the first step is to collect information with the head of CJU about the business needs and relate it with the literature review, to obtain more knowledge about the best methods that can be used to reach this project objectives making use of the technology available in the institution.

3.1. CJU NEEDS AND DECISIONS

Within AGU strategic planning published in the report Relatório de Gestão (Advocacia Geral da União, 2020), there are two topics that relates technology with public governance in the institution:

- Strengthen governance: To promote and disseminate good practices of public governance, aiming at improving the capacity to implement, monitor and evaluate the institutional strategy, with a focus on innovation and reducing bureaucracy.
- Fostering Digital Transformation: Disseminate the use of information technology to improve governance, work processes, and knowledge management, focusing on improving the services provided by AGU.

When consultants realized some repeated errors in different documents from a same agency, they understood that the public tender cycle was resulting on rework and becoming too time-consuming. The main consequence is the bureaucracy for the government while making contracts with third parties. As CJU is the institution responsible for analyzing those documents from the agencies, there was a need of improving the whole bidding cycle by raising the documentations quality. By making use of technology, it was possible to create a business intelligence system subject of this study, in accordance with CJU needs and decisions, as related as follows.

3.1.1. Cycles and Macrocycles

To monitor the agencies performance over the year based on the workflow, the consultants decided that the reports should be published in cycles of two or three months, explained next. Every time one cycle finishes, two dashboards must be published, one for each bid modality: one for SRP [*Price Registration System*] and another for PE [*Electronic Auction*]. In the end of each macrocycle, one main dashboard is published covering the information from the whole period. The first macrocycle was called 2019-2020 because the project started to collect data in October 2019. The cycles and macrocycles were divided as follows.

For the 2019-2020 macrocycle:

- 1st Cycle: From October to December 2019
- 2nd Cycle: From January to March 2020
- 3rd Cycle: From April to June 2020

- 4th Cycle: From July to September 2020
- 5th Cycle: October and November 2020

For 2021 macrocycle, the program cycles were:

- 1st Cycle: From December 2020 to February 2021
- 2nd Cycle: From March to May 2021
- 3rd Cycle: From June to August 2021
- 4th Cycle: From September to October 2021
- 5th Cycle: November and December 2021

From 2022 onwards there are still five cycles, but it begins to comprise the exact period of the year, as the macrocycle 2021 above finished in December.

3.1.2. Meetings and Prizes

When each cycle and macrocycle end, CJU consultants schedule meetings with all the supported agencies to show the dashboards and discuss the results with everyone. If on the one hand this institutional program came up from the agencies' repeated mistakes, on the other hand the top ranked agencies in a macrocycle are kindly invited to make a presentation to spread their good practices to all participants. This is the main reason why this project is called *Institutional Learning Program*, to promote the knowledge sharing among the supported agencies to raise the processes' quality overall.

The macrocycle dashboard is published in the beginning of the next year based on the data collected from the cycles produced in the previous year. Beside the meetings to share knowledge among the participants, to motivate the stakeholders the best ranked agencies in a macrocycle are prized by CJU in two categories, defined as the institutions with the highest evolution, and the winner in each bid modality.

3.2. RECORDING THE BIDDING PROCESSES DATA

The data warehouse is populated after the Extraction-Transform-Load process, fulfilled after the Load stage finishes by running the SQL Integration Services in MS Visual Studio. The database architecture is designed in MS SQL Server, and the data is finally loaded inside it when the ETL process ends. Once the DW is ready and the tables are populated, Power BI connects to it and the dashboards are designed by making use of Power Query tool to answer the business needs and then display the information as required. Next there is the explanation how this flow works since the beginning of the document's analysis made by the CJU consultants until the dashboards' publication.

3.2.1. The Analysis Flow of the Bidding Processes

Once the documentation is received at CJU office, it is distributed (delivered) to a lawyer to make the analysis and to produce a report on it. After it is done, this report and the original bidding document is sent back to the requesting agency containing the results of the analysis, approving, disapproving, requiring corrections as needed or asking for missing documents. Meanwhile, before returning the results to the institution, administrative peers record the bidding analysis report in an intranet MS Form. Whenever a form is sent, one new line is created in a Sharepoint list making use of MS Power Automate tool.

When every cycle is opened, one form and one list are made available for each type of bid – SRP [*Price Registration System*] and PE [*Electronic Auction*]. These lists serve as the initial data source in each cycle, and later are exported in a compatible format to be handled through ETL processes, to finally populate the data warehouse.

Once the data is ready, the system make use of MS Power BI to collect the information stored in the DW and trains it to produce two dashboards from a closed cycle, one for each bid modality. The dashboards are finally published on AGU national website and presented on further meeting with the supported agencies about the obtained results.

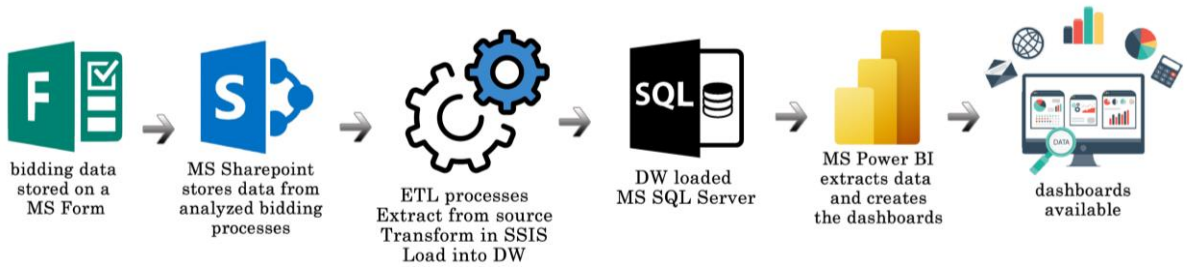


Figure 3-1 - The BI system flow from the MS Forms to the dashboards

3.2.2. Web Forms and Matrices

The web forms are provided as checklists designed by CJU consultants based on the Law 8.666 (1993) and in documents analysis. It is organized in main topics or themes named “matrices”, with its intrinsic related items. Those matrices are the most important part of the forms as they list the items that are evaluated to produce the process’ score and the agencies ranking. In addition to the matrices, the forms have other main fields such as the Process Number (unique number that legally identifies the bidding documentation), Distribution Date (when the consultant received the documents), Analysis Date (report date), and the Agency Name (which requested the analysis). Within the form there is also a Disapproval field – and its related items, checked when the bid is reported as completely unusable. In this case, the report orders the applicant institution to produce a brand-new documentation for that specific bid subject.

Even though those main fields and matrices have the same nomenclature for both SRP and PE web forms, the items evaluated by the consultants to be checked inside the matrices are not the same for each modality. For both forms there are six numbered matrices, but some of them are subdivided according to the bidding type, and specifically in the Price Registration System form there is one

extra matrix name “Acts” that must be part of the documentation. Below there are tables explaining only the main fields in SRP and PE forms – both forms are fully detailed in the annexes 8.1 and 8.2.

Field	Description
Process Number	The number of the legal bidding process
Distribution Date	When the consultant received the documents
Analysis Date	Report date
Legal Consultant	Lawyer responsible for the analysis
Institution Name	Which required the analysis
Disapproval	Relevant items that totally disapprove the documentation
Matrix 1 [M1]	Formalities and instruction of the process
Matrix 2 [M2]	Contract planning
Matrix 3 [M3]	Statements
Matrix 4 [M4]	Quotation and budgets
Matrix 5 [M5 Q]	Public notice qualification drafts
Matrix 5 [M5 E]	Public notice elements drafts
Matrix 6 [M6 TR]	Term of reference
Matrix 6 [M6 TC]	Term of contract
Matrix 6 [M6 A]	Acts

Table 3-1 - Main fields available for the SRP web form

Field	Description
Process Number	The number of the legal bidding process
Distribution Date	When the consultant received the documents
Analysis Date	Report date
Legal Consultant	Lawyer responsible for the analysis
Institution Name	Which required the analysis
Disapproval	Relevant items that totally disapprove the documentation
Matrix 1 [M1]	Formalities and instruction of the process
Matrix 2 [M2]	Contract planning
Matrix 3 [M3]	Statements
Matrix 4 [M4]	Quotation and budgets
Matrix 5 [M5 Q]	Public notice qualification drafts
Matrix 5 [M5 E]	Public notice elements drafts
Matrix 6 [M6 TR]	Term of reference drafts
Matrix 6 [M6 TC]	Term of contract drafts

Table 3-2 - Main fields available for the PE web form

The main fields Process Number, Distribution Date, Analysis Date, Legal Consultant, and Institution Name are single line option. The remaining fields such as the Disapproval field and all the Matrices are made of several items disposed as checklists, and on the top of each matrix' items list there is one "Able" single option that must be checked when a matrix complies all the requirements. In the case of the whole documentation disapproval, one single item checked in the Disapproval related field turns the process score to zero, but still there is a need of checking the remaining matrices to point the errors for the statistics.

One bidding process may be approved even with some mistaken documents, but it is completely disapproved in case of missing some meaningful documents. An approved documentation reported may have one, many, all, or no matrix fully approved, so it is not necessary to the process to have 100% approval in any matrix for the documentation to be approved. When recording a report in the web-form, the errors checked in the web forms affects the process' score and consequently the agency ranking. In this case when a documentation is approved but a matrix is not totally approved, the agency is able to publicize the process by making the necessary corrections on it according to the report, without the need of sending it back to CJU Office to be reanalyzed.

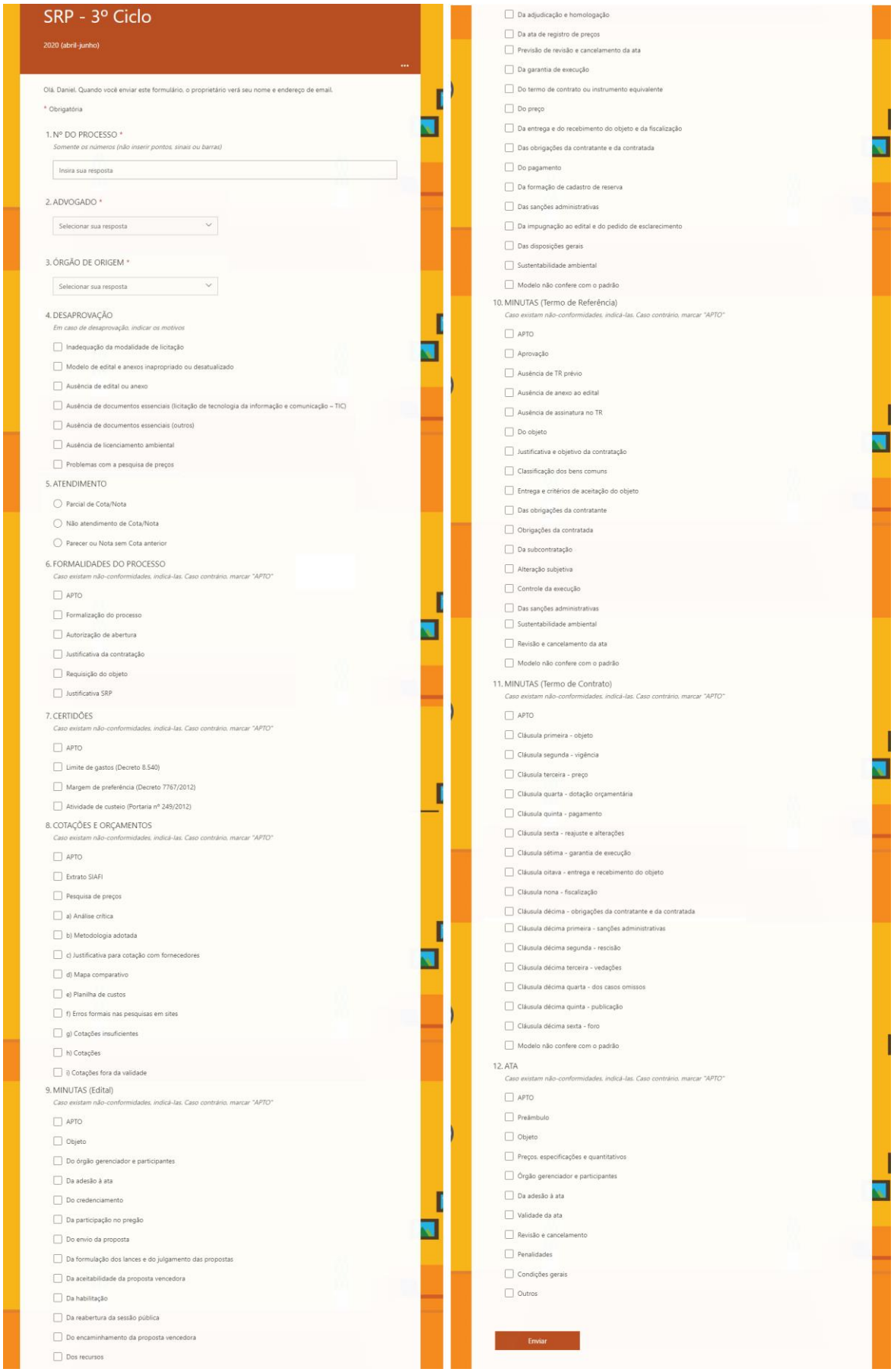


Figure 3-2 - SRP modality web form

3.2.3. Sharepoint list

All the information sent from the web forms is recorded in two Sharepoint lists, one for each bidding type, made available in each cycle. Every time one SRP form is sent, the data goes to an SRP Sharepoint list, and so it happens to PE. By making use of MS Power Automate tool, the web-form fields are sent straight to the list column related, where the lists headers are the main forms fields.

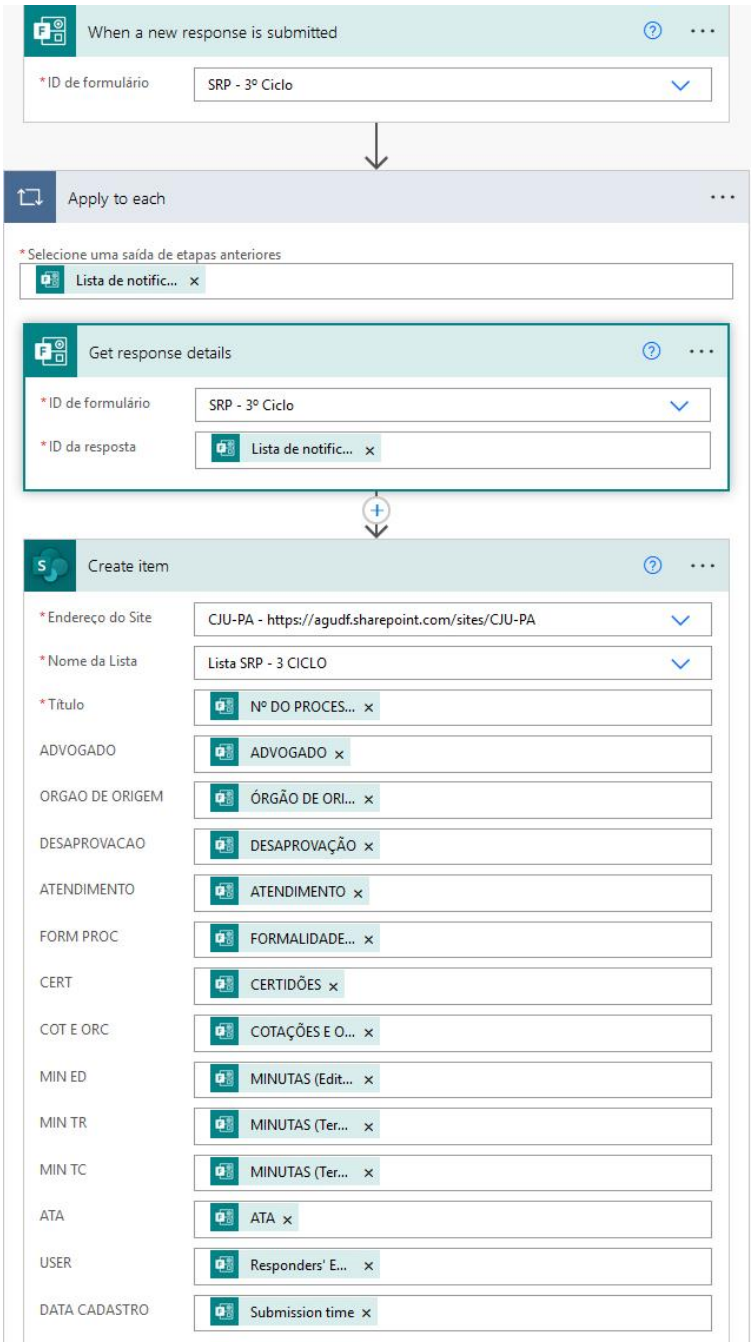


Figure 3-3 - MS Power Automate flow to create a new line in the SRP Sharepoint list

Inside Power Automate tool, the first instruction is to identify the web form where the data came from and get the responses, then designate the Sharepoint list that will receive it. In the “Create Item” box in the Figure 3-3 above, the fields on the left are the ones from the list, and the right ones are the web-form items related.

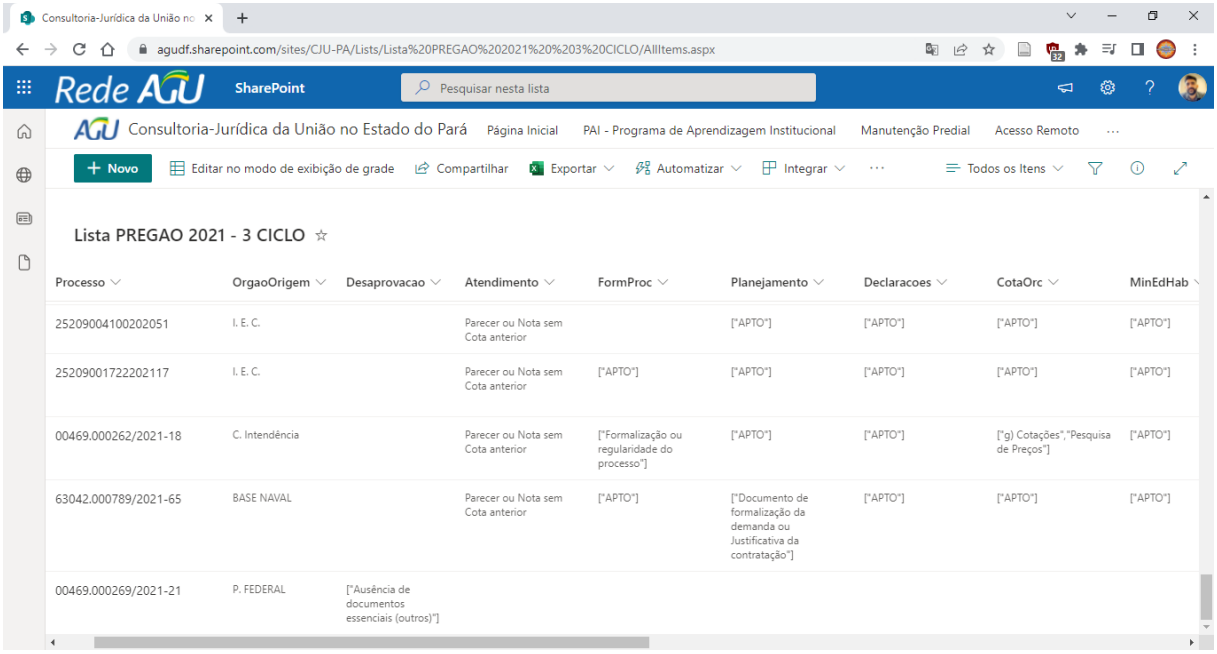


Figure 3-4 - The Sharepoint list for PE modality

The figure 3-4 above displays how the data is filed inside the lists, in this example it is the 2021's 3rd cycle for the Pregão Eletrônico [*Electronic Auction*] modality. As matrices have several items to be check listed, those items are semi-colon separated inside the related cell in the Sharepoint list. When the cycle ends and the list is done, it is exported in an Excel format that serves as source file to be extracted and used in SSIS, where the ETL processes runs to make the transformations and then load the data into the data warehouse.

3.3. DATA WAREHOUSE

The data warehouse is loaded at the end of ETL processes, after the transformation stage, and to finally be loaded, the data collected must match the data warehouse as designed. As the institution is regularly dealing with papers, there are no measures to be collected from the bid documentation analysis, and in this case, what is needed to improve is the quality of documentation from CJU supported agencies.

Although the bid modalities have different items to be analyzed, full detailed in the annexes 8.1 (SRP) and 8.2 (PE), one single table Dim_Process is created to store information from both with a column named 'Modality' to distinguish them. The field related to the extra matrix (M6A) in the SRP modality is ignored when dealing with the PE dashboard. The data warehouse is incrementally loaded in the end of each cycle, meaning that there are five loads within a year or macrocycle.

As there are no numerical fields in legal documents to be measured, the fact is not transactional. The schema used in both is a star schema where the dimensions connect to a central factless fact table. The dimensions designed are Process, Consultant, Agency, and Date, with the Fact table central in a star schema model as proposed by Ralph Kimball.

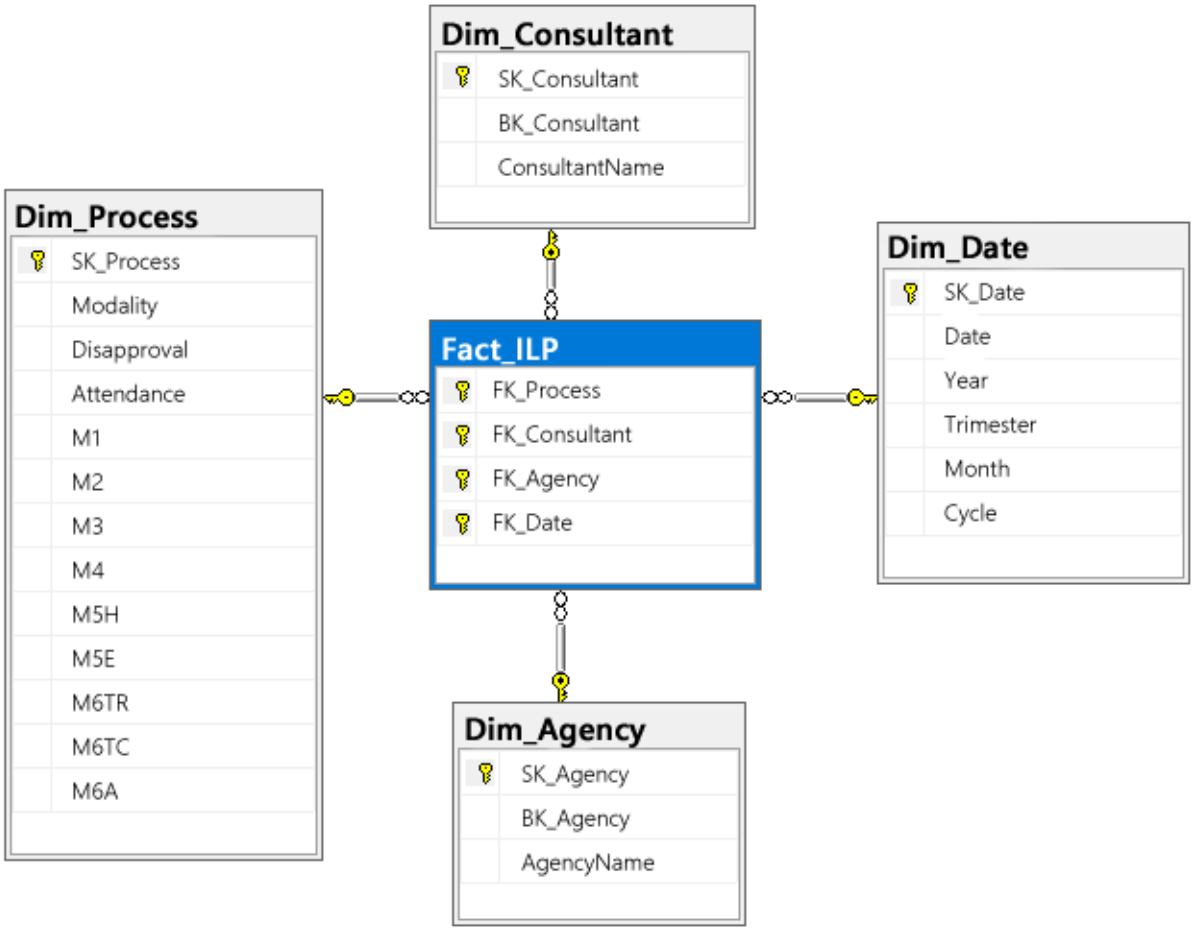


Figure 3-5 - Data warehouse diagram

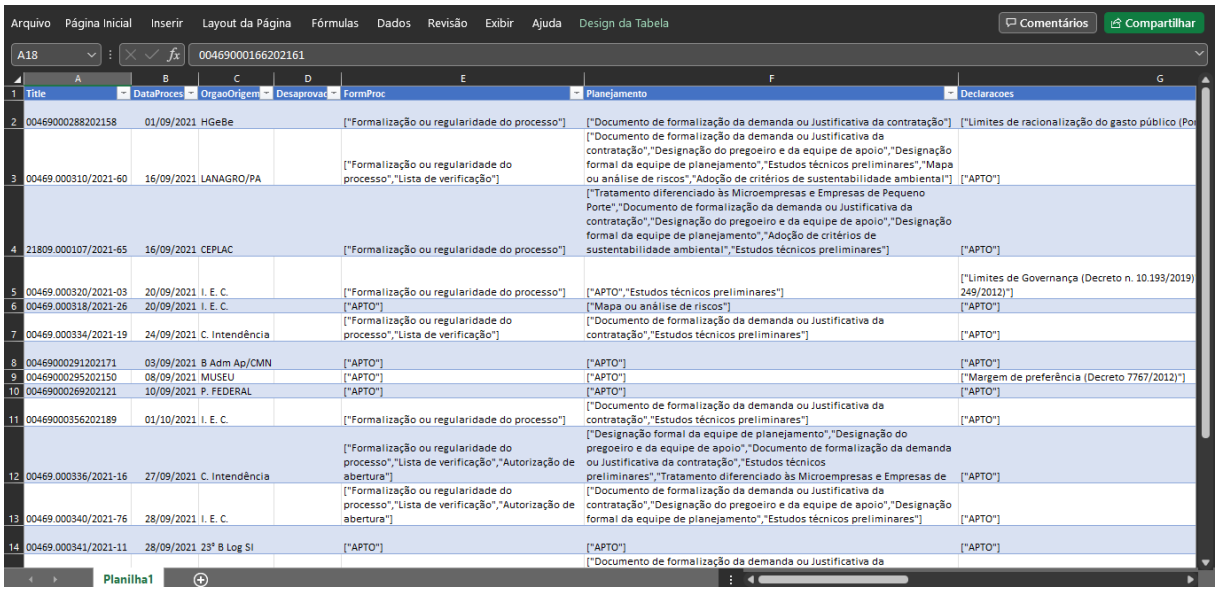
3.4. ETL PROCESSES

The classic ETL processes used, refers to three steps that are Extraction, Transform and Load. The final goal is to load the database, but first there is a need to treat the data after its extraction from the source. The ETL processes runs in the SSIS (SQL Server Integration Services), a platform that creates high-performance data integration solutions inside the MS Visual Studio.

For both modalities, when one cycle finishes, the Sharepoint list is exported serving as source file in a MS Excel format compatible with the SSIS tool. The transformation is the most detailed stage as it takes more time to be accomplished, and it occurs to make the changes needed to match the target data warehouse design, that is finally loaded in the end of this process.

3.4.1. Loading the Staging Area

Inside the Sharepoint list, there is a simple option to export the data in a MS Excel format. Whenever a cycle finishes, the lists of both modalities are exported, and then extracted in Visual Studio SSIS. The figure 3-6 below shows part of a PE [*Electronic Auction*] extracted file, where it is possible to visualize that there are some problems with the initial data, but the file is ready to be handled and so the problems to be fixed.



	A	B	C	D	E	F	G
1	Title	DataProces	OrgaoOrigem	Desaprovac	FormProc	Planejamento	Declaracoes
2	00469000288202158	01/09/2021	HGeBe		["Formalização ou regularidade do processo"]	["Documento de formalização da demanda ou Justificativa da contratação"]	["Limites de racionalização do gasto público (Po
3	00469.000310/2021-60	16/09/2021	LANAGRO/PA		["Formalização ou regularidade do processo","Lista de verificação"]	["Documento de formalização da demanda ou Justificativa da contratação","Designação do pregoeiro e da equipe de apoio","Designação formal da equipe de planejamento","Estudos técnicos preliminares","Mapa ou análise de riscos","Adoção de critérios de sustentabilidade ambiental"]	["APTO"]
4	21809.000107/2021-65	16/09/2021	CEPLAC		["Formalização ou regularidade do processo"]	["Tratamento diferenciado às Microempresas e Empresas de Pequeno Porte","Documento de formalização da demanda ou Justificativa da contratação","Designação do pregoeiro e da equipe de apoio","Designação formal da equipe de planejamento","Adoção de critérios de sustentabilidade ambiental","Estudos técnicos preliminares"]	["APTO"]
5	00469.000320/2021-03	20/09/2021	I. E. C.		["Formalização ou regularidade do processo"]	["APTO","Estudos técnicos preliminares"]	["Limites de Governança (Decreto n. 10.193/2019) 249/2012)"]
6	00469.000318/2021-26	20/09/2021	I. E. C.		["APTO"]	["Mapa ou análise de riscos"]	["APTO"]
7	00469.000334/2021-19	24/09/2021	C. Intendência		["Formalização ou regularidade do processo","Lista de verificação"]	["Documento de formalização da demanda ou Justificativa da contratação","Estudos técnicos preliminares"]	["APTO"]
8	00469000291202171	03/09/2021	B Adm Ap/CMN		["APTO"]	["APTO"]	["APTO"]
9	00469000295202150	08/09/2021	MUSEU		["APTO"]	["APTO"]	["Margem de preferência (Decreto 7767/2012)"]
10	00469000269202121	10/09/2021	P. FEDERAL		["APTO"]	["APTO"]	["APTO"]
11	00469000356202189	01/10/2021	I. E. C.		["Formalização ou regularidade do processo"]	["Documento de formalização da demanda ou Justificativa da contratação","Estudos técnicos preliminares"]	["APTO"]
12	00469.000336/2021-16	27/09/2021	C. Intendência		["Formalização ou regularidade do processo","Lista de verificação","Autorização de abertura"]	["Designação formal da equipe de planejamento","Designação do pregoeiro e da equipe de apoio","Documento de formalização da demanda ou Justificativa da contratação","Estudos técnicos preliminares","Tratamento diferenciado às Microempresas e Empresas de	["APTO"]
13	00469.000340/2021-76	28/09/2021	I. E. C.		["Formalização ou regularidade do processo","Lista de verificação","Autorização de abertura"]	["Documento de formalização da demanda ou Justificativa da contratação","Designação do pregoeiro e da equipe de apoio","Designação formal da equipe de planejamento","Estudos técnicos preliminares"]	["APTO"]
14	00469.000341/2021-11	28/09/2021	23ª B Log SI		["APTO"]	["APTO"]	["APTO"]
						["Documento de formalização da demanda ou Justificativa da	

Figure 3-6 - Sharepoint list exported in an Excel format

The project created in the SSIS contains two packages running, one for the Staging Area and the other to finally load the Data Warehouse. Although the quantity of analyzed items is different among the modalities, the data from both source files receive the same treatment, except for the matrix M6A that only exists in the SRP modality. For this reason, the images exposed next were taken randomly between the two modalities.

Due to the absence of a previous design as the source of this project are lists, the Staging Area had a different dimensional model from the DW design. In the Staging Are there is no fact table but a central table with the data from the Process connecting to the other three dimensions that are Date, Agency, and Consultant. The normalization occurs when loading the data warehouse, where a central factless fact table is created linking to the dimensions as previously illustrated in the Figure 3-5.

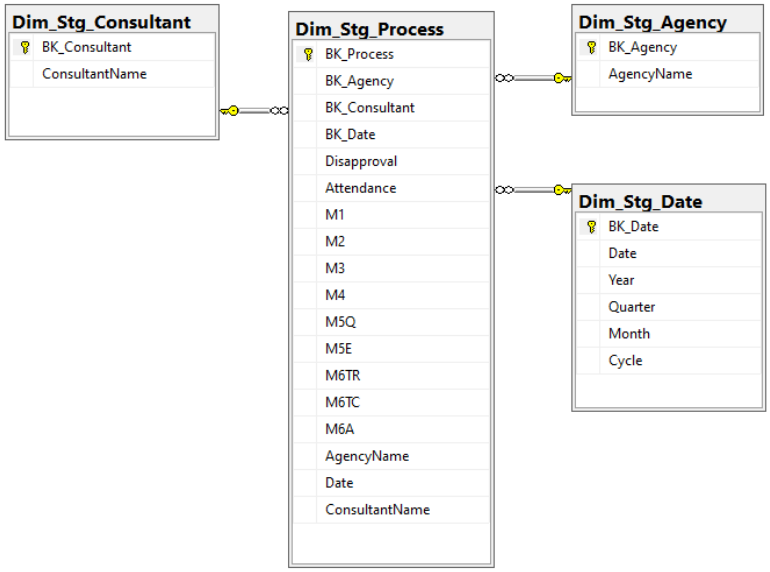


Figure 3-7 - The Staging Area diagram

Although the Staging Area diagram is not the same designed for the data warehouse, the SA database not only collects the data from the source, but it runs the most significant transformation to accomplish the final format and design required for loading the DW.

- **Dim_Stg_Consultant and Dim_Stg_Agency**

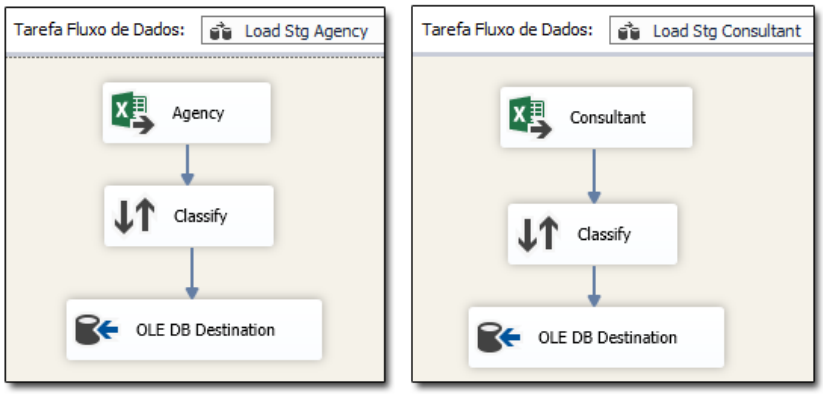


Figure 3-8 - Staging Area loading Dim_Stg_Consultant and Dim_Stg_Agency

The dimensions Stg_Consultant and Stg_Agency are fully loaded from flat files apart without suffering any relevant treatment, except the Classify tool ConsultantName and AgencyName, to order it alphabetically and to remove its duplicated values, if any. In the SSMS database these DIMs receive an automatic Business Key specified as identity while loading the Name fields.

- **Dim_Stg_Date**

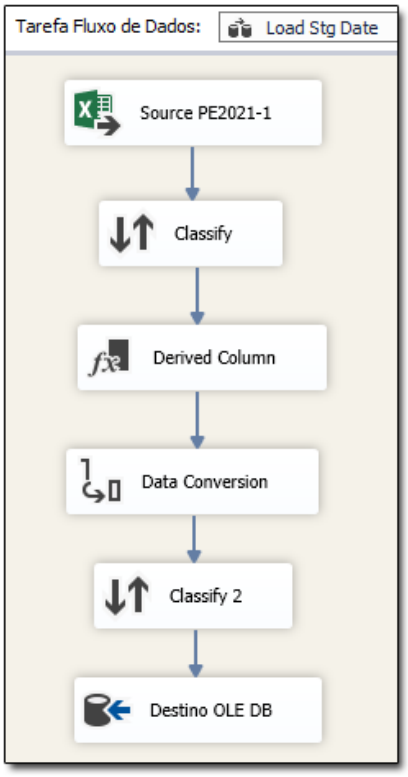


Figure 3-9 - Staging Area loading Dim_Stg_Date

Together with those dimensions, the Dim_Stg_Date is loaded in the first container, where the first Classify gets the field DataProcesso [*Process Date*] that is set as ascending, and to remove the duplicates. The Derived Column has the most significant transformations as shown in the image below:

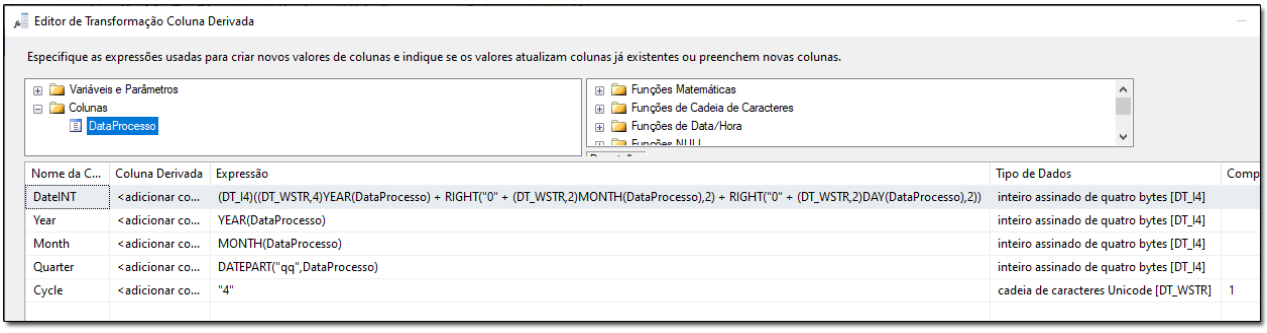


Figure 3-10 - Dim_Stg_Date derived column

- DateINT concatenates the original date as an integer that becomes the BK
- Year, Month and Quarter columns gets these data from the original date
- Cycle column receive the value of the cycle related, in this case “4”.

The next two steps are the Data Conversion that modifies the data types to match the database, and the Classify 2 where the final or new columns are chosen, and again remove the duplicates if any. Later, the Dim_Stg_Date is finally loaded.

- **Dim_Stg_Process**

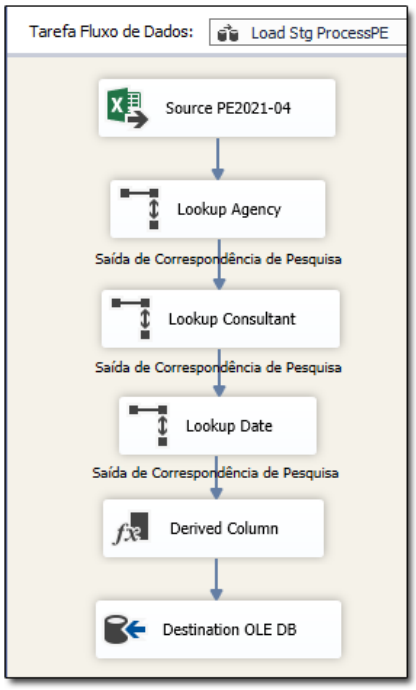


Figure 3-11 - Staging Area loading Dim_Stg_Process

In the Staging Area this table is the one with the most significant transformations. It is loaded at last, on a third sequence container. The Lookup tool gets the Business Keys from the previously loaded dimensions in the Staging Area. It runs fine for the Agency and Consultant dimensions, but for the Date there was an error message that the data types did not match between the source and the destination table. The original date in the column DataProcesso from the extracted file is set as DT_DATE format, and in the Dim_Stg_Process it is DT_DBDATE. To solve it, the data type in the source file properties was changed to DT_DBDATE.

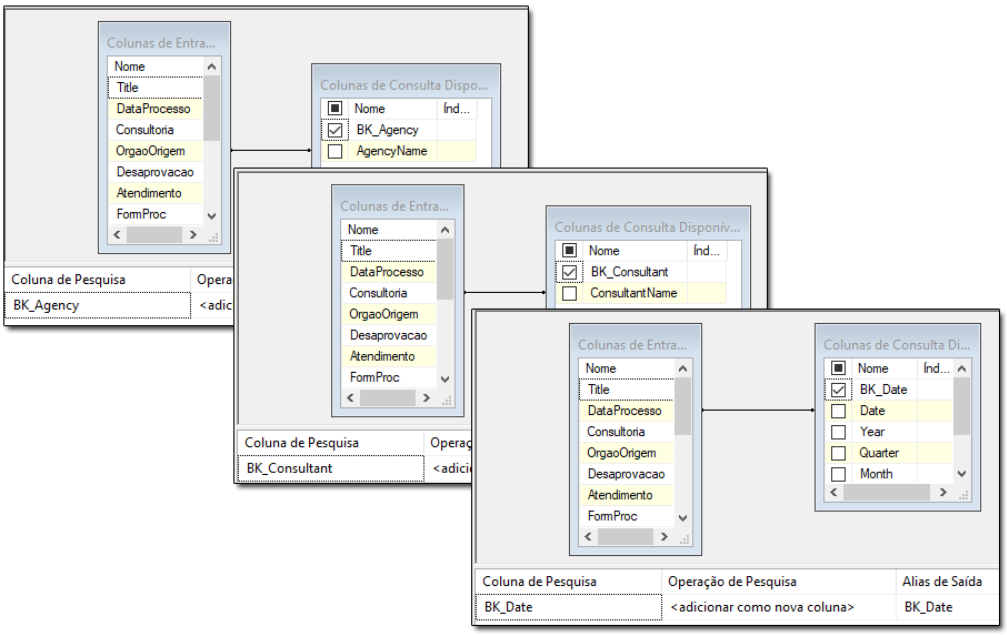


Figure 3-12 - Dim_Stg_Process Look Up tool columns

The next step is to make the transformation in the columns with the information that will populate the matrices. By looking at the source file it is possible to realize that the data need cleansing and transformation to deal with dots, slashes, quotes, dashes, brackets, commas, and others similar characters, as shown in the Figure 3-13 below:

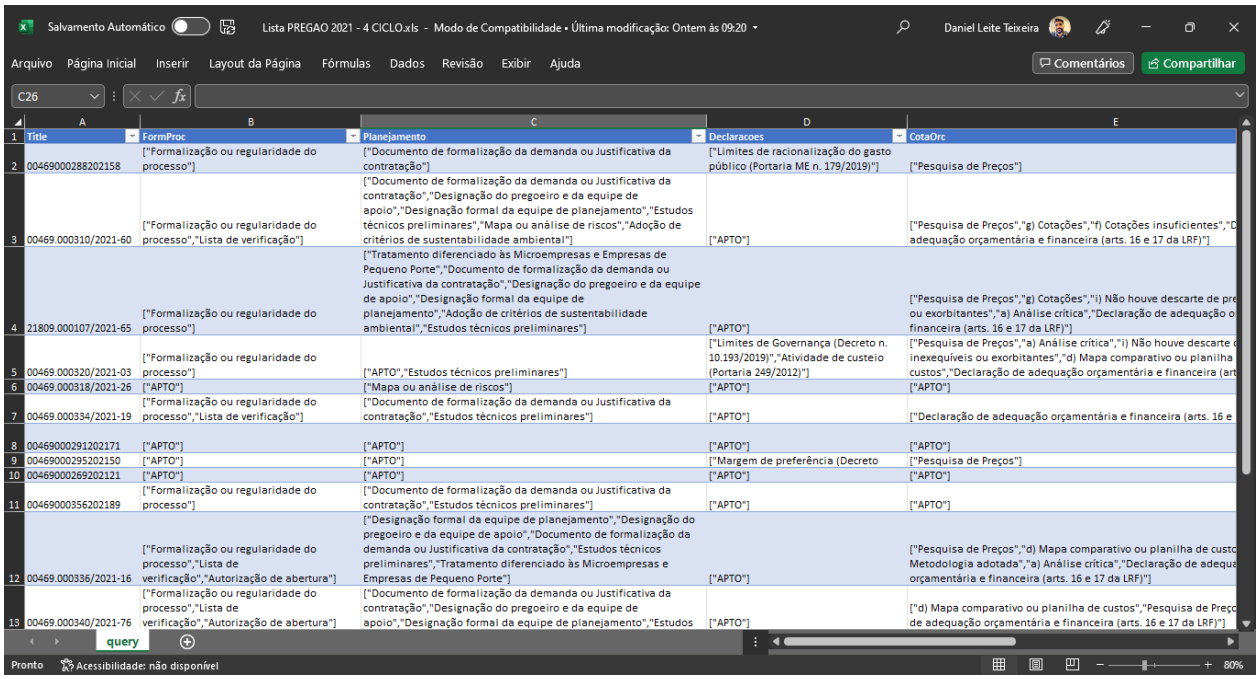


Figure 3-13 - Data in the source file

The transformations take place in the Derived Column tool, specially making use of the REPLACE function to clean and standardize the data from the source file. Despite having texts, some columns have different types in the extracted file. While applying the REPLACE function, there was an error message alerting that this function does not support DT_NTEXT or DT_TEXT types. For the PE

modality, this error was shown in the Planejamento [*Planning*] column (that will be the matrix M2), and for the SRP modality the same alert was displayed for Planejamento [*Planning*], Minutas do Edital Habilitação [*Public notice qualification drafts*], and Ato [*Act*] (matrices M2, M5Q and M6A). Similar to the work done in the Date field within the Dim_Stg_Date, the data type in these columns was changed from DT_NTEXT or DT_TEXT to DT_WSTR in the source file properties.

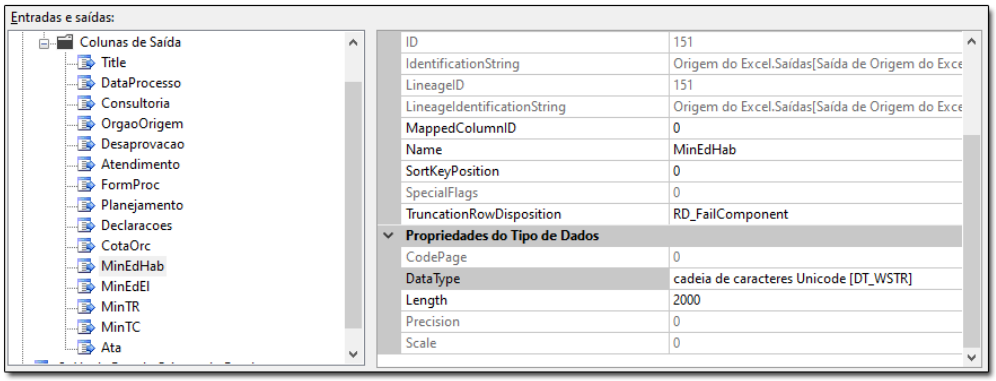


Figure 3-14 - The changed column DataType in the source file properties

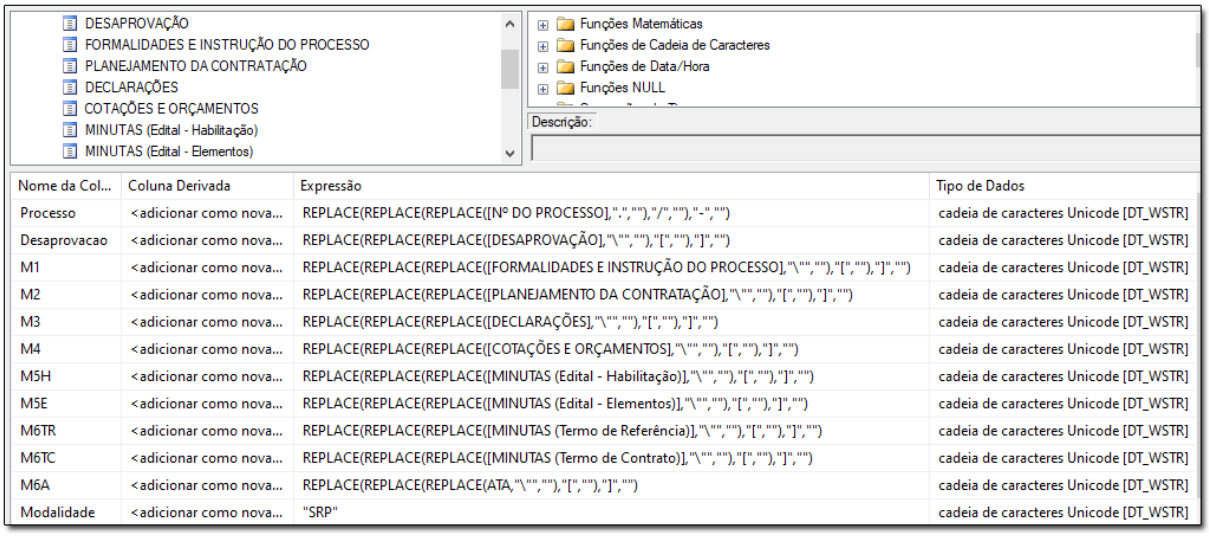


Figure 3-15 - Derived Column tool transformation

The REPLACE function is used to remove or substitute some words or characters. In this case, it is used to do several changes at once, removing unwanted characters and creating new derived columns, that in the last stage are mapped to the related column/matrix in the destination table Dim_Stg_Process, finally loading all the tables and finishing the ETL work in the Staging Area.

3.4.2. Loading the Data Warehouse

The second package created in the SSIS is to the data warehouse. At this time the source is the Staging Area already populated, with no need of use the extracted Excel file. As most of the ETL process was done in the SA, the tables in the DW basically need to map the columns and load it.

- **Loading the Dimensions**

The dimensions are loaded in the first sequence container. Dim_Agency and Dim_Consultant does not receive any significant transformation as these tables were previously treated in the Staging Area. The work done is only to map the source table in the SA to the destination column in the related table in the data warehouse, where the surrogate keys are automatically generated specified as identity in the SSMS corresponding column.

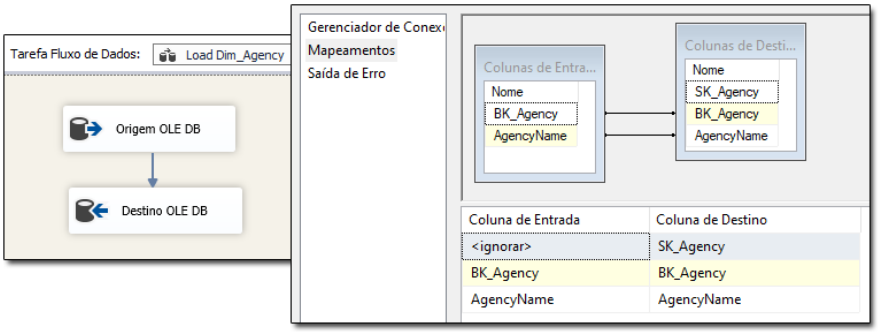


Figure 3-16 - Loading the Dim_Agency

The most relevant fact while loading the DW dimensions is that the BK_Date and the BK_Process turns to be the SK_Date and the SK_Process in the Dim_Date and the Dim_Process. The date field is in an integer format like '20210903', but still the original date is maintained in the Date field. The process has a unique national identifier generated when the documentation is sent by an Agency to the CJU Office, this number is used as the SK_Process.

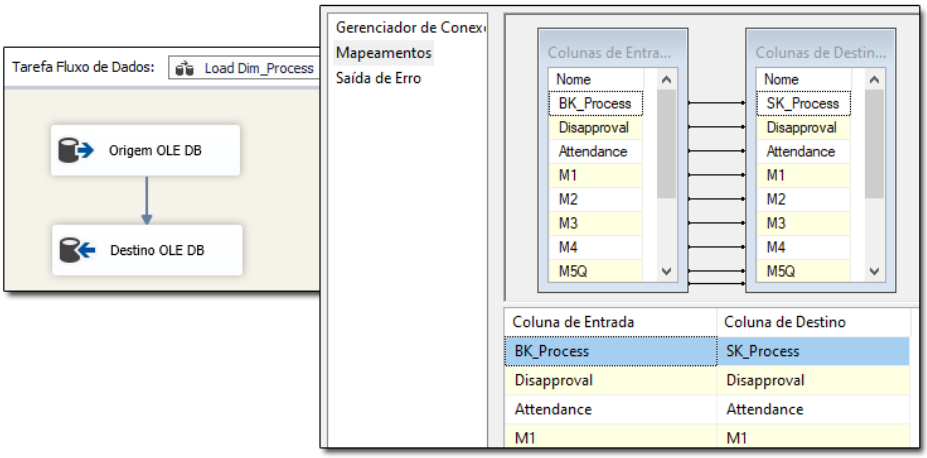


Figure 3-17 - Loading the Dim_Process

- **Loading the Fact**

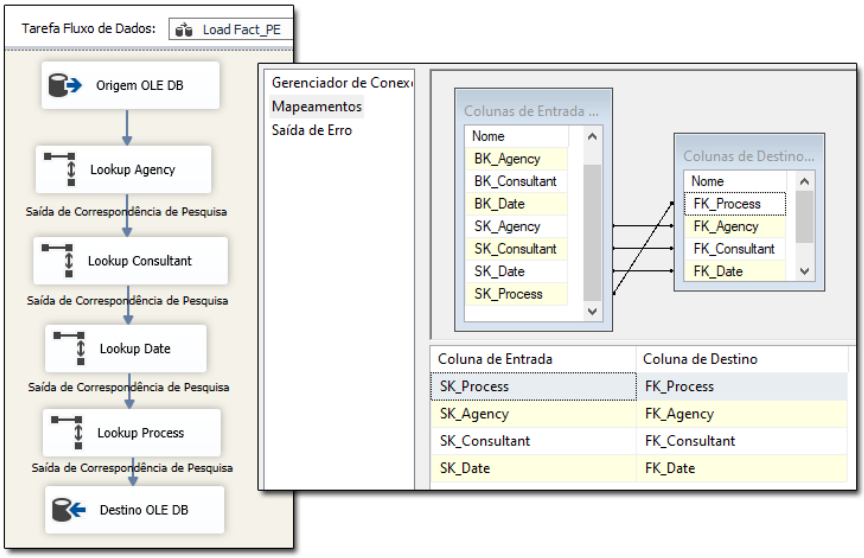
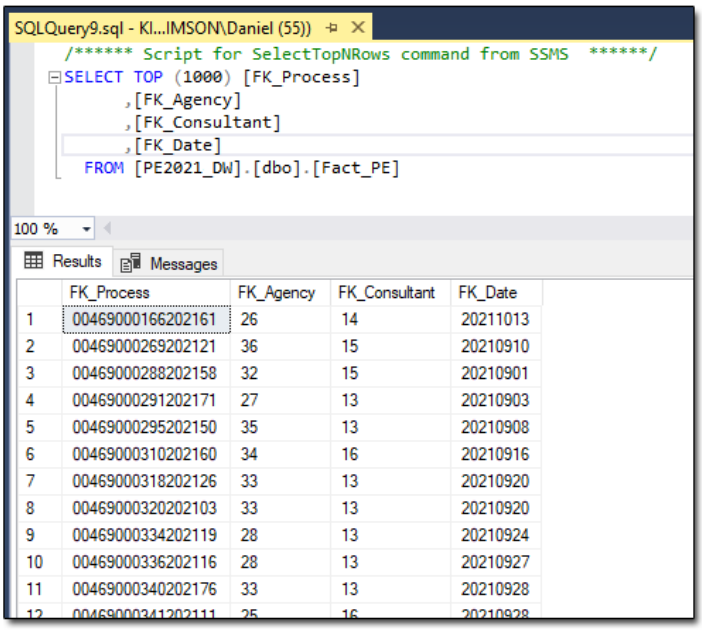


Figure 3-18 - Loading the Fact table

The central Fact table is the last one to be loaded. It is linked to the dimensions by connecting its foreign keys to their surrogate keys, previously joined using the Lookup tool in each dimension. After the fact table is fully loaded, the data warehouse is finally populated.



	FK_Process	FK_Agency	FK_Consultant	FK_Date
1	00469000166202161	26	14	20211013
2	00469000269202121	36	15	20210910
3	00469000288202158	32	15	20210901
4	00469000291202171	27	13	20210903
5	00469000295202150	35	13	20210908
6	00469000310202160	34	16	20210916
7	00469000318202126	33	13	20210920
8	00469000320202103	33	13	20210920
9	00469000334202119	28	13	20210924
10	00469000336202116	28	13	20210927
11	00469000340202176	33	13	20210928
12	00469000341202111	25	16	20210928

Figure 3-19 - Loaded PE DW Fact table

3.5. THE PROCESS' SCORE

Once the data is loaded inside the DW, Power BI connects to the data base and by making use of an algorithm it is possible to obtain each process' score and finally generate the agencies ranking for the cycle. At the end of the year when the macrocycle finishes, the top ranked agencies in the period are prized by the heads of CJU based on their good practices and results on the Institutional Learning Program, while conducting legal bidding processes.

3.5.1. Matrices' Scores

The items related in the web form to each bidding type vary in quantities as detailed in the annexes 8.1 and 8.2. The SRP modality has more items to be analyzed than PE, but there is no difference in the maximum score obtained for a same named matrix because its score is turned to percentual. The scoring process is done for each process by making several transformations in Power Query, a tool that runs inside Power BI. The maximum value of achievements that can be obtained for each matrix is detailed in the tables below, displayed by modality.

SRP Matrix	Maximum Score
M1 – Formalities	9
M2 – Planning	9
M3 – Statements	5
M4 – Quotation	12
M5 Q – Qualification	3
M5 E – Elements	25
M6 TR – Term of reference	20
M6 TC – Term of contract	18
M6 A – Acts	10

Table 3-3 - SRP maximum matrices values

PE Matrix	Maximum Score
M1 – Formalities	5
M2 – Planning	8
M3 – Statements	5
M4 – Quotation	12
M5 Q – Qualification	3
M5 E – Elements	22
M6 TR – Term of reference	20
M6 TC – Term of contract	17

Table 3-4 - PE maximum matrices values

When a matrix gets the “Able” option checked in the web form, it automatically receives the maximum score. On the other hand, one single item checked in the Disapproval main field sets the whole bidding process’ score to zero. Each matrix value is calculated at a time, separately, and the matrixes comes from duplications of the Dim_Process in Power Query.

The matrix’ score is obtained by splitting the columns in lines by the separator (comma), using group by SK_Process and counting distinct lines, then merging with Dim_Process to collect the respective matrix and the disapproval columns. A conditional column is generated counting the quantity of items checked, and its value is subtracted from the maximum score available for the matrix. The obtained value is set to absolute to remove the negative numbers and then turned to percentual in a new column. For the matrices M5 and M6, that are subdivided, the final value is the average obtained from the subdivision.

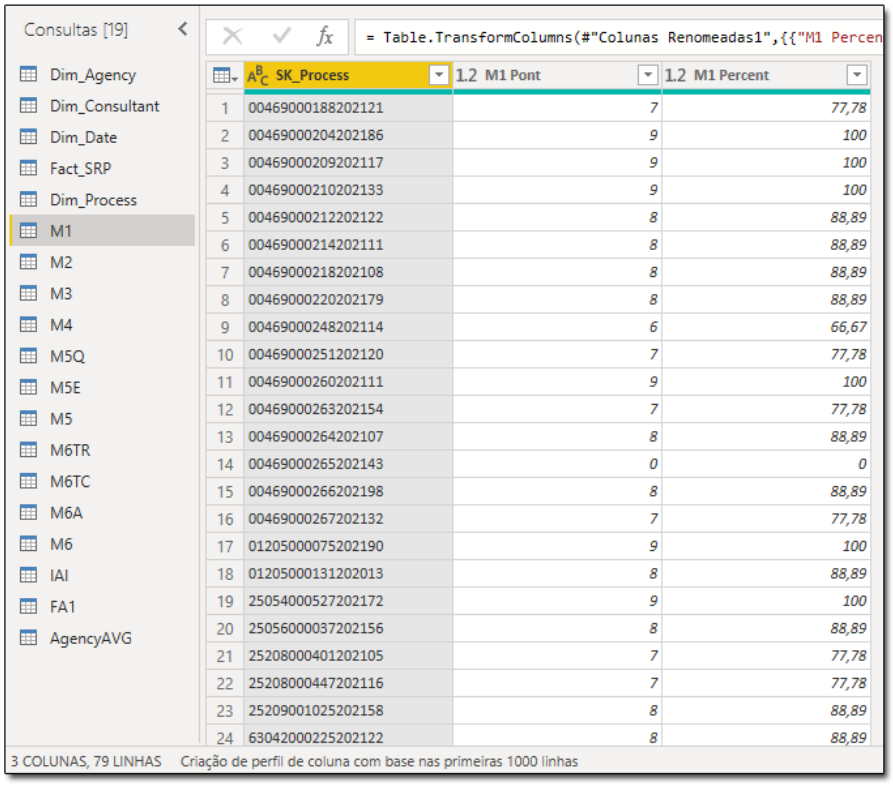


Figure 3-20 - The processes' score in the M1 matrix

From Power Query, the Figure 3-20 above displays the scores of each process obtained in the matrix M1 after the calculation explained. The left column SK_Process identifies the legal process number, the “M1 Pont” shows the quantity of achievements reached per process, and the column “M1 Percent” displays these achievements in percentual for this matrix (M1). Each matrix is treated at a time in PowerQuery, and later their values are merged to get the Agency average and the “IAI” value obtained from the algorithm, explained as follows.

3.5.2. Algorithm

As previously explained, each process' score is obtained individually for each documentation analyzed, and the agencies ranking can only be generated when the cycle ends because the algorithm uses some other factors in addition to the value resultant from the matrices. After each process' score obtention, the algorithm named "Índice de Aprendizagem Institucional (IAI)" [*Institutional Learning Index (ILI)*], runs to generate the agencies ranking in the cycle. The ILI formula below is the same for both bidding modalities (SRP and PE):

ILI = [(M1 + M2 + M3 + M4 + M5 + M6)] – [(RF1 – RF2 – RF3) + LF1 + LF2 + JF]

The value in the first brackets is directly obtained from the SUM of the matrices' scores. For the second brackets, CJU consultants defined some aspects considered most relevant on the bidding documentation analysis, that shall increase or decrease the process' final score in case of reaching or not a minimum value in some matrices. These elements in the algorithm are called Reduction Factors (RF), Learning Factors (LF) and Judicialization Factor (JF). These values are obtained in comparison with the scores resulting from some matrices as explained next:

- RF1 (Reduction Factor 1): if Matrix 5 < 80%, then RF1 receives -5. Else, RF1 receives 0.
- RF2: if M6 < 80%, then RF2 = -2. Else, RF2 = 0.
- RF3: if M2 < 80%, then RF3 = -3. Else, RF3 = 0.
- JF: if M5Q = 100%, then JF = 15. Else, JF =0.
- LF1 (Learning Factor 1): Counts in how many Matrices a given agency has improved in comparison with the previous cycle. If the institution has improved from 1 to 2 matrices, LF1 = 5; if it has improved from 3 to 4 matrices, LF1 = 10; if the agency has improved from 5 to 6 matrices, LF1 = 20. Else, LF1 = 0.
- LF2: if IAI < 90%, then LF2 = 15. Else, LF2 = 0.

	SK_Process	M Score	RF 1	RF 2	RF 3	LF 1	LF 2	FJ	IAI
1	00469000188202121	82,74	0	0	-3	5	0	15	99,74
2	00469000248202114	87,26	0	0	0	5	0	15	107,26
3	00469000204202186	95,99	0	0	0	0	15	15	125,99
4	00469000209202117	91,2	0	0	0	0	15	15	121,2
5	00469000210202133	100	0	0	0	0	15	15	130
6	64046001282202151	94,81	0	0	0	0	15	15	124,81
7	00469000212202122	81,64	0	0	-3	0	0	15	93,64
8	00469000214202111	90,76	-5	0	0	10	15	0	110,76
9	00469000218202108	88,07	-5	0	0	0	0	0	83,07
10	64046002942202111	88,07	-5	0	0	0	0	0	83,07
11	00469000220202179	89,3	0	0	-3	0	0	15	101,3
12	00469000251202120	88,06	0	0	0	0	0	15	103,06
13	00469000260202111	90,13	0	0	-3	0	15	15	117,13
14	00469000263202154	90,59	0	0	-3	0	15	15	117,59
15	00469000264202107	96,98	0	0	0	0	15	15	126,98
16	00469000265202143	0	-5	-2	-3	0	0	0	0
17	00469000266202198	88,5	0	0	-3	0	0	15	100,5
18	00469000267202132	92,06	0	0	0	0	15	15	122,06
19	01205000075202190	85,03	-5	0	-3	0	0	0	77,03
20	01205000131202013	91,4	-5	0	0	0	15	0	101,4
21	25054000527202172	98,3	0	0	0	20	15	15	148,3
22	25056000037202156	88,78	-5	0	-3	5	0	0	85,78

Figure 3-21 - Algorithm results

Later gathering the values from the matrices, the algorithm runs to obtain the ILI final value for each legal bid process in IAI query. In the Figure 3-21 above, this value is shown in the right column with the header “IAI”, translated as ILI in this project. The values in the column “M Score” comes from the matrices average (within the first brackets in the algorithm formula), then this value is summed with the ones in the remaining columns (in the second brackets from the formula) that are the Learning Factors FA1, FA2 [*LF1, LF2*], the Reduction Factors FR1, FR2 and FR3 [*RF1, RF2 and RF3*], and the Judicialization Factor FJ [*JF*], finally obtaining IAI [*ILI*] value.

The maximum IAI [*ILI*] obtainable score by a single process is 150 points, of which 100 points are from the matrices (that are percentual), and the remaining 50 points come from the mentioned Learning, Reduction and Judicialization Factors. However, in the first cycle of the year the maximum score that may be obtained is 130, as there is no previous cycle to compare the values and generate the LF1 that is worth 20 points.

3.5.3. Agencies’ Ranking

	AgencyName	AgencyAVG
1	HGeBe	103,5
2	H. N. Belém	125,99
3	C. Intendência	103,65
4	8° D SUP	110,76
5	8° BEC	105,24
6	H Gu Ma	92,18
7	B Adm Ap/CMN	108,54
8	MUSEU	89,22
9	Dsei/Tapajós	148,3
10	Dsei/Guatoc	85,78
11	CENP	83,68
12	I. E. C.	98,35
13	BASE NAVAL	129,6
14	2° BIS	55,26
15	51° BIS	131,29
16	53° BIS	140,08
17	23ª Bda Inf Sl	143,08
18	Pq R Mnt/8ª RM	94,33
19	23° Esqd C Sl	146,61
20	CMBel	125,38
21	23° B Log Sl	125,09
22	BABE	127,77

Figure 3-22 - Agencies’ average

The last stage in Power Query is to generate the agencies’ ranking. The IAI query is duplicated and merged with the Dim_Process to get the AgencyName column, then use the function group by “AgencyName” and set the operation to calculate the average value in the IAI column, generating the AgencyAVG column. After collecting this value, the whole work in Power Query is finished and the data is ready to develop the charts in the dashboards.

3.6. DASHBOARDS

CJU needs to display detailed information of all processes and agencies to discuss the results with the supported agencies whenever a cycle end. The biggest challenge when designing it is that the matrices have lots of items to be displayed, such as M3 that has more than almost thirty items as shown in the 8.1 and 8.2 annexes.

Considering the need of displaying the whole information of each matrix, what would be at least eight charts only for the matrices containing several items each, the dashboards would be too dense and difficult to understand, the solution was to split the matrices among several pages as proposed by Kosara & Mackinlay (2013) stating this type of dashboard as slideshow.

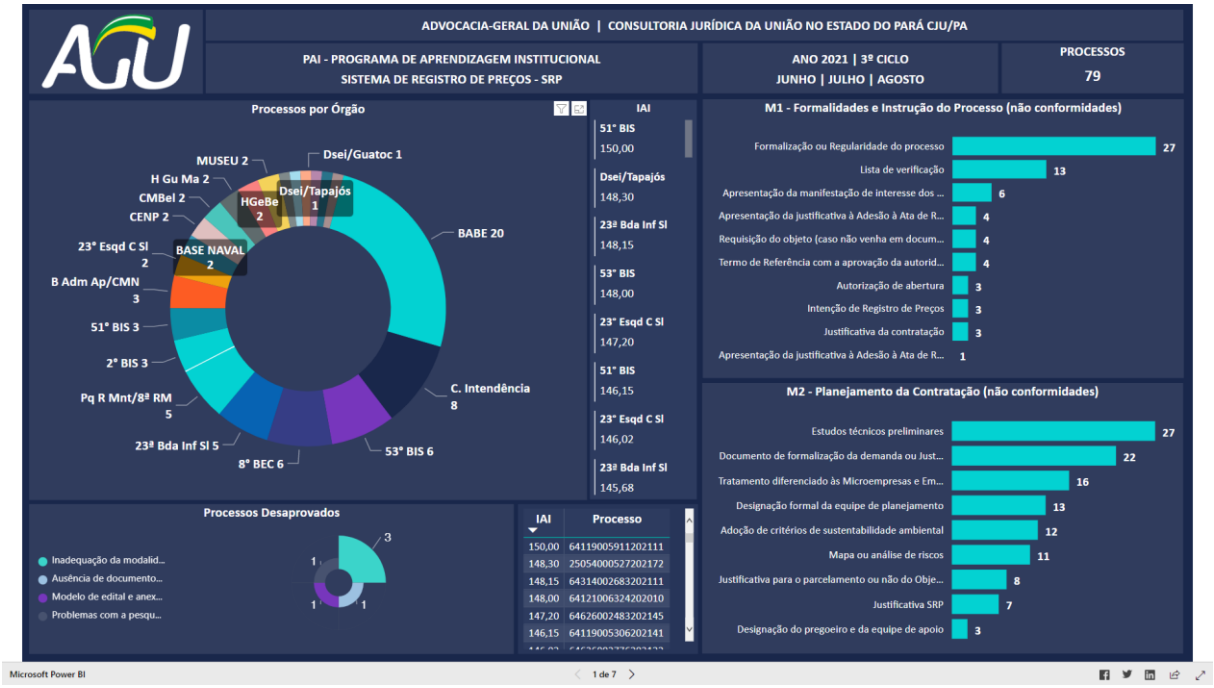


Figure 3-23 - 2021's 3rd cycle dashboard for SRP modality

The figure 3-23 displays the first page of a published dashboard where the right side brings horizontal bar charts with the matrices M1 and M2 errors in the period, and the left side brings general information of disapproved processes in the lower left, the processes' score in cards and table in the middle, and the main donut chart showing the agencies and its processes in the cycle.

From the 1st to the 5th page, the dashboards look very similar. The only change among these pages is the right side that shows the matrices. All the dashboard pages are disposed in the annex 8.3.

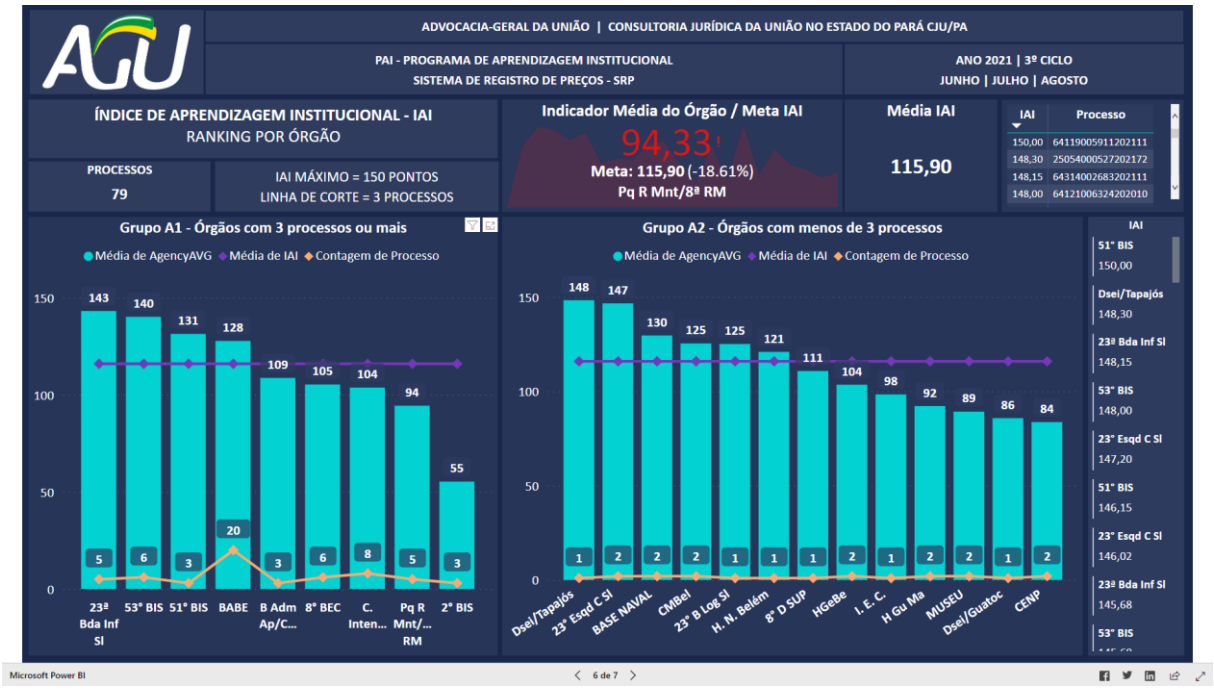


Figure 3-24 - 2021's 3rd cycle ranking page for SRP modality

The Figure 3-24 is the 6th dashboard page which contains the ranking for this cycle. A threshold shown as "Linha de corte" is defined to split the number of agencies in two different groups based on their processes volume in the period. The threshold value is obtained from the quantity of processes divided by the number of agencies in the cycle, in this case the value is 3.

Based on this value, the two main bar charts are designed using this threshold as filter where the A1 group displays the agencies that had a number of processes equal or higher than the threshold, and the A2 group shows the agencies with a quantity of processes lower than this value. For both charts, the green bars are each agency average in comparison with the purple line that shows the ILI overall average in the cycle, and the orange line displays the processes amount of each agency.

Figure 3-25 - KPI and IAI average

As the quantity of processes is uncertain because it depends on the supported agencies need to bid, it is possible that an institution that appears in one cycle, may not appear in another. For the same reason the KPI cannot be pre-defined, and it is stated by the ILI average in the cycle. The image above was taken from the previous Figure 3-24 to highlight two relevant boxes. The right one "Média IAI" box shows the IAI [*ILI*] average in the cycle, which is the value that becomes the KPI for this period. The left box "Indicador Média do Órgão / Meta IAI" [*Agencies Average Indicator / ILI Target*] is the KPI that displays the agency average in the bigger number (in red), the white number below compares the value with the indicator/target and shows the difference inside the parentheses, and the text below brings the agency name.

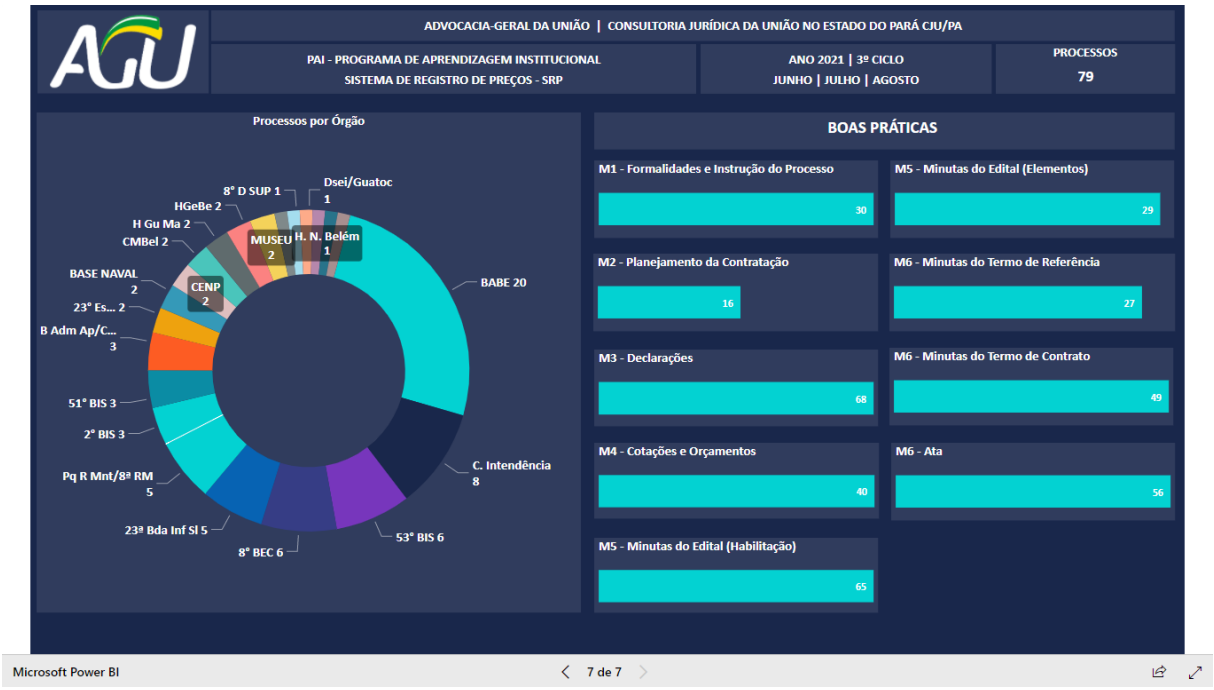


Figure 3-26 - Best practices page

The figure 3-26 displays the 7th and last dashboard page, dedicated to show the best practices where there are several horizontal bar charts showing the achievements in each matrix, and the left donut chart with the data from the agencies and their processes amount, similar to the first page. The bars charts bring only the amount of the “Able” option checked in each matrix, representing processes with full achievement on it.

3.7. Further Remarks

3.7.1. Incremental Load

The created data warehouse receives the first load at the end of the first cycle. The incremental load is done at the end of each remaining cycle without the need of loading the dimensions Consultant and Agency, that were previously loaded only in the first cycle. It is possible to make the incremental load in two ways:

- Creating new incremental packages in SSIS where:
    - In the Staging Area: deleting and loading only the Date and Process dimensions
    - In the Data Warehouse: loading Date and Process first, and then the Fact table
- Using the same packages from the first cycle with the following adjustments:
    - In the Staging Area: disabling the delete of Date and Process dimensions, and the load of Consultant and Agency dimensions

- In the Data Warehouse: disabling the load of Consultant and Agency dimensions and loading only the remaining dimensions Date and Process, and then the Fact at last

The SSIS images containing the Staging Area and the Data Warehouse ETL processes are displayed in the annex 8.4.

3.7.2. Filtering a cycle in Power Query

CJU needs to publish dashboards per cycle to display and discuss the results with the supported agencies, identifying the best ranked and inviting them to schedule a further session to spread their best practices. As previously stated, the data is incrementally loaded whenever a cycle ends. For instance, when working on the third cycle, the previous cycles are already in the database. In this case, in the beginning of the work in Power Query, it is necessary to filter the 'Cycle' column in the Dim_Process and then start the processes' scoring. This same procedure is done to choose the bidding modality in Power Query.

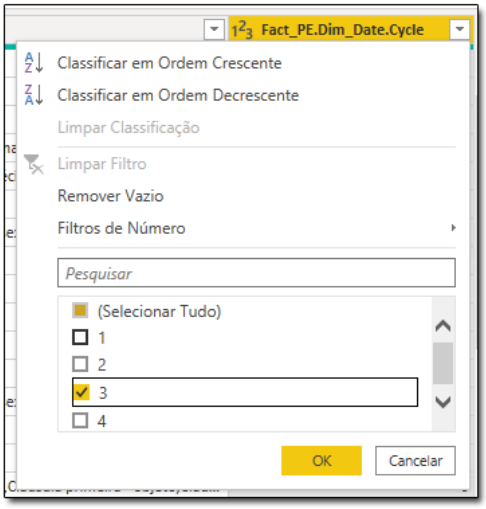


Figure 3-27 - Filtering a cycle in Power Query

4. RESULTS

The prevention of errors in bidding documents is to avoid the processes going to the court. For this reason, the project applied the concept of institutional learning in the public governance. Institutionally, this project inception was to identify errors patterns in public tenders and to offer to stakeholders statistical and strategical information about the need of improvement in the bid documentation. The regular system use turned it into a relevant tool for decision making based in evidence.

From the created and loaded data warehouse it is possible to collect the data to work in Power BI using the Power Query tool, to run the algorithm and obtain the needed information to display in the dashboards. These dashboards enable the stakeholders to have insights regularly based on a comprehensive view of documents quality produced in each period. In this chapter, charts are disposed instead of full dashboards pages to provide a better view of the information.

4.1. PROCESSES' QUALITY EVOLUTION

CJU considers that the BI system promoted digital transformation as together with the supported agencies, they became able to make business decisions based on accurate and current information, something that did not exist before the system implementation. The dashboards are publicized in the institution's website and became usable also in regular meetings that CJU have with the supported agencies, and together they conduct an analysis over the data. By making use of the system, it is possible to analyze the current situation, the most common and repeated mistakes, or missing documents, to help agencies to improve their bid notices acting in critical points of errors.

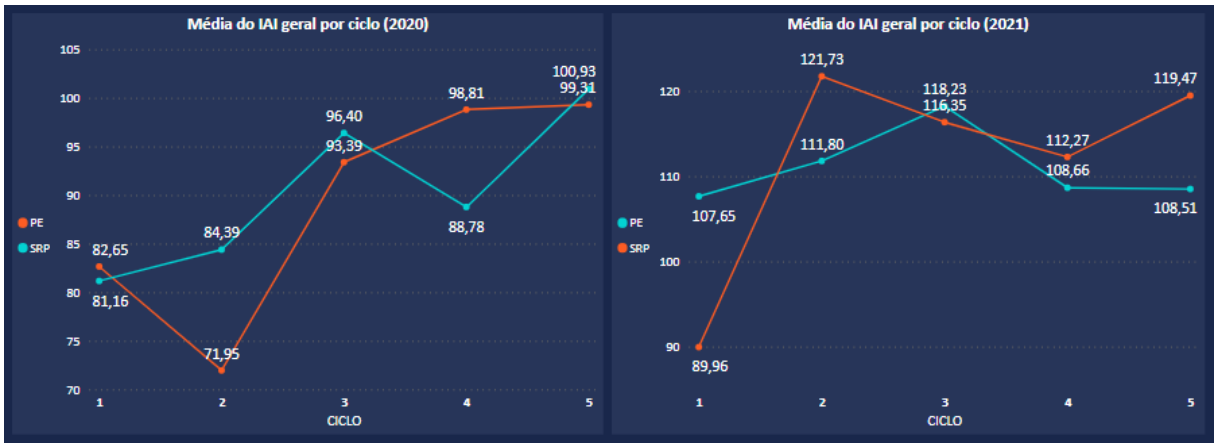


Figure 4-1 - ILI general evolution

The historical statistics obtained shows how the system improved the documents quality in public tenders. The ILI algorithm measures the documents quality by scoring first the matrices based on its errors and achievements, and then apply some reduction and learning factors that may increase or decrease the process final score. From the charts in the figure 4-1 above it is possible to visualize how was the situation in the beginning of the project, and how it is in the end of 2021. PE grew up from 82,65 to 108,51, and SRP modality from 81,16 to 119,47 points.

CJU consider the ILI average the most relevant factor to be analyzed, because it represents not only the processes' score, but essentially an evolution in the documents' quality. The ILI raising express that the documents have more achievements than errors, evidencing how the system helped to enhance bureaucracy, as raising the bidding quality means that the governmental contracts cycle became faster, with less mistakes and rework from the parties.

4.2. MATRICES' EVOLUTION

Figure 4-2 - The Matrices Average in 2021

The matrices displayed in Figure 4-2 show how much they evolved by modality in the period. The decrease of errors and the increase of achievements, evidence the overall processes' quality evolution displayed previously in this macrocycle. ILI's score in PE is quite stable in 2021, but the SRP modality went up mainly by the growth of achievements in matrices M1 and M5 in the period.

4.3. BEST PRACTICES

Another significant chart from the system is related to best practices. If the initial idea was that the system could provide a descriptive analysis identifying patterns of errors, this graphic displays the opposite as the system shows the processes that had 100% of achievements in each matrix by bid modality. This information is significant for CJU to implement institutional learning in processes' management, creating a collaborative and healthy competition, fostering the best practices sharing.

Figure 4-3 - Best Practices

These charts are available to identify the best practices in every cycle, as shown previously in the Figure 3-27 in the methods chapter, more focused in the supported agencies. In the macrocycles' dashboards, the system provides this information generally measured in the whole period. The Figure 4-3 above displays these data collected in 2020 and 2021.

4.4. PRIZING AND KNOWLEDGE SHARING

To stimulate the continuous improvement, annually CJU prizes the best agencies in categories that are the winners and the institutions that had the highest evolution, in each bid modality. The prizing occurs in the end of macrocycles when these dashboards are published containing the results from the whole period. Considered an important governance aspect by the institution, they created a collaborative community inviting the best ranked to share knowledge among all participating agencies. This was only possible for CJU with the implementation of this project, providing analytics, identifying trends found in data, helping the stakeholders to make better decisions based on data-driven insights, that is what business intelligence stands for.

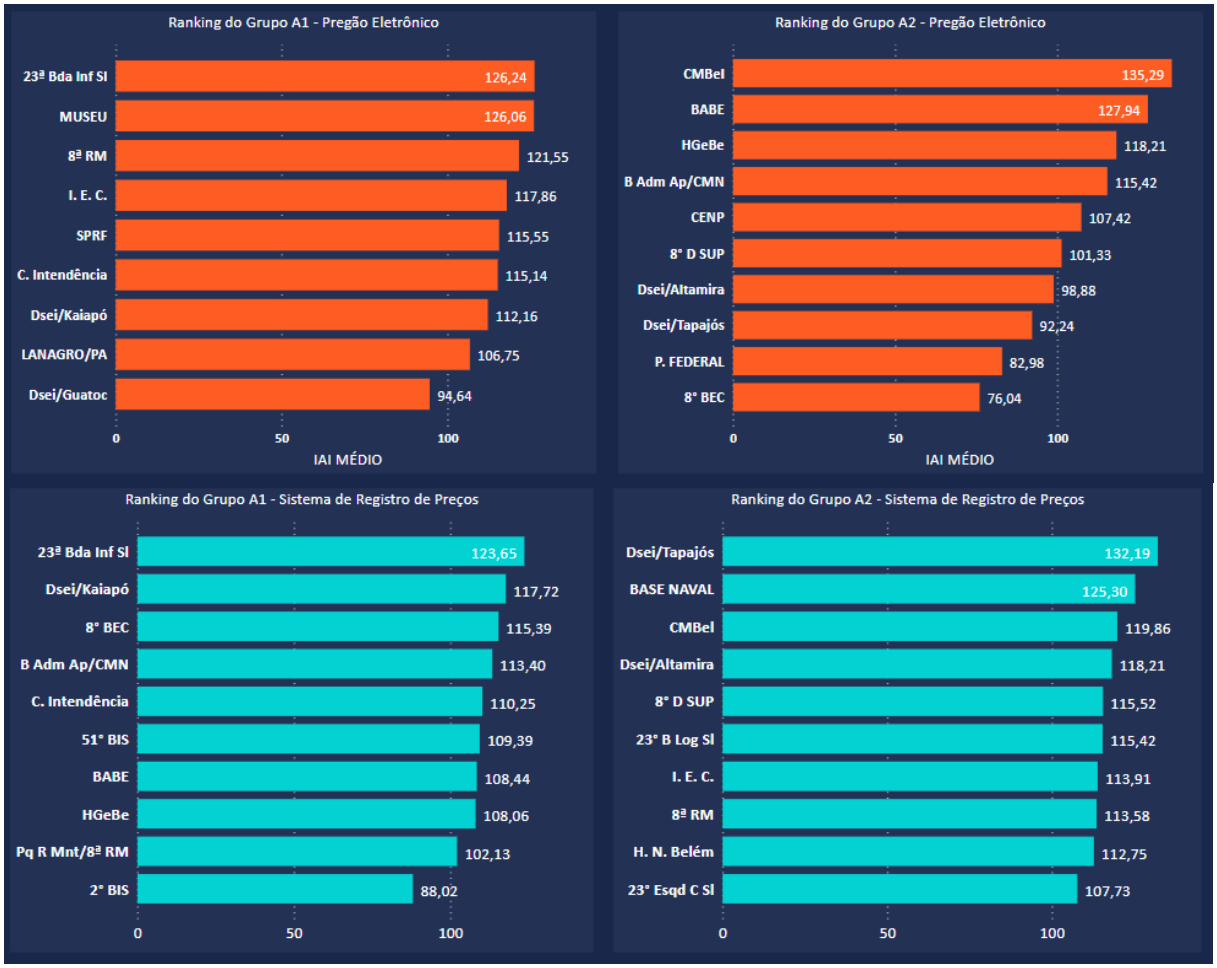


Figure 4-4 - The Agencies Ranking in 2021

The Figure 4-4 displays information about the ranking of agencies in 2021. In the end of each macrocycle, these charts are made available from the system, that are used by CJU to identify the best ranked institutions and prize them. Each modality is splitted in two groups (A1 and A2), in accordance with the quantity of processes evaluated by each agency in the period. This average value defines the threshold in the macrocycle.

Figure 4-5 - The Agencies Evolution in 2021

The charts in the Figure 4-5 above shows each agency ILI's score in the first and in the last cycle, producing the third column in blue displaying the difference between these values. This data obtained from the system is used by CJU to prize the other category that is the agencies with the greatest evolution per modality, in the period. Whereas the maximum score is 150 points, the overall results are meaningful despite some agencies' involution in PE (left) modality, but it is considerable that these organizations initially had a high score.

5. CONCLUSION

Institutionally, the biggest challenge is to change the governmental mentality to focus on problem solving instead of its prevention, to create a culture of decision making based in evidence. Technologically the challenge was how to transform legal reports in measurable data, to create charts and indexes representing the agencies performance in multiples aspects. After two years of system use, the proposed objectives were accomplished by this project, that were to:

- Build a scalable data warehouse to store information from bid documents
- Create the dashboards to provide analytics
- Improve the bidding processes

The research in literature made it possible to find a solution that better adapts to the business. The data warehouse created stores data from both modalities that are Pregão Eletrônico [*Electronic Auction*] and Sistema de Registro de Preços [*Price Registration System*], incrementally loaded at the end of each cycle. SRP differs from PE because it has one extra matrix and distinguished items analyzed. The DW was designed making use of a star schema and a not so common factless fact table, regularly exemplified when there are no measures on it, likewise the used model where legal processes are composed mainly by texts fields, as the only numbers are the unique processes' identification. The full DW considered at the end of a macrocycle is not heavy, and the whole system implementation is documented to the IT team continue the work.

From the Sharepoint list, the ETL processes made it possible to extract the data from the source file, transform it to match the designed data warehouse and then load it. There is no need of using DAX formulas in Power Query as processes are scored by using common functions like counting distinct lines, group by, and the algorithm runs to generate the agencies ranking and the ILI average. Together with the ETL processes, the Power Query is the hardest work as it takes up to 80% of the job time as well as related by the literature review. After gathering the process quantities, the matrices' errors and achievements, and the algorithm scores, the last stage is to carry out the graphical design and publish the dashboards to provide analytics to the stakeholders regularly whenever a cycle and macrocycle finish. The applied methodology fully achieved the system needs.

This work main objective for the business is to improve the legal bidding processes by making use of technology, and it was possible to accomplish it not only from the regularly analytics provided, but because of the involvement of the heads of CJU and the agencies' stakeholders that believed in the project since its release. The first published dashboard provided a snapshot of the current workflow situation at that time, and the results reached by this project proved the increasing in documents quality, mitigating the rework and the disapprovals, and reducing the bureaucracy by public institutions when making contracts. Since the beginning of the project, the local CJU/PA Office believed that it was possible to make the supported agencies to learn from mistakes committed by themselves or by the others. Discussing the results and encouraging the institutions to share their knowledge regularly, and prizing them based on their performance, proved that sharing best practices brings good results to everyone, as the average processes scores increased significantly across the time. That's the reason why this project is called PAI – Programa de Aprendizagem Institucional [*ILP - Institutional Learning Project*].

Figure 5-1 - CJU/PA website post of the award delivery to one of the prized agencies

6. LIMITATIONS AND FUTURE WORKS

The project by itself continues to evolve as the business need. In 2021 the consultants' workflow suffered a relevant change with the addition of new themes to be analyzed in the SRP modality, resulting in the inclusion of a new matrix (M6) in the algorithm, together with the Judicialization Factor that adds 20 possible points to a process. The judicialization approval is significant to the evaluation as it reduces the possibility of corruption or bringing the process to the court, and that is the reason of its inclusion in the analysis and in the algorithm formula.

Once the data warehouse is working fine, it is possible to extract some other information from it, not strictly related to this project but to the CJU workload, such as the processes' distribution to control how many processes are received and reported by each consultant or office, and a productivity control of the time spent to produce a report since the documents' distribution for each consultant or office. This type of work is not done but it is possible to obtain this kind of information from the DW as the business need. As the project is still evolving and opened to changes, there are some aspects that can be improved in the future, also some limitations as disposed next.

6.1. DIMENSIONS ATTRIBUTES

Some dimensions created containing few information like the Dim_Date where there is no weekday for example. This dimension does not have that granularity because the institution does not need it when dealing with dates. The same explanation applies to other dimensions Dim_Consultant and Dim_Agency that only contains their respective names. Despite meeting the needs there is an improvement that can be done adding more details on it like address, city, and region.

6.2. DUPLICATED RECORDINGS

Administrative peers are the ones responsible for recording the legal process in the provided web-forms, and sometimes happens that the same process is recorded twice in the web forms. The ETL tool remove duplicates, but there is a need to find a technological manner to avoid it inside the forms or even before this stage, in the process registration control.

6.3. DATA TYPES IN THE SOURCE FILE

One error found in the SSIS is the mismatch between the data types in the extracted Excel file and the designed database schema. When creating derived columns by running the ETL processes, it required some adjustments in the source file properties, changing some columns' data types from DT_DATE to DT_DBDATE and DT_NTEXT to DT_TEXT, because some functions like the REPLACE used does not support the mentioned original data type. There is a need to spend some time to investigate its origin, if it is from the fields relationship in the Automate tool, if it comes from the column definition in the Sharepoint list, or if it is a pattern while converting the lists to an Excel format.

7. BIBLIOGRAPHY

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *Management Information Systems Quarterly, 28*(1), 6.

Casters, M., Bouman, R., & Van Dongen, J. (2010). *Pentaho Kettle solutions: building open source ETL solutions with Pentaho Data Integration*. John Wiley & Sons.

Yessad, L., & Labiod, A. (2016, November). Comparative study of data warehouses modeling approaches: Inmon, Kimball and Data Vault. In *2016 International Conference on System Reliability and Science (ICSRS)* (pp. 95-99). IEEE.

Fourati-Jamoussi, F., & Niamba, C. N. (2016). An evaluation of business intelligence tools: a cluster analysis of users' perceptions. *Journal of Intelligence Studies in Business, 6*(1).

Kakish, K., & Kraft, T. A. (2012). ETL evolution for real-time data warehousing. In *Proceedings of the Conference on Information Systems Applied Research ISSN* (Vol. 2167, p. 1508).

Golfarelli, M., & Rizzi, S. (2009). A Survey on Temporal Data Warehousing. *International Journal of Data Warehousing and Mining (IJDWM), 5*(1), 1-17. http://doi.org/10.4018/jdwm.2009010101

Levene, M., & Loizou, G. (2003). Why is the snowflake schema a good data warehouse design?. *Information Systems, 28*(3), 225-240

Demarest, M. (1997). The politics of data warehousing. June, *http://www.hevanet.com/demarest/marc/dwpol.html, 6*(03), 1998.

Devlin, B. (1997). Managing time in the data warehouse. *InfoDB*, *11*, 7-11.

Larson, B., English, D., & Purington, P. (2012). *Delivering business intelligence with Microsoft SQL server 2016*. New York: McGraw-Hill.

Reddy, G. S., Srinivasu, R., Rao, M. P. C., & Rikkula, S. R. (2010). Data Warehousing, Data Mining, OLAP and OLTP Technologies are essential elements to support decision-making process in industries. *International Journal on Computer Science and Engineering*, *2*(9), 2865-2873.

Inmon, W. H. (1996). Building the Data Warehouse, John Wiley & Sons. *Inc., ISBN*, (0471-14161), 5.

Kimball, R. (1996). *The data warehouse toolkit: practical techniques for building dimensional data warehouses*. John Wiley & Sons, Inc.

Ross, M., & Kimball, R. (2013). *The data warehouse toolkit: the definitive guide to dimensional modeling*. John Wiley & Sons.

Kimball, R., & Merz, R. (2000). The data webhouse toolkit: Building the Web-enabled data warehouse. *Industrial Management & Data Systems.*

Ballard, C., Herreman, D., Schau, D., Bell, R., Kim, E., & Valencic, A. (1998). *Data modeling techniques for data warehousing* (p. 25). San Jose: IBM Corporation International Technical Support Organization.

Breslin, M. (2004). Data warehousing battle of the giants. *Business intelligence journal, 7*, 6-20.

Watson, H. J. (2009). Tutorial: Business Intelligence – Past, Present, and Future. *Communications of the Association for Information Systems, 25*. https://doi.org/10.17705/1CAIS.02539

Adzic, J., Fiore, V., & Spelta, S. (2001, September). Data warehouse population platform. In *International Workshop on Databases in Telecommunications* (pp. 9-18). Springer, Berlin, Heidelberg.

El-Sappagh, S. H. A., Hendawi, A. M. A., & El Bastawissy, A. H. (2011). A proposed model for data warehouse ETL processes. *Journal of King Saud University-Computer and Information Sciences, 23*(2), 91-104.

Nath, R. P. D., Hose, K., Pedersen, T. B., & Romero, O. (2017). SETL: A programmable semantic extract-transform-load framework for semantic data warehouses. *Information Systems, 68*, 17-43.

Dayal, U., Castellanos, M., Simitsis, A., & Wilkinson, K. (2009, March). Data integration flows for business intelligence. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology* (pp. 1-11).

Marín-Ortega, P. M., Dmitriyev, V., Abilov, M., & Gómez, J. M. (2014). ELTA: new approach in designing business intelligence solutions in era of big data. *Procedia technology, 16*, 667-674.

Vassiliadis, P. (2009). A survey of extract–transform–load technology. *International Journal of Data Warehousing and Mining (IJDWM), 5*(3), 1-27.

Vassiliadis, P., Simitsis, A., & Skiadopoulos, S. (2002, November). Conceptual modeling for ETL processes. In *Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP* (pp. 14-21).

Inmon, W. H., & Nesavich, A. (2007). *Tapping into unstructured data: integrating unstructured data and textual analytics into business intelligence.* Pearson Education.

Vassiliadis, P., & Simitsis, A. (2005). Extraction-Transformation-loading, (ETL) Processes in Data Warehouse Environments. *Encyclopedia of Database Technologies and Applications, Idea Group, 2009.*

Sadiku, M., Shadare, A. E., Musa, S. M., Akujuobi, C. M., & Perry, R. (2016). Data visualization. *International Journal of Engineering Research And Advanced Technology (IJERAT), 2*(12), 11-16.

Healy, K. (2018). *Data visualization: a practical introduction.* Princeton University Press.

Few, S. (2005). Dashboard Design: Beyond Meters, Gauges, and Traffic Lights. *Business intelligence journal,* 18-24.

Cairo, A. (2012). *The Functional Art: An introduction to information graphics and visualization*. New Riders.

Cairo, A. (2015). Graphics lies, misleading visuals. In *New challenges for data design* (pp. 103-116). Springer, London.

Komorowski, M., Marshall, D. C., Salciccioli, J. D., & Crutain, Y. (2016). Exploratory data analysis. *Secondary analysis of electronic health records*, 185-203.

Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. (2013). What makes a visualization memorable?. *IEEE transactions on visualization and computer graphics, 19*(12), 2306-2315.

Kosara, R., & Mackinlay, J. (2013). Storytelling: The next step for visualization. *Computer, 46*(5), 44-50.

do Brasil, S. F. (1988). Constituição da República Federativa do Brasil. *Brasília: Senado Federal, Centro Gráfico.*

Doing Business. (2020). *Contracting with the Government*. World Bank. Retrieved from https://archive.doingbusiness.org/en/data/exploretopics/contracting-with-the-government

Lei de Acesso à Informação, 12.527 Brasil (2011). http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm

Lei da Transparência, 12.527 Brasil, art. 3rd (2011). http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm

Lei Complementar, 73 Brasil, art. 11 (1993). http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp73.htm

Lei de Licitações e Contratos, 8.666 Brasil, art. 2nd (1993). http://www.planalto.gov.br/ccivil_03/leis/l8666cons.htm

Decreto de Regulamentação do Sistema de Registro de Preços, 7.892 Brasil (2013). http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2013/decreto/d7892.htm

Nova Lei de Licitações e Contratos, 14.133 Brasil, cap. 3 XLI (2021). https://www.in.gov.br/en/web/dou/-/lei-n-14.133-de-1-de-abril-de-2021-311876884

Decreto de Regulamentação do Pregão Eletrônico, 3.555 Brasil, art. 5 (2000). http://www.planalto.gov.br/ccivil_03/decreto/d3555.htm

Advocacia Geral da União. (2020). *Relatório de Gestão 2020*. AGU. https://www.gov.br/agu/pt-br/acesso-a-informacao/auditorias/processo-de-contas-anuais/relatorio_2021_03_31_completo_final_media.pdf

8. ANNEXES

The annexes display detailed information of all items within the web-forms available as checklists for each bid type, one for SRP and another for PE. The web-forms' fields and its items related are exhibited below as tables. The difference between them is that SRP has one extra field (Act) and more items to be analyzed in some of the matrices.

8.1. SRP WEB-FORM ITEMS

Field	Checklist
Disapproval	☐ Bidding method inadequate ☐ Inappropriate or outdated model of bid invitation and attachments ☐ Lack of call for bid or attachment ☐ Absence of essential documents (bidding for information and communication technology) ☐ Absence of essential documents (others) ☐ Lack of environmental licensing ☐ Problems with price research
M1 – Formalities and instruction of the process	☐ ABLE ☐ Formalization or Regularity of the process ☐ Authorization for opening ☐ Hiring justification ☐ Requisition of the object ☐ Checklist ☐ Term of Reference with the approval of the higher authority ☐ Intention of Price Register ☐ Presentation of the expression of interest of the participating agencies in the price register if any ☐ Presentation of the justification for the adhesion to the Price Register Act when allowed
M2 – Contract planning	☐ ABLE ☐ Document formalizing the requirement or justification of the hiring ☐ Preliminary technical studies ☐ Risk map or analysis ☐ SRP justification ☐ Justification for the objective's parceling or not (can be

	found in the Preliminary Technical Studies) ☐ Differentiated treatment for Micro and Small Companies ☐ Designation of the auctioneer and support team ☐ Formal designation of the planning team ☐ Adoption of environmental sustainability criteria
M3 – Statements	☐ ABLE ☐ Spending limit ☐ Preference margin ☐ Costing activity ☐ Governance limits ☐ Limits for rationalization of public spending
M4 – Quotation and budgets	☐ ABLE ☐ SIAFI Statement ☐ Price Survey ☐ Critical analysis ☐ Methodology adopted ☐ Justification for quotation with less than three suppliers ☐ Comparison chart or cost spreadsheet ☐ Formal errors in website searches ☐ Insufficient quotations ☐ Quotes ☐ Outdated quotations ☐ There was no discarding of unfeasible or exorbitant prices ☐ Budgetary and financial adequacy statement
M5 Q – Public notice qualification drafts	☐ ABLE ☐ Of the opening of the session/classification of proposals and formulation of bids ☐ The acceptability of the winning bid ☐ The qualification
M5 E – Public notice elements drafts	☐ ABLE ☐ Object ☐ Price registration ☐ Accreditation ☐ Participation in the bidding process ☐ Submission of the proposal and qualification documents ☐ Filling out the proposal ☐ Forwarding the winning proposal ☐ Appeals ☐ Reopening the public session

	☐ The awarding and ratification ☐ The performance guarantees ☐ The contractual guarantee of the goods ☐ Minutes of price registration ☐ The term of contract or equivalent instrument ☐ Adjustment in general ☐ Delivery and receipt of the object and inspection ☐ Contractor's and hired obligations ☐ Payment ☐ Administrative penalties ☐ Formation of a reserve list ☐ Refutation of the tender protocol and request for clarification ☐ General provisions ☐ Environmental sustainability ☐ Review and cancellation of the minutes ☐ Template mismatches the standard
M6 TR – Term of reference drafts	☐ ABLE ☐ Approval ☐ Absence of attachment to the bidding notice ☐ Object ☐ Justification and objective of the contracting ☐ Description of the solution ☐ Classification of common goods ☐ Delivery and criteria for acceptance of the object ☐ Contractor's obligations ☐ Hired obligations ☐ Subcontracting ☐ Subjective alteration ☐ Control of the execution ☐ Payment ☐ Readjustment ☐ Performance guarantee ☐ Contractual guarantee of the goods ☐ Administrative penalties ☐ Price estimates and reference prices ☐ Environmental sustainability ☐ Template does not match the standard
M6 TC – Term of contract drafts	☐ ABLE ☐ Clause one - object

	☐ Clause two - validity ☐ Clause three - price ☐ Clause four - budget allocation ☐ Clause five - payment ☐ Clause six - readjustment ☐ Clause seven - performance Guarantee ☐ Clause eight - delivery and receipt of the object ☐ Clause nine - supervision ☐ Clause ten - obligations of the contracting party and the contractor ☐ Clause eleventh - administrative penalties ☐ Clause twelve - termination ☐ Clause thirteen - prohibitions and permissions ☐ Clause fourteen - amendments ☐ Clause fifteen - of omitted Cases ☐ Clause sixteen - publication ☐ Clause seventeen - forum ☐ Template does not match the standard
M6 A – Act	☐ ABLE ☐ Preamble ☐ Object ☐ Prices specifications and quantities ☐ Managing agency and participants ☐ Adherence to the act ☐ Act expiration date ☐ Revision and cancellation ☐ Penalties ☐ General conditions ☐ Other

Table 8-1 - SRP form checklist items

8.2. PE WEB-FORM ITEMS

Field	Checklist
Disapproval	☐ Bidding method inadequate ☐ Inappropriate or outdated model of bid invitation and attachments ☐ Lack of call for bid or attachment ☐ Absence of essential documents (bidding for information and communication technology) ☐ Absence of essential documents (others) ☐ Lack of environmental licensing ☐ Problems with price research
M1 – Formalities and instruction of the process	☐ ABLE ☐ Formalization or Regularity of the process ☐ Authorization for opening ☐ Requisition of the object ☐ Checklist ☐ Term of Reference with the approval of the higher authority
M2 – Contract planning	☐ ABLE ☐ Document formalizing the requirement or justification of the hiring ☐ Preliminary technical studies ☐ Risk map or analysis ☐ Justification for the objective's parceling or not (can be found in the Preliminary Technical Studies) ☐ Differentiated treatment for Micro and Small Companies ☐ Designation of the auctioneer and support team ☐ Formal designation of the planning team ☐ Adoption of environmental sustainability criteria
M3 – Statements	☐ ABLE ☐ Spending limit ☐ Preference margin ☐ Costing activity ☐ Governance limits ☐ Limits for rationalization of public spending
M4 – Quotation and budgets	☐ ABLE ☐ SIAFI Statement ☐ Price Survey ☐ Critical analysis

	☐ Methodology adopted ☐ Justification for quotation with less than three suppliers ☐ Comparison chart or cost spreadsheet ☐ Formal errors in website searches ☐ Insufficient quotations ☐ Quotes ☐ Outdated quotations ☐ There was no discarding of unfeasible or exorbitant prices ☐ Budgetary and financial adequacy statement
M5 Q – Public notice qualification drafts	☐ ABLE ☐ Of the opening of the session/classification of proposals and formulation of bids ☐ The acceptability of the winning bid ☐ The qualification
M5 E – Public notice elements drafts	☐ ABLE ☐ Object ☐ Price registration ☐ Accreditation ☐ Participation in the bidding process ☐ Submission of the proposal and qualification documents ☐ Filling out the proposal ☐ Forwarding the winning proposal ☐ Appeals ☐ Reopening the public session ☐ The awarding and ratification ☐ The performance guarantees ☐ The contractual guarantee of the goods ☐ Minutes of price registration ☐ The term of contract or equivalent instrument ☐ Adjustment in general ☐ Delivery and receipt of the object and inspection ☐ Contractor's and hired obligations ☐ Payment ☐ Administrative penalties ☐ Refutation of the tender protocol and request for clarification ☐ General provisions ☐ Environmental sustainability ☐ Template mismatches the standard

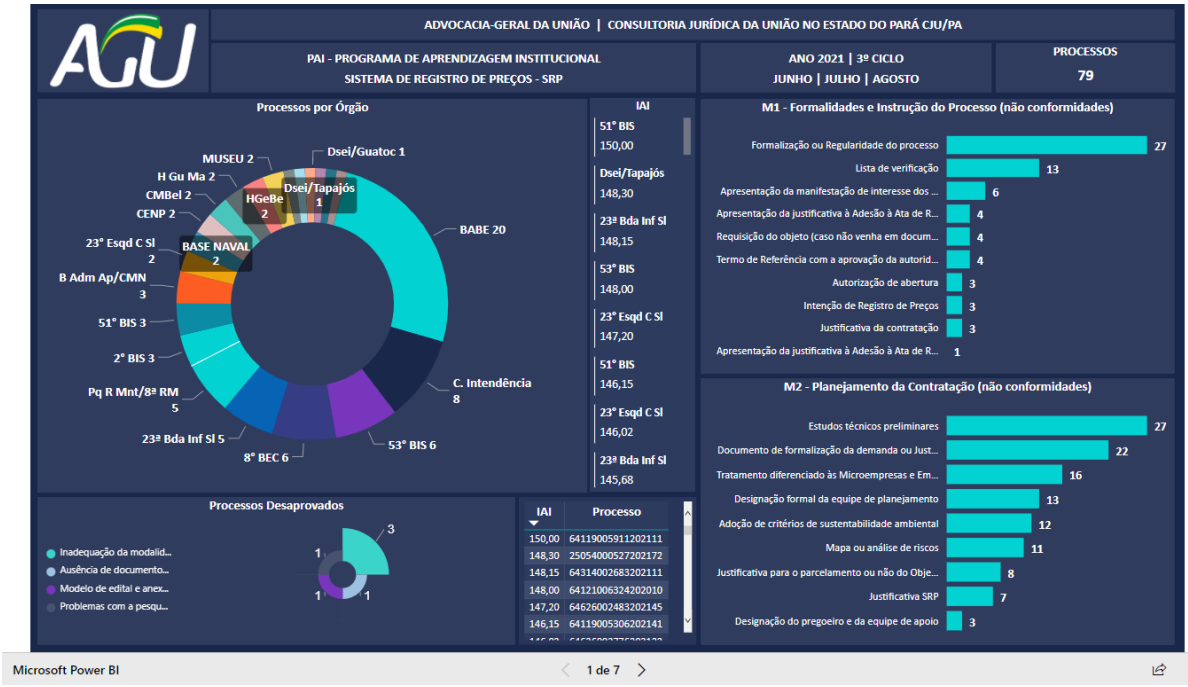
M6 TR – Term of reference drafts	☐ ABLE ☐ Approval ☐ Absence of attachment to the bidding notice ☐ Object ☐ Justification and objective of the contracting ☐ Description of the solution ☐ Classification of common goods ☐ Delivery and criteria for acceptance of the object ☐ Contractor's obligations ☐ Hired obligations ☐ Subcontracting ☐ Subjective alteration ☐ Control of the execution ☐ Payment ☐ Readjustment ☐ Performance guarantee ☐ Contractual guarantee of the goods ☐ Administrative penalties ☐ Price estimates and reference prices ☐ Environmental sustainability ☐ Template does not match the standard
M6 TC – Term of contract drafts	☐ ABLE ☐ Clause one - object ☐ Clause two - validity ☐ Clause three - price ☐ Clause four - budget allocation ☐ Clause five - payment ☐ Clause six - adjustment and amendments ☐ Clause seven - execution guarantee ☐ Clause eight - delivery and receipt of the object ☐ Clause nine - supervision ☐ Clause ten - obligations of the hired and the contractor ☐ Clause eleventh - administrative penalties ☐ Clause twelve - termination ☐ Clause thirteen - voidance ☐ Clause fourteen - of omitted cases ☐ Clause fifteen - publication ☐ Clause sixteen - jurisdiction ☐ Template mismatch the standard

Table 8-2 - PE form checklist items

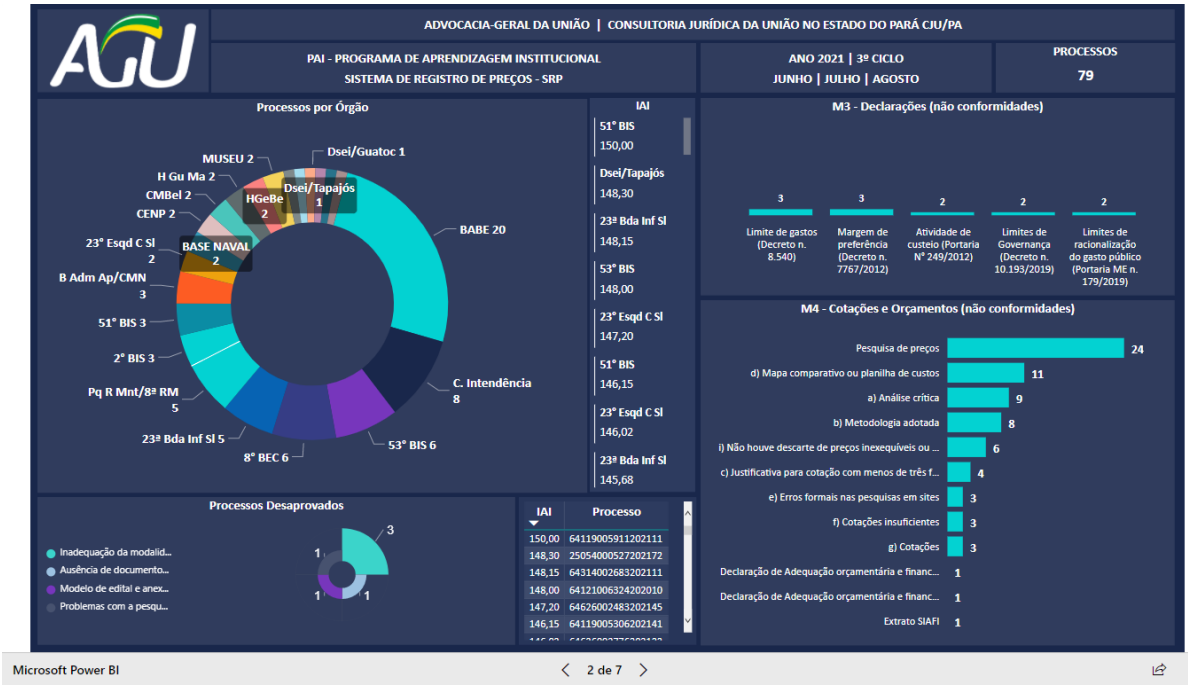
8.3. DASHBOARD PAGES

The images next were taken from the 2021's third cycle but only the SRP pages are displayed in the below because the PE dashboard has the same design. The only difference between them is that PE modality does not have the Matrix 6 "Acts" (M6A) and its chart.

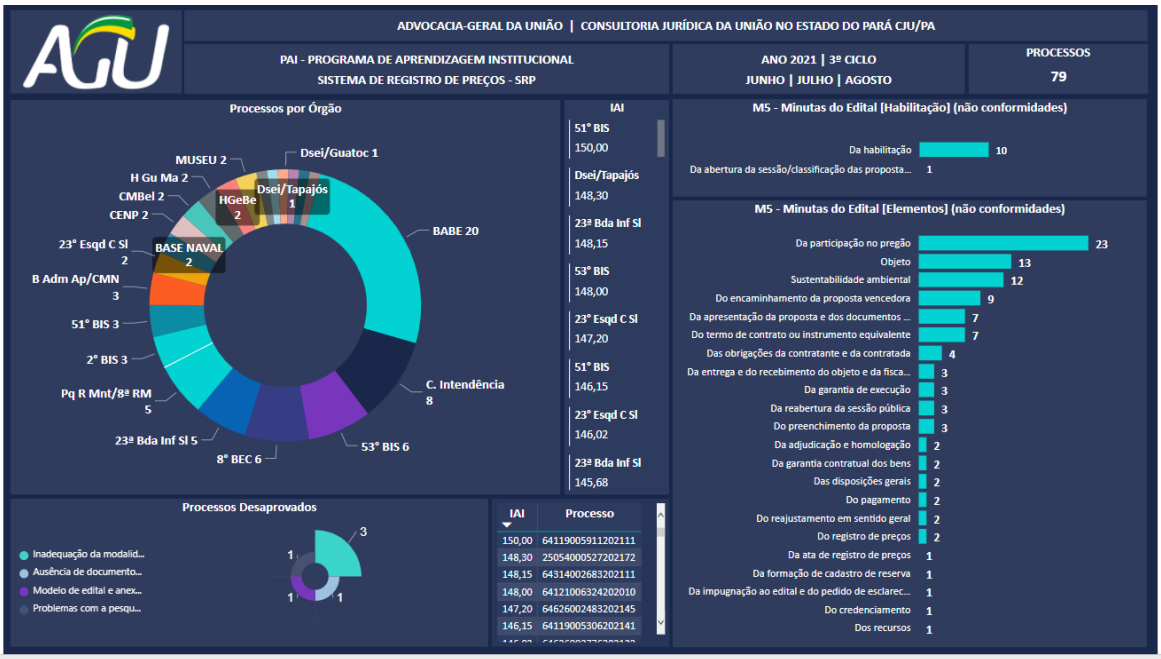
- 1st page with the matrices M1 and M2



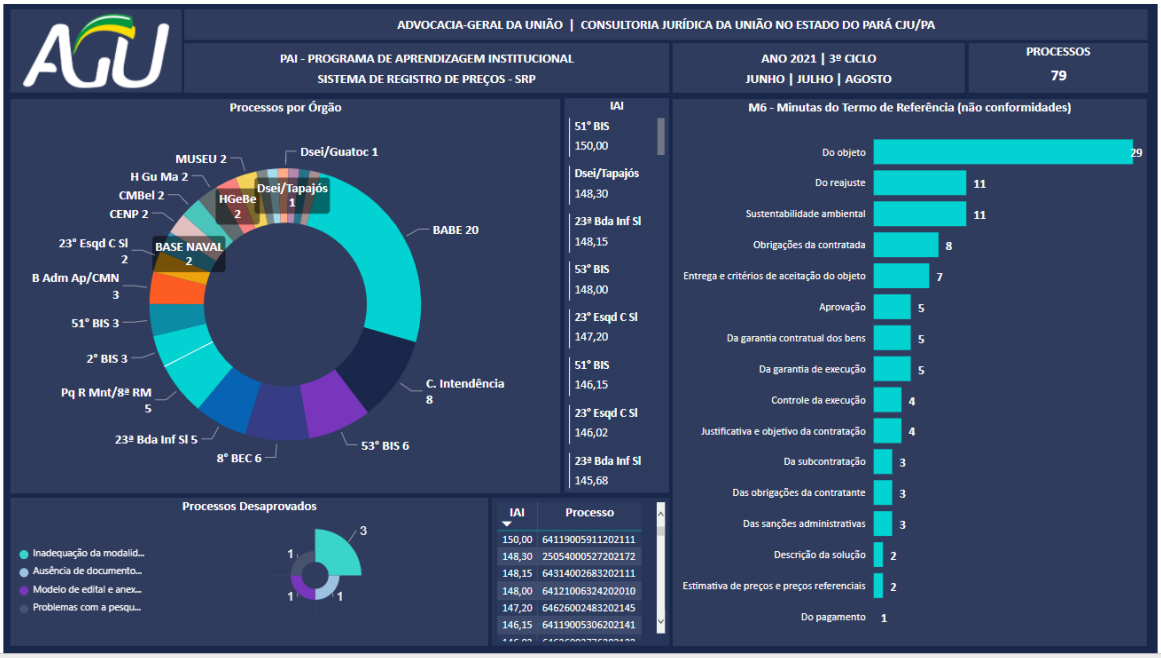
- 2nd page displaying the matrices M3 and M4



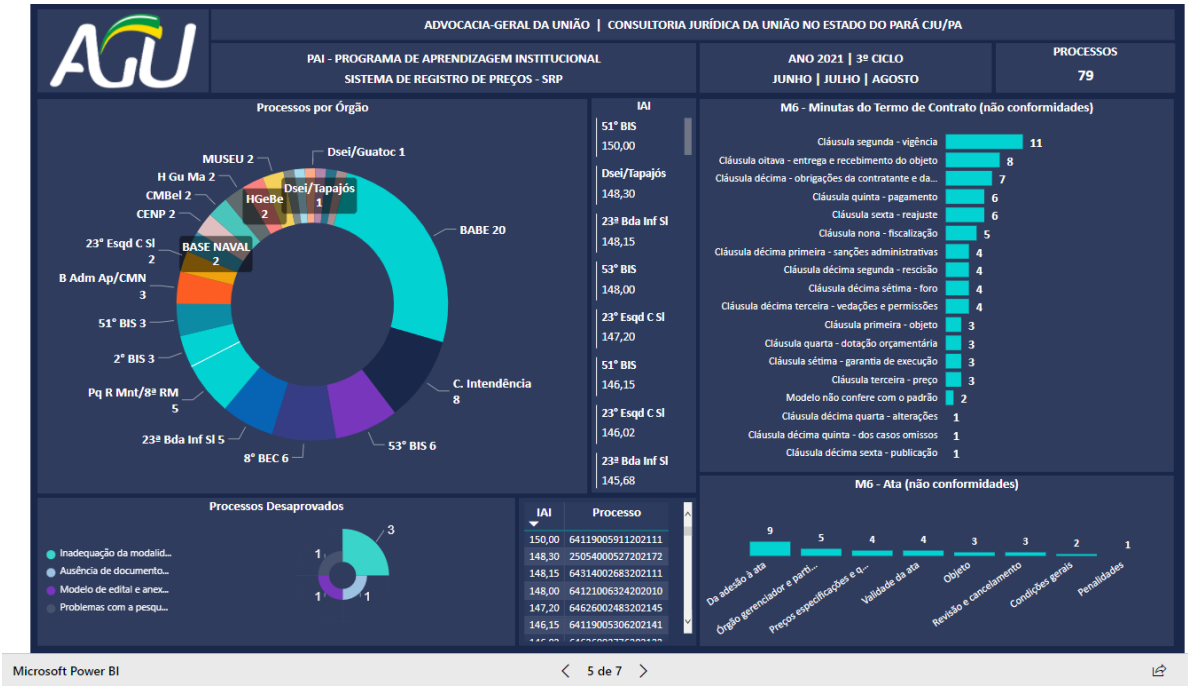
- 3rd page containing the matrices M5Q and M5E



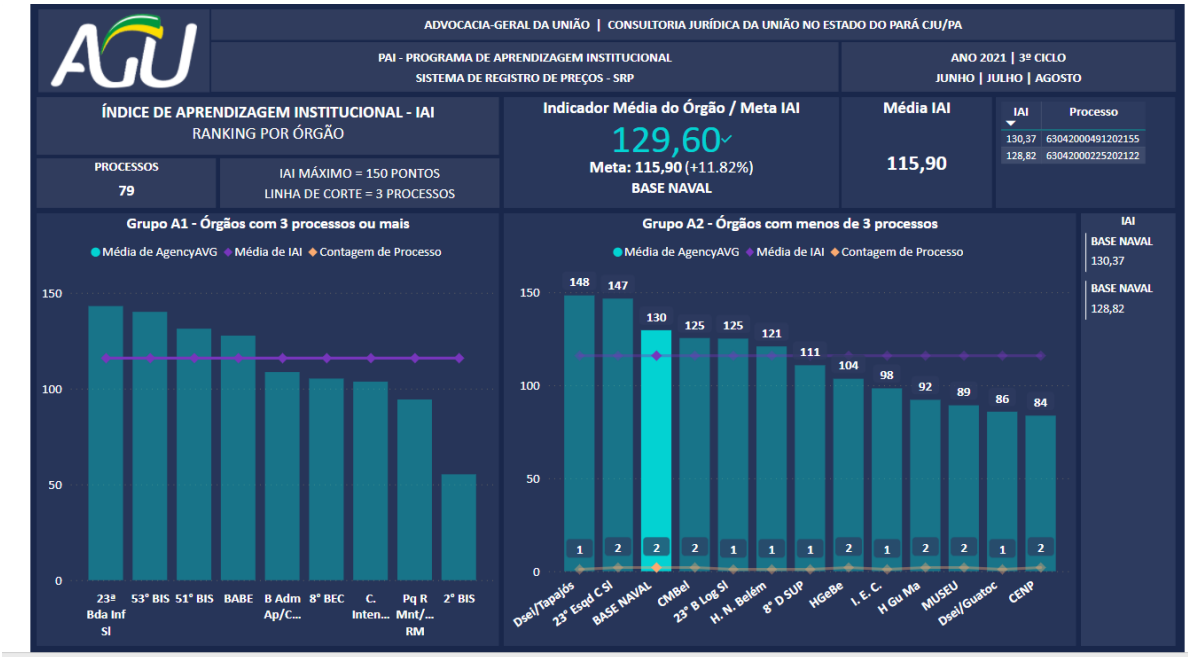
- 4th page with the M6TR matrix



- 5th page with the matrices M6TR and M6A, that is the matrix that does not exist for the PE modality where this page only displays the M6TR.



- 6th page with the ranking highlighting BASE NAVAL agency



- 7th and last page, showing the Good Practices in the cycle, highlighting the BABE agency numbers

8.4. SSIS ETL PROCESSES

The images disposed next were taken from the Visual Studio SSIS, showing the ETL process.

- Staging Area ETL

- Data Warehouse ETL