RegBR: AN APPLICATION OVERVIEW

Leticia Moreira Valle[1]

Stefano Giacomazzi Dantas[1]

Pedro Masson Sesconetto Souza[2]

Daniel Guerreiro e Silva[1]

Ugo Silva Dias[1]

[1]Universidade de Brasília (UnB), Brasília – DF, Brazil
[2]Escola Nacional de Administração Pública (Enap), Brasília – DF, Brasil

Government openness and transparency are key elements to build an accountable and trustworthy state, which are essential concepts to functioning democracies and market economies. Transparency in government processes increases citizen understanding and allows them to get involved by supervising and auditing government actions. RegBR was conceived as a framework to improve the legislative transparency, giving the citizens a tool to observe and monitor the Brazilian regulatory process and its characteristics. In this context, RegBR uses structured information to both create a historical regulatory flow and several regulatory metrics, which present how relevant are different sectors of the economy, how restricted are the normative acts, how linguistically complex are the regulations and, finally, how popular are theses normative acts through the eyes of the citizen and for the government.

Keywords: transparency; open government; natural language processing (NLP); machine learning; government regulation.

RᴇɢBR: UMA VISÃO GERAL DAS APLICAÇÕES

A abertura e a transparência do governo são elementos-chave para construir um estado responsável e confiável, conceitos essenciais para o funcionamento das democracias e economias de mercado. A transparência nos processos governamentais aumenta a compreensão dos cidadãos e permite que eles se envolvam na supervisão e auditoria das ações governamentais. O RegBR foi concebido como um framework para melhorar a transparência legislativa, dando ao cidadão uma ferramenta para observar e monitorar o processo regulatório brasileiro e suas características. Nesse contexto, o RegBR utiliza informações estruturadas tanto para criar um fluxo regulatório histórico quanto para conceber diversas métricas regulatórias, que apresentam quão relevantes são os diferentes setores da economia, quão restritos são os atos normativos, quão linguisticamente complexas são as regulamentações e, por fim, quão populares são esses atos normativos pelos olhos do cidadão e para o governo.

Palavras-chave: transparência; governo aberto; processamento de linguagem natural (PLN); aprendizado de máquina; regulamentação governamental.

RᴇɢBR: UNA VISIÓN GENERAL DE LAS APLICACIONES

La apertura y transparencia del gobierno son elementos clave para construir un estado responsable y confiable, que son conceptos esenciales para el funcionamiento de las democracias y las economías de mercado. La transparencia en los procesos de gobierno aumenta la comprensión de los ciudadanos y les permite involucrarse supervisando y auditando las acciones gubernamentales. RegBR fue concebido como un *framework* para mejorar la transparencia legislativa, brindando a los ciudadanos una herramienta para observar y monitorear el proceso regulatorio brasileño y sus características. En este contexto, RegBR utiliza informaciones estructuradas para crear un flujo regulatorio histórico y también varias métricas regulatorias, que presentan qué tan relevantes son los diferentes sectores de la economía, qué tan restringidos son los actos normativos, qué lingüísticamente complejas son las regulaciones y, finalmente, qué tan populares son estos actos normativos a los ojos del ciudadano y para el gobierno.

Palabras clave: transparencia; gobierno abierto; procesamiento del lenguaje natural (NLP); *machine learning*; regulación gubernamental.

1. INTRODUCTION

In the past decade, governments worldwide have been focusing efforts on trying to predict the impacts of new regulation before they are published. However, governments have been paying remarkably little attention to analyzing regulations after adoption or evaluating the impacts of procedures and practices that govern the regulatory process itself (Cary, 2012).

As Brazil ranks at 46th place out of 48 countries evaluated by OECD in terms of regulatory performance, the country urgently needs scientific contributions that study regulatory policies and help to metrify the legislative processand allow policymakers to understand whether regulations efficiently achieve their intended goals and prioritize regulations that may need reforms.

In addition to promoting studies that support improvements to regulatory performance, the government should use information and communication technologies (ICTs) to promote openness and transparency around their legislative operations and decisions. The ICT concept is an umbrella term that includes any sort of technological tools and resources used to store, process, transmit, share or exchange information, e.g. software applications and operational systems, web based information and applications such as websites and blogs, broadcast technologies including radio, television or podcasts and others infrastructure and components that enable modern computing (Pandey, 2021). Besides being essential for modern life, these tools are a cost-effective and convenient way to promote openness and transparency by allowing the citizen to track activities of common interest and enabling monitoring, controlling and discussing behaviours and tendencies.

The Open Government Idea can be defined as using compatible standards and architectures to improve people's access to data (Fishenden & Thompson, 2013). Finally, being an open government is not only related to making information available but also opening a communication channel with the citizens, which informs and receives feedback, and, more importantly, acts based on feedback. This process puts governments into a more active and collaborative role, and enables them to reach the full open government state (Abu-Shanab, 2015). For instance, designing more intuitive websites with a focus on creating reliable and publicly accessible infrastructure that exposes the underlying data is essential for the success of open government initiatives (Robinson et al., 2008).

The culture of Open Government promotes the principles of transparency, openness and accountability, which builds trust in government and provides citizens with a better understanding of governmental processes that impact the economy and, consequently, their lives. As a tool that aims to increase transparency in the Brazilian legislative processes, the RegBR framework is based on the American initiative RegData (Al-Ubaydli & Mclaughlin, 2017). The RegData was created by the Mercatus Center[1] at George Mason University, to quantify federal

---

1   https://www.mercatus.org/publications/regulation/regdata

regulations by industry and by the regulatory agency for all federal regulations of the United States. The metrics expanded in RegData include a measurement of the applicability of each regulation to each one of the industries that comprise the US economy using information from the regulatory text. This work acts as both a framework and a database to analyze regulations in the US. This methodology was also expanded to other countries, such as Australia (McLaughlin et al., 2019a) and Canada (McLaughlin et al., 2019b). This genre of work is fascinating as once researchers achieve a methodology to objectively measure regulations, policymakers can use this information to gather insights about their policy impact and even evaluate one's own work, in terms of whether the legislative characteristics are in accordance with what was initially planned by the government.

Various other organizations are also developing frameworks to help countries evaluate the design and implementation of their regulatory policy. One of the best examples is the Regulatory Policy Committee from the Organization for Economic Cooperation and Development (OECD). This Committee's objective is to assist countries in building and strengthening capacity for regulatory quality and regulatory reform (Parker & Kirkpatrick, 2012).

Among several indicators compiled by OECD that capture the level of anti-competitive regulation in the economy, the most relevant is the Product Market Regulation (PMR). The economy wide PMR indicator covers state control, barriers to entrepreneurship, and barriers to trade and investment (Egert & Wanner, 2016). However, in the context of developing economies, the regulation impacts are not as well known and studied as in developed countries. Specifically, there is no framework to study and analyze regulations in Brazil. The country figured on the bottom three countries in the economy-wide PMR indicator for 2018 (Vitale et al., 2020).

In this context, this paper presents the RegBR framework and its various applications in the Federal Government, including use cases related to regulatory impact studies, feedback mechanism for regulatory agencies, monitoring tool for citizens and policymakers, comparison tool to compare against other countries that also apply similar metrics. The RegBR framework is also an instrument of transparency, presenting information that was previously decentralized, now in an easily accessible and understandable way.

With this purpose, this article is structured in five parts: the subject contextualization and introduction; a review of the national and international literature about transparency, regulatory impact studies and use of artificial intelligence in legal domain in the second part; the methodology and the database, presented in the third section; discussion of the applications of the framework on the Federal Government in the fourth section and our concluding remarks in the fifth section.

2. LITERATURE REVIEW

This section presents a brief review of related literature around the use of information and communication technologies to promote openness and transparency tools for society, the impact of regulations as policy tools and a review of artificial intelligence across the legal domain.

2.1 ICTs as transparency tools

The usage of ICTs to promote openness and transparency has been increasing in recent years, as it is often considered cost-effective and convenient (Bertot et al., 2010). In particular, e-government initiatives aim to improve the efficiency, effectiveness, and quality of government services and activities by employing ICTs. Institutions have modified not only the organization of activities and processes within organizations but also the public perception of governmental institutions and policy-making decisions (Orlikowski & Barley, 2001; Mitchell & Zmud, 1999). Some benefits of ICT adoption include promoting Food Governance, strengthening reform-oriented initiatives and enhancing relationships between government and citizens, with successful cases from governments across the Americas, Asia, and Europe (Shim & Eom, 2008).

The scope of e-government studies is expanding to consider not only government basic operations and service delivery, but also to enable citizen participation and engagement using technology tools (Abu-Shanab, 2015). In other words, the interactions between governments and non-state stakeholders is gaining more importance. For example, some governments are adopting social media to disseminate complementary information and provide participation channels to citizens, with the goal of increasing perceptions of government transparency (Song & Lee, 2016). This effect is a consequence of making up-to-date government information easily accessible and providing more interactions between the government agencies and the public. The aforementioned are some efforts that are part of the open government concept, which is based on making public information access freely and more efficiently.

Intuitive web applications, such as dashboards, are essential in bridging the gap between the government and the public (Matheus et al., 2020; Velcu-Laitinen & Yigitbasioglu, 2012). The use of these data instruments can influence government policy-making and provide more interaction with the public, as they can be interactive and can be used to release information for both governmental decision-makers and the public (Maheshwari & Janssen, 2014; Brown et al., 2017). Additionally, dashboards can also be used to verify data integrity and quality, an essential pre-requisite for decision-making contexts (Chengalur-Smith, et al., 1999). ICT tools help to increase transparency promotes good governance and instill trust in the government (Allio, 2012). Moreover, ICTs also enable citizens to participate in the decision-making process that help craft new public policies.

2.2   Review of regulatory impact studies

Government regulations are considered a crucial policy tool for addressing market inefficiencies (Pigou, 2013), as they can affect economic agents differently. Regulations are intended to correct market failures and, therefore, increase economic efficiency and growth. In practical terms, the government can also intervene and regulate cases where people should have access to certain services and goods regardless of the ability to pay, such as health care and education services.

Most of the impacts are the consequence of constraints or expansions of their legal rules, and this intervention can play a critical role in the successful development efforts of the economy (Stiglitz, 1989; Coates & Heckelman, 2003). However, the complexity of the market and other political processes often result in regulations that are not honourably created (Stigler, 1971; Mcchesney, 1987). In other words, regulations driven by the particular interest of specific groups lobbying for legislative changes that result in personal gain, e.g. *rent-seeking* behavior (Rowley et al., 1988). Even when this is not the case, regulations may result in unintentional consequences (Peltzman, 1975) or may do so at an unsatisfactory cost in terms of economic distortion.

Regulations can act as a potential aid or risk to every industry in the economy. By acting as a force that has the power to stimulate or restrain, to take or give resources, the government can help or hurt a vast number of industries at its discretion. Thus, it is crucial to evaluate the causal effect of regulations on the economic sectors. However, the studies that examine these impacts are often focused on one specific regulation or sector. Some examples include examining the impact of liquidity regulation on the banking sector (Banerjee & Mio, 2018), the differences in regulation for collaborative economy peer-to-peer accommodation in different European cities (Dredge et al., 2016), and regulatory risk and the resilience of new sustainable business models in the energy sector (Robin et al., 2019) in Germany. Compared to thousands of existing regulations that govern a country's economy, these studies are comparatively limited in scope.

One alternative is to study and analyze the impact of regulations and create indicators based on the time evolution of the number of legal documents that specify regulations. Works such as (Crews, 2002; Coffey et al., 2012) use data from the Code of Federal Regulations (CFR), an annual publication that contains all regulations issued at the federal level, to create quantitative measurements of federal or state regulations created or in effect in the United States, each year.

Nonetheless, there have been noteworthy developments to address regulatory issues in some sectors. One example is the logistic infrastructure sector (Lodge et al., 2017). Despite these developments, it is still imperative to adopt a more strategic perspective on logistics infrastructures. The authors provide several recommendations for enhancing regulation in logistics infrastructure in Brazil.

On a different note, the role of government regulation in ecological restoration is discussed in (Aronson et al., 2011). More specifically, the work focuses on the state of São Paulo's goal to

increase the effectiveness of tropical forest restoration projects. The authors argue that some points about ecological restoration are still unclear. In other words, there is no firm consensus. As such, these regulations are still determing whether or not to set goals for preservation of legal instruments or if legislation on similar topics should be delayed until adequate scientific knowledge is available.

Another sector in which regulations are constantly being debated in Brazil is the sanitation sector. The main challenge exists at the government level where different entities have not conceded authority as to where control should reside and how private operators can fulfill social objectives (Da Motta & Moreira, 2006). According to Da Motta and Moreira, these issues are not the crucial barriers to the sector's development when one looks at the operators' productivity performance. Instead, they suggest that operators dissipate their productivity potential and apply higher tariffs in the absence of efficiency incentives. That is to say that the debate over the regulatory framework should be redirected to focus on what instruments should be put in place to create incentives to efficiency and increase sharing of the resulting gains with users.

2.3   Review of Artificial Intelligence on legal domain

In recent months, the world has seen the emergence of several new initiatives to define how AI systems should be regulated. OEDC (Organization for Economic Development and Cooperation) by itself has already identified 700 AI policy initiatives in 60 countries, with new Artificial Intelligence legislations. It is expected that AI systems will be explicitly and comprehensively regulated in the future. However, AI systems can not only be regulated, rather, they can also be used to help policymakers to analyze regulations after their adoption or to evaluate the impacts of the procedures and practices that govern the regulatory process itself. One of the most emblematic works in this area is RegData (Al-Ubaydli & McLaughlin, 2017) by Mercatus Center, defined as both a methodology and a database focused on the quantification of various dimensions of regulation on various jurisdictions in the United States, Canada, and Australia.

For the private sector, Artificial Intelligence can also be used to assist industries and markets to ensure regulatory compliance. There are already companies (IBM, 2018) that use analytics to create real-time frameworks to track, monitor, visualize, and alert against any violation across the organization. Regulatory analytics, a growing category of information analysis, involves gathering and storing relevant data and leveraging statistical models and machine learning to mine for patterns, discrepancies, and anomalies (Mcgovern, 2017). The goal of using regulatory analytics is to detect and prevent potentially improper transactions or wrongful actions before they create peril for the firm.

Artificial Intelligence is also being leveraged to assure legal text classification or predict court decisions. Based on the association between a legal text and its domain label in a database of legal texts, the authors (Boella et al., 2011) present a classification approach to identify the

relevant domain to which a specific legal text belongs. A Support Vector Machine (SVM) classified the data with a polynomial kernel, and the authors evaluated the results via identifying topics covered by a piece of legislation and via the classification of individual articles.

Some of the works presented in the literature aim to predict court ruling decisions. For example, (Katz et al., 2017) used extremely randomized trees to predict the US Supreme Court's rulings and, more recently, (Wongchaisuwat et al., 2017) tackled the task of predicting patent litigation and time to litigation. Finally, (Aletras et al., 2016) proposed a model to predict the verdicts of the European Court of Human Rights (ECRH). Another example includes (Sulea et al., 2017), where the authors present results of machine learning algorithms in the task of predicting the decisions of the French Supreme Court and the law area to which a case belongs.

In (Yuan et al., 2019), the authors use three different approaches for judgment prediction. The models achieved many results on the accuracy of charges prediction and relevant law articles prediction, but poor results in recall and precision scores. In a more related approach to our work, (Tuggener et al., 2020) creates a multilabel corpus of legal provisions in contracts. That is achieved by first crawling and scraping Security and Exchange Commission (SEC) filings and then using machine learning to classify the extracted corpus.

With this in mind, we corroborate the need for new research that involves the use of disruptive technologies in studying the legal domain, specifically the regulation sector in emerging economies, such as Brazil. Hence, we propose a framework applicable to Brazil, which aims to produce relevant information on the national regulatory situation, promoting active transparency, centralizing information of different sources, compiling results and reducing regulatory information access costs.

3. METHODOLOGY

RegBR's methodology is separated into two parts. The first part is related to classifying regulatory texts in different sectors of the economy using machine learning models. This is necessary as the laws produced in Brazil do not formally mention which sector of the economy they regulate, and the legislative body is too large to be classified manually. The second part uses this labeled information to apply metrics that measure different aspects of regulation, which can be monitored by different economic sectors, such as linguistic complexity, restrictiveness, citation influence and popularity over time.

In the context of normative acts text classification, we can formally define eighteen different classes based on the Brazilian Institute for Geography and Statistics (*Instituto Brasileiro de Geografia e Estatística* - IBGE) economic sector classification. Hence, we define Text Document Classification (TDC) as the task of assigning a label for each normative act in our dataset.

Additionally, we propose the following metrics to allow regulators and policymakers to identify better and prioritize regulations that may need reform:

- **Restrictiveness**: indicates the occurrence of restrictive words/terms in the body of text. It can be used as an indication of the restrictiveness of regulation over time. This metric is adapted from Al-Ubaydli & McLaughlin (2017).

- **Industry citation relevance:** Calculates the relevance of regulations to economic sectors and industries, according to keywords defined by the machine learning model and their frequencies considering the regulatory body of all sectors, for a given year.

- **Popularity:** Indicates how "popular" a law is. This is evaluated for two audiences: the population, based on the active search for that law on Google, and the government, based on the frequency of citations of the law in the Official Gazette of the Federal Government. This metric is a novel contribution from RegBR.

- **Linguistic complexity:** This set of metrics aims to indicate whether or not regulations are employing more complex wording in their text. It uses statistics from the words to measure the linguistic complexity of each document. These metrics are adapted from McLaughlin et al. (2019b).

3.1 Dataset annotation

One of the main contributions of RegBR is the centralization of all federal regulatory acts in one place, which reduces the barrier to data access and increases government transparency. This infrastructure is available to researchers and the general public as a database in https://infogov.enap.gov.br/regbr.

As previously mentioned, the normative acts are not originally labeled according to the economic sector they affect. Hence, we need to employ classification methods to label them. Nonetheless, a subset of norms has to be manually labeled in order to train the machine learning models.

These labels were defined according to the National Classification of Economic Activities (*Classificação Nacional de Atividades Econômicas*, CNAE) (IBGE, 2007). The CNAEis the official categorization adopted by IBGE for the production of statistics by type of economic activity and by Public Administration, identifying economic activities in registrations of legal entities.

The original CNAE is structured in twenty-one main categories and four hierarchical levels. Due to the project scope, only the first hierarchy level is used in this work and some sectors were grouped together according to their similarity. Specifically:

- Classes 7 (trade and repair of vehicles), 9 (accommodation and meals) and 12 (real estate activities) of the CNAE were merged in 7: Commerce, Accommodation and Food, and Real Estate Services.

- Classes 19 (Other service activities), 20 (Domestic services) and 21 (International organizations) were merged in 17: Other services.

In addition to the 17 sectors, an extra label was added to indicate laws that do not deal with regulatory activities. The final classes are the following:

Table 1: Final Classes

Class	Definition	Class	Definition
Class 1	Agriculture, livestock and forestry	Class 10	Financial, insurance and related services
Class 2	Extractive industry	Class 11	Professional, scientific and technical activities
Class 3	Transformation industry	Class 12	Administrative activities and complementary services
Class 4	Electricity and gas	Class 13	Public administration, defense and social security
Class 5	Water, sewage and waste	Class 14	Education
Class 6	Construction	Class 15	Human health and social service
Class 7	Commerce, accommodation and food, and real state services	Class 16	Arts, culture, sports and recreation
Class 8	Transportation, storage and mail	Class 17	Other services
Class 9	Information and communication	Class 18	Non-regulatory

Source: elaborated by the authors.

The manual labeling of the dataset was performed by a consultant with expert domain knowledge on normative acts structure and content, following the definitions illustrated in Table 1. This dataset contained about 8 thousand labeled acts divided into 18 classes, with publication dates ranging from the end of the 19th century to recent years.
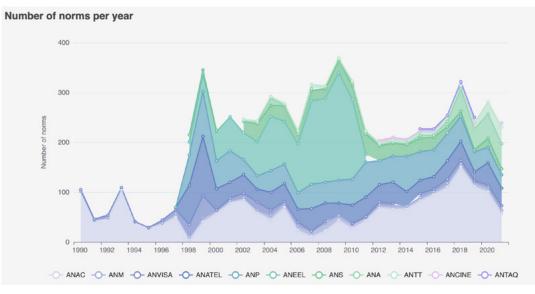
We evaluated different models and approaches using a standard data split of 75% training, 25% testing. More specifically, we tested statistical methods based on word frequency and/or word embedding techniques to transfer learning approaches. Surprisingly, statistical models emerged as the best performing approaches on both accuracy and macro averaged F1 scores. It is also important to notice that the statistical models are significantly faster for training and testing when compared to the implementations using deep neural networks.
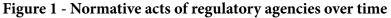
3.2 Regulatory stock in numbers

Since the publication of Law no. 10.139, of 28 November 2019, which imposed the review and consolidation of all normative acts with a hierarchy lower than the decree by the end of 2021, the regulatory stock management became more relevant for the Brazilian government.

To adopt these new requirements, some regulatory agencies are defining specific working groups for mapping their regulatory stock. In this context, RegBR presents a bird's-eye view of the Brazilian regulatory stock that can be filtered by the economic sector or by the regulatory agency. This aggregation aims to assist regulatory authorities in managing the country's regulatory stock

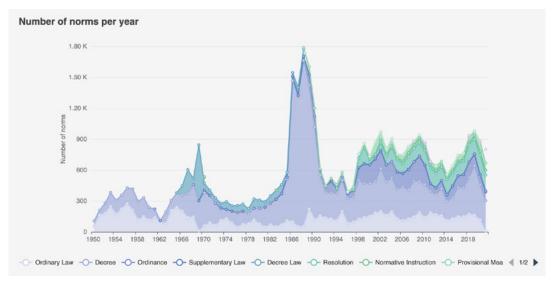and to better adjust the regulatory process to international quality standards.

Figure 1 illustrates the evolution over time of the normative acts *Resolutions* and *Ordinances* by agency. Using the same tool, Figure 2 presents the number of normative acts divided by type. It is possible to correlate the number of norms per year with the creation of new regulatory agencies or important national events in Figure 1. For instance, ANATEL was created in 1997, and this agency lead the number of norms creation in 1998. Another example is the increase of average number of norms created by ANEEL (Brazilian Electricity Regulatory Agency) during the period of frequent energy blackouts and the need for energy rationing at the beginning of the 2000s.

Figure 1 - Normative acts of regulatory agencies over time

Source: InfoGOV

Similarly, it is easy to note that the peak in the number of new norms in Figure 2 correlates with the redemocratization of Brazil in 1985 and the promulgation of the 1988 Constitution.

Figure 2 - Regulatory stock over time

Source: InfoGOV

3.3 Regulatory Metrics

Since the aftermath of the 2007-2008 financial crisis, discussions about regulation and the relationship of private and public interest are taking new shapes (Castro, 2014). As a consequence, researchers are aiming to create metrics that are simple yet efficient in quantifying regulations and performing ex-post evaluation (OECD, 2003).

In this context, RegBR provides a variety of metrics that can be used by regulators and policymakers to identify better and prioritize regulations that may need reforms. More specifically, the following are provided: restrictiveness, industry citation relevance, popularity, and linguistic complexity metrics.

3.3.1 Restrictiveness

Due to the complexity often found in regulatory texts, the first attempts of quantifying regulations were rather simple (Al-Ubaydli & McLaughlin, 2017). For instance, Mulligan & Shleifer (2005) used the file-size data to quantify regulations. However, this metric can be affected by large graphics or other embedded files in the document.

A new method to quantify regulation is presented in Al-Ubaydli & McLaughlin (2017), where the word count of restrictive words such as 'shall' and 'must' was calculated for each document. These words often indicate specific obligations or prohibition in a regulatory context (McLaughlin et al., 2019b), so their presence in the text represents restrictiveness.

In RegBR, this metric is adapted in two ways. First, a list of Portuguese words that intend to restrict legal scopes is used. This list[2] was proposed by the authors and validated by law professionals with vast experience in the legislative field. Second, the number of restrictive words was normalized each year according to the number of new publications. This normalization allows the comparison of restrictive word frequencies across years with differing number of normative acts.

3.3.2 Industry citation relevance

Another metric employed in RegBR is the industry citation relevance. This measure attempts to represent the influence of each economic sector, defined by CNAE, for a given year, which is calculated based on the use of keywords from each sector in the whole corpus of normative acts for that particular year.

The key idea is that if words directly related to a sector are often used, this sector tends to have a higher importance or prioritization in the context of regulatory legislation. The keywords were derived from the most relevant features[3] of the machine learning model used to classify the corpus.

---

2   In Portuguese, the restrictive words are: vetado, vedado, defeso, proibido, negado, determina, obriga, ordena, impõe, limita, delimita, demarca, restringe, confina, reduz, define, deve, deverá, precisa and necessita.
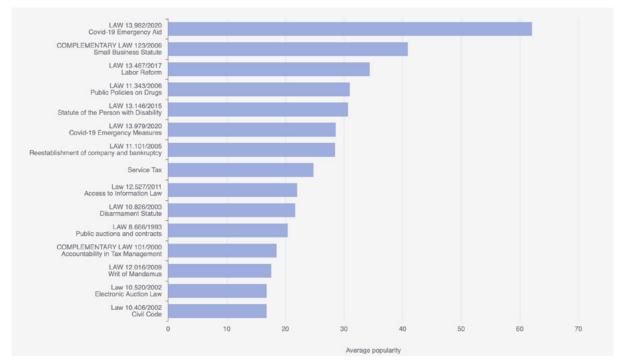
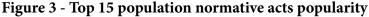3   More specifically, these are words whose corresponding weights had largest values.

After calculating this metric for each sector and year, the values were normalized in the range of 0 to 1, where 1 represents the most relevant sector in the year. This type of analysis can help citizens to assess government's priorities, increasing transparency. For example, after a pandemic, one would expect a higher industry citation relevance on the human health sector, or during an economic crisis, an increase in the relevance of the financial sector.

3.3.3 Normative act popularity

Metrics such as relevance metric or restrictiveness can help agencies to understand the trend of new regulations. However, this analysis alone does not offer any indication on how these regulations relate to the general public. In this context, we propose a novel metric, which is the normative act popularity. This new metric indicates how popular a normative act is for a specific group. Specifically, we targeted the general population with access to the internet and the federal government.

The population normative act popularity is calculated based on historical data from search queries containing normative acts on Google, the most used search engine. Specifically, the search strings used to calculate normative acts popularity were formed by the following elements: the act type followed by its number, a slash, and publication year. Examples of the search strings are "Law 8.112/1990" and "Constitutional Amendment 20/1998". In this context, we used information from the Google Trends engine alongside rules to search for normative acts and get their search frequencies, ranging from 0 to 100 for the most popular matching result.

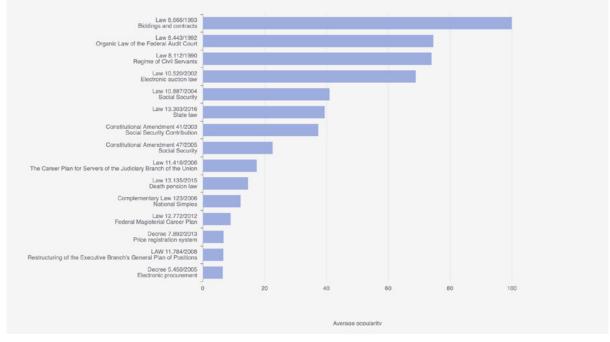**Figure 3 - Top 15 population normative acts popularity**

Source: InfoGOV

As a result, Figure 3 shows the average popularity of the 15 more popular normative acts. Not surprisingly, the most popular normative acts for the general public concern important social aspects and issues such as the COVID-19 pandemic, economic relief, the statute of small enterprises and labor reform.

To calculate the popularity metric in the context of the Federal Government, we employed a different approach. Instead of using Google Trends, we scrapped the data from the Official Gazette of the Federal Government (*Diário Oficial da União* - DOU). All contents of sections 2 and 3 from DOU were extracted since it was made digitally available in 2001. After this process, the text content is formatted and the frequency of each normative act is calculated and later normalized to obtain a metric ranging from 0 to 100.

Figure 4 - Top 15 government normative acts popularity

Source: InfoGOV

The results for the top 15 most popular normative acts on DOU are shown in Figure 4. The most popular normative acts tend to be related to important aspects of public administration. For instance, acts that regulate bidding processes and contracts, and an act that establishes the legal regime of public civil employees.

It is important to note that both popularity metrics are not comparable as they represent different information from distinct sources. Additionally, the data from Google Trends does not directly represent the number of citations of a search term. Instead, it represents a frequency calculation that considers factors such as geography and dates (Google, 2022).

3.3.4 Linguistic complexity

The final metric introduced is concerned with the linguistic complexity of each regulation. As regulations get more complex, there might be a need for more personnel to study them and they become less accessible to the general public understanding. Inspired by [7], we employ three different metrics to compare regulations' complexity.

The first metric is the median sentence length in a document, as longer sentences tend to be more challenging to understand, thus increasing their complexity. To avoid the effect of outliers in the metric, we used the median of this measure.

The second metric employed is Shannon's entropy (Cover & Thomas, 2005), that can be interpreted as measuring the frequency that new ideas (or words) are introduced in documents. Therefore, documents that employ simple and direct terms tend to have a lower entropy score than complex ones.

Finally, the last metric is related to branching words, i.e., conditional words in the text that connect clauses or ideas. Some examples of conditional words in English are: "if", "but", and "provided". We adapted the conditional terms to words that denote a similar branching idea[4] in Portuguese.

It is important to note that these complexity metrics can not be related directly as they incorporate different aspects of a document in terms of linguistic complexity. Thus, these metrics can present different behavior and correlation overtime. So, in order to assess if the regulation text from an economic sector is becoming more or less complex, one has to analyze the three metrics together.

As an example of their possible uncorrelation, imagine that the number of conditional words for a given sector increases. If the remaining aspects are kept equal, the entropy can decrease as the idea is being presented using fewer different words.
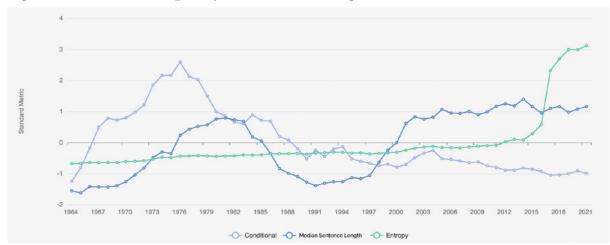
Figure 5 - Median of complexity metrics considering all sectors

Source: InfoGOV

---

4   We used the following words as conditional words: "se", "caso", "quando", "dado que", "desde que", "a menos que", "a não ser que", "embora", "ainda que", "mesmo que", "posto que" and "em que".

Consequently, the complexity analysis has to be conducted considering how each metric evolved over time. We present in Figure 5 the median of each metric over time across all sectors. The median is used to reduce the influence of outliers, and the results were standardized (zero mean and unit variance) in order to simplify the analysis. It follows that, based on the metrics, the language encountered in the regulations tends to get more complex overtime. It is also important to note that each sector has its own specificities, so an individual analysis of each one of them is necessary before concluding the analysis. RegBR allows the user to filter by sector all aforementioned regulatory metrics

In summary, it is essential to notice that these results are based on labels derived from a machine learning model that can (and most likely will) make mistakes, and on noisy metrics that are used as proxy to represent the linguistic complexity of a document. Therefore, the results presented in this section should not be used as facts, instead, they can help policymakers to gain insights on how regulations evolved over time and how they can be improved. Hence, the regulatory stock flux and the resultant metrics described above are available for consultation at https://infogov.enap.gov.br/regbr.

4. APPLICATIONS FOR THE FEDERAL GOVERNMENT

In this section, we discuss five of many applications that RegBR has on the federal government: RegBR as a regulatory base for studies; RegBR as a feedback mechanism to regulatory agencies; RegBR as a monitoring tool; RegBR as a comparative apparatus and RegBR as a transparency instrument. All of these applications are related to the usage of RegBR by citizens, organizations, and by the Federal Government itself.

Even though RegBR is a framework with all the features described in previous sections, the applications stated in this section bring another look to its usage and focus on how this tool can be useful in a practical way.

4.1 RegBR as a regulatory base to studies

Since RegBR is a framework that quantifies regulations produced by the Brazilian government across the years, this tool became a critical dataset that subsidizes regulatory studies. For instance, the actual call for studies that *Revista do Serviço Público* (Public Service Journal) opened, which this article is part of, uses RegBR as its data source.[5]

Regulatory impact analysis (RIA) inside the Brazilian Executive branch, especially in the regulatory agencies, is gaining ground and becoming a good practice before deliberating new regulations. Also, regulatory agencies are developing regulatory results assessments (RRA), that occur after the implementation of a regulation.

---

5   https://revista.enap.gov.br/index.php/RSP/announcement/view/40

One case of success in the public sector is the National Agency of Electric Energy (*Agência Nacional de Energia Elétrica - ANEEL*) which is adopting these practices of evaluating its regulations on the energy economic sector (Martins *et.al.*, 2022).

For both practices, RIA and RRA, RegBR can be a useful source of data for scholars that aim to better understand the impact of regulations in Brazil.

4.2   RegBR as a feedback mechanism to regulatory agencies

The four different metrics that are inside the RegBR framework, as described in Section 3, evaluate the quality of regulations produced by Brazilian regulatory agencies according to their: popularity; restrictiveness; influence on the economic sector; and linguistic complexity.

The last metric, linguistic complexity, is especially sensitive to the agencies because it is an indicator of transparency, and also of effectiveness. This metric is related to the capacity of citizens and regulated entities to understand the applications of the normatives. A normative with high linguistic complexity is less intelligible to its interlocutor. As a consequence, the effectiveness of this normative can be affected by the lack of understanding of what is being regulated.

Concerned about this fact, the National Agency of Waters (*Agência Nacional de* Águas - ANA) in 2021 used the RegBR metrics to evaluate the linguistic complexity of their normatives. After a meeting with the RegBR developers at Enap, the staff of ANA confirmed a feeling that they had about their normatives: the linguistic complexity of their regulations could be affecting the public understanding.

This fact was one example of how RegBR can be used, and probably is being used, by the regulatory agencies as a feedback mechanism.

4.3   RegBR as a monitoring tool

RegBR can also be used as a monitoring tool. The framework allows the Brazilian Federal Government to monitor its regulatory production, measuring lengthwise the number of acts that have been produced.

Regarding the task of monitoring the number of normatives that are being produced and have been produced, the federal government signed the Decree nº 10.139 from 2019 November (*Decreto* nº 10.139 de *Novembro de 2019*), that establishes that every organization in the Federal Government must revoke normatives that no longer had applicability. This initiative aimed to reduce the number of normatives in the legal system.

Since RegBR quantifies the number of normative and their legal status, this tool can be used by the Federal Government to analyze the impact of the Decree 10.139 and other initiatives that have similar goals.

Moreover, RegBR can assist decision makers in measuring their own work, providing a framework that allows the heads of regulatory agencies to measure what their organization produces in terms of volume and characteristics of regulations. The use of standard metrics can be interesting to have a clearer view of historical trends in the context of a specific regulatory agency or to compare different regulatory agencies in terms of their produced normatives characteristics.

4.4   RegBR as a comparative apparatus

RegBR is an initiative inspired by RegData from Mercatus Center. RegData, such as RegBR, quantifies the regulation produced by the U.S Federal Government, analyzing the data by the regulator and by industry. Also, RegData quantifies the regulatory restrictions, the same metric developed at RegBR with the name of restrictiveness.

Nowadays, updates in RegData are launched at "quantgov.org". This website brings information about other English-speaking countries, such as Canada and Australia.

With RegBR the Brazilian Federal Government can now compare its normatives production by industry, by the regulator, and by metrics such as regulatory restrictions, with distinct countries. RegBR made it possible for the Brazilian Public Administration to look at itself but also to look around and see how it is performing in comparison with other countries.

4.5   RegBR as a transparency instrument

Last but not least, RegBR achieves one proposal of Law nº 12.527, from 2011 November (*Lei nº 12.527 de Novembro de 2011*), known in Brazil as "Access to Information Law", which states that:

> "art. 3 The procedures provided for in this Law are intended to ensure the fundamental right of access to information and must be carried out in accordance with the basic principles of public administration and with the following rules:
> …
> IV - fostering the development of a culture of transparency in public administration;"

Based on the principles of "fostering the culture of transparency", RegBR gathers and centralizes federal regulations on a public database, with intuitive navigation and easy data visualization.

It is important to note that all the normative analyzed by RegBR had already been published in the Official Gazette of the Federal Government (DOU) and gathered as text on the government's official website[6]. Nevertheless, RegBR goes a step further given its an emphasis on accessibility. The tool transforms text into a data frame and subsequently displays the information from the data analysis in a simple format.

---

6   http://www4.planalto.gov.br/legislacao/

Probably the most important application of this framework is to ensure easy access to data and to foment active transparency, which means providing information to the public without being requested by a specific citizen.

5. CONCLUSIONS

This paper introduces RegBR, an active transparency framework that supports regulatory analysis and improves the legislative transparency and the understanding of the Brazilian regulatory process and its characteristics. This paper briefly introduces the framework and presents its main characteristics and the definition of the regulatory metrics used.

Then, we present a set of relevant applications of the framework in the Federal Government, such as: a regulatory base for studies; a feedback mechanism to regulatory agencies; a monitoring tool; a comparative apparatus, a transparency instrument. These applications indicate how the framework has been used and can be used in the future to support the concept of open government.

The authors expect to keep increasing the number of regulation metrics gathered by the project as well as increase the types of legislation analyzed, the scope of the framework and its impact on the Brazilian government. Moreover, the set of possible applications introduced here is far from complete, and we expect new approaches to arise over the next few years, to increase visibility and dissemination of the framework.

REFERENCES

Abu-Shanab, E. (2015). Reengineering the open government concept: An empirical support for a proposed model. *Government Information Quarterly*, 32, 453-463. https://doi.org/10.1016/j.giq.2015.07.002

Aletras, N., Tsarapatsanis, D., Preotiuc-Pietro, D, Lampos, V. (2016). Predicting judicial decisions of the european court of human rights: A natural language processing perspective. *PeerJ Computer Science*, 2(e.93). https://doi.org/10.7717/peerj-cs.93

Allio, M. K. (2012). Strategic dashboards: Designing and deploying them to improve implementation. *Strategy & Leadership*, Vol. 40 (5), 24-31. https://doi.org/10.1108/10878571211257159

Al-Ubaydli, O. & Mclaughlin, P.A. (2017). Regdata: A numerical database on industry-specific regulations for all united states industries and federal regulations, 1997–2012. *Regulation & Governance*, 109-123. https://doi.org/10.1111/rego.12107

Aronson, J., Brancalion, P.H.S., Durigan, G., Rodrigues, R.R., Engel, V.L., Tabarelli, M., Torezan, J.M.D., Gandolfi, S., De Melo A.C.G., Kageyama, P.Y., Marques, M.C.M., Nave, A.G., Martins,S.V. , Gandara, F.B., Reis, A., Barbosa, L.M., Fabio R. Scarano, F.R. (2011). What role should government regulation play in ecological restoration? *Restoration Ecology*, 19, 690-695. https://doi.org/10.1111/j.1526-100X.2011.00815.x

Banerjee, R. & Mio, H. (2018). The impact of liquidity regulation on banks. *Journal of Financial Intermediation*, 30-44. https://doi.org/10.1016/j.jfi.2017.05.008

Bertot, J. C., Jaeger, P. T. & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government information quarterly*, 264-271. https://doi.org/10.1016/j.giq.2010.03.001

Boella, G., Di Caro, L., Humphreys, L., Robaldo, L., Rossi, P. & Torre, L. (2011). Using classification to support legal knowledge engineers in the eunomos legal document management system. *Fifth International Workshop on Juris-Informatics*, 245-283. https://doi.org/10.1007/s10506-016-9184-3

Brown, A., Fishenden, J., Thompson, M. & Venters, W. (2017). Appraising the impact and role of platform models and Government as a Platform (GAAP) in UK Government public service reform: Towards a Platform Assessment Framework (PAF). *Government Information Quarterly*, 167-182. h https://doi.org/10.1016/j.giq.2017.03.003

Cary, C. (2012). *Measuring regulatory performance – Evaluating the impact of regulation and regulatory policy.* OECD Publishing.

Castro, M. (2014). New legal approaches to policy reform in Brazil. *University of Brasília Law Journal*, v. 1, n.01.

Chengalur-Smith, I.N., Ballou, D.P. & Pazer, H.L. (1999). The impact of data quality information on decision making: an exploratory analysis. *IEEE Transactions on Knowledge and Data Engineering*, 853-86. https://doi.org/10.1109/69.824597

Coates, D. & Heckelman, J.C. (2003). Interest groups and investment: a further test of the Olson Hypothesis. *Public Choice*, 333-340. https://doi.org/10.1023/B:PUCH.0000003737.58212.a7

Coffey, B., McLaughlin, P.A. & Tollison, R.D. (2012). Regulators and redskins. *Public Choice*, 153, 191-204. https://doi.org/10.1007/s11127-011-9781-3

Cover, T.M. & Thomas, J.A. (2005). *Elements of information theory*, Second Edition. John Wiley & Sons, Inc.

Crews, C. (2002). *Ten thousand commandments: An annual snapshot of the federal regulatory state.* Cato Institute.

Da Motta, R.S. & Moreira, A.R. (2006). Efficiency and regulation in the sanitation sector in Brazil. *Utilities Policy*, 14 (3), 185-195. https://doi.org/10.1016/j.jup.2006.03.002

Dredge, D., Gyimóthy, S., Birkbak, A., Jensen, T.E. & Madsen, A. (2016). The impact of regulatory approaches targeting collaborative economy in the tourism accommodation sector: Barcelona, Berlin, Amsterdam and Paris. *Impulse Paper*, n.9.

EGERT, B. & Wanner, I. (2016). *Regulations in services sectors and their impact on downstream industries. The OECD 2013 regimpact indicato*r. Tech. rep., OECD.

Fishenden, J. & Thompson, M. (2013). Digital government, open architecture, and innovation: why public sector it will never be the same again. *Journal of public administration research and theory*, 977-1004. https://doi.org/10.1093/jopart/mus022

Google. (2022). *FAQ about Google Trends data*. Available at: https://support.google.com/trends/answer/4365533?hl=en. Last access on 30 May 2022.

IGBE. (2007). Introduction to the national classification of economic activities (CNAE). *IBGE Repository*, v. 2.

IBM. (2018). *The future of compliance: how cognitive computing is transforming the banking industry*. Whitepaper.

Katz, D.M., Bommarito II, M.J. & Blackman, J. (2017). A general approach for predicting the behavior of the supreme court of the United states. *PLOS ONE*, 12, 1-18. https://doi.org/10.1371/journal.pone.0174698

Lodge, M., Van Stolk, C.V., Batistella-Machado, J., Schweppenstedde, D. & Stepanek M. (2017). *Regulation of logistics infrastructure in Brazil*. RAND.

Maheshwari, D. & Janssen, M. (2014). Dashboards for supporting organizational development: principles for the design and development of public sector performance dashboards. *Proceedings of the 8th international conference on theory and practice of electronic governance*, p. 178-185. https://doi.org/10.1145/2691195.2691224

Martins, A.R.S., Sanches, C.S. & Pinheiro, T.M.M. (2022). Iniciativas para a institucionalização do uso de evidências no processo regulatório na ANEEL - um estudo de caso de agência reguladora. In *Políticas públicas e o uso de evidência no Brasil: conceitos, métodos, contextos e práticas*, chapter 18. Brasília: IPEA.

Matheus, R., Janssen, M. & Maheshwari, D. (2020). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. *Government Information Quarterly*. https://doi.org/10.1016/j.giq.2018.01.006

Mcchesney, F. (1987). Rent extraction and rent creation in the economic theory of regulation. *The Journal of Legal Studies*,101-118. https://doi.org/10.1086/467825

Mcgovern, K. (2017). *Regulatory analytics: Keeping pace with the SEC: Five insights into compliance*. Delloite.

Mclaughlin, P.A., Sherouse, O. & Potts, J. (2019a). *Regdata: Australia*. Mercatus Center, Research Paper.

Mclaughlin, P.A., Strosko, s. & Jones, L. (2019b). *Regdata Canada: A snapshot of regulatory restrictions in Canada's provinces*. Mercatus Center, Research Paper, George Mason University, Regulatory Snapshot.

Mitchell, V.L. & Zmud, R. Z. (1999). The effects of coupling it and work process strategies in redesign projects. *Organization Science*, 424-438. https://doi.org/10.1287/orsc.10.4.424

Mulligan, Casey & Shleifer, Andrei. (2005). The extent of the market and the supply of regulation. *Quarterly Journal of Economics*, 120 (4), 1445-1473. https://doi.org/10.1162/003355305775097579

OECD. (2003). *Regulatory performance: ex-post evaluation of regulatory policies*. Proceedings from OECD expert meeting.

Orlikowski, W.J. & Barley, S.R. (2001). Technology and institutions: what can research on information technology and research on organizations learn from each other? *MIS quarterly*, V. 25 (2) 145-165. https://doi.org/10.2307/3250927

Parker, David C & Kirkpatrick, Colin. (2012). *Measuring regulatory performance, The economic impact of regulatory policy: A literature review of quantitative evidence*. OECD Publishing.

Peltzman, S. (1975). The effects of automobile safety regulation. *Journal of political Economy*, 677-725. https://doi.org/10.1016/0001-4575(76)90005-1

Pigou, A.C. (2013). *The economics of welfare*. Palgrave Macmillan,.

Leisen, R., Steffen, B. & Weber, C. (2019). Regulatory risk and the resilience of new sustainable business models in the energy sector. *Journal of Cleaner Production*, v. 219, 865-878. https://doi.org/10.1016/j.jclepro.2019.01.330

Robinson, D., Yu, H., Zeller, W.P. & Felten, E.W. (2008). *Government data and the invisible hand. Yale JL & Tech*, v 11, p. 160.

Rowley, C.K, Tollison, R.D. & Tullock, G. (1988). *The political economy of rent-seeking*. Springer Science & Business Media.

Shim, D.C. & Eom, T.H. (2008). E-government and anti-corruption: empirical analysis of international data. *International Journal of Public Administration*, 298-316. https://doi.org/10.1080/01900690701590553

Song, C. & Lee, J.. (2016). Citizens' use of social media in government, perceived transparency, and trust in government. *Public Performance & Management Review*, 39 (2) 430-453. https://doi.org/10.1080/15309576.2015.1108798

Stigler, G. (1971). The theory of economic regulation. *The Bell Journal of Economics and Management Science*, 2 (1) 3-21. https://doi.org/10.2307/3003160

Stiglitz, J.E. (1989). Markets, market failures, and development. *The American Economic Review*, 79 (2) 197-203.

Sulea, O., Zampieri, M. Vela, M. & Van Genabith, J. (2017). *Predicting the law area and decisions of french supreme court cases*. Proceedings of the International Conference Recent Advances in Natural Language Processing. RANLP, 716-722. https://doi.org/10.26615/978-954-452-049-6_092

Tuggener, D., Von Däniken, P., Peetz, T. & Cieliebak, M. (2020). Ledgar: *A large-scale multi-label corpus for text classification of legal provisions in contracts*. Proceedings of the 12th Language Resources and Evaluation Conference (LREC), 1228-1234. https://aclanthology.org/2020.lrec-1.155

Velcu-Laitinen, O. & Yigitbasioglu, O. (2012). The use of dashboards in performance management: Evidence from sales managers. *International Journal of Digital Accounting Research*, 12, 39-58. https://doi.org/10.4192/1577-8517-v12_2

Vitale, C., Bitetti, R.. Danitz, E., Moiso, C. & Wanner, I. (2020). *The 2018 edition of the oecd pmr indicators and database: Methodological improvements and policy insights*. OECD Publishing. https://dx.doi.org/10.1787/2cfb622f-en

Wongchaisuwat, P., Klabjan, D. & McGinnis, J.O.. (2017). *Predicting litigation likelihood and time to litigation for patents*. Proceedings of the 16th edition of the International Conference on Artificial Intelligence and Law, 257-260. https://doi.org/10.1145/3086512.3086545

Yuan, L.; Wang, J., Fan, S., Bian, Y., Yang, B., Wang, Y. & Wang, X. (2019). *Automatic legal judgment prediction via large amounts of criminal cases*. IEEE 5th International Conference on Computer and Communications, 2087-2091. https://doi.org./10.1109/ICCC47050.2019.9064408

Leticia Moreira Valle

https://orcid.org/0000-0001-9181-5974

PhD degree in Machine Learning at the Department of Electrical Engineering (UnB) and Data Science Manager, Universidade de Brasília (UnB). M.Sc in Electronic Systems with emphasis in Complex Systems Engineering, University of Bordeaux. B.Sc. degree in Communication Network Engineering, UnB. Electronic Engineering, ENSEIRB-MATMECA.

leticia.valle@redes.unb.br

Stefano Giacomazzi Dantas

https://orcid.org/0000-0002-6118-0143

M.Sc. in Electrical and Computer Engineering, McGill University. B.Sc. degree in Electrical Engineering, Universidade de Brasília (UnB). Researcher, UnB. Data Scientist, GSTS.

sgdantas1@gmail.com

Pedro Masson Sesconetto Souza

https://orcid.org/0000-0002-3191-4118

Master and Graduated in Political Science, Universidade de Brasília (UnB).

pedro.masson@enap.gov.br

Daniel Guerreiro e Silva

https://orcid.org/0000-0003-2858-1110

Ph.D. degrees in Electrical Engineering, M.Sc. and B.Sc. degree in Computer Engineering, Universidade de Campinas (Unicamp). Professor, Department of Electrical Engineering, University of Brasília (ENE/UnB).

danielgs@ene.unb.br

Ugo Silva Dias

https://orcid.org/0000-0003-0121-1523

B.Sc. degree in Electrical Engineering, Universidade Federal do Pará (UFPA). Ph.D. and M.Sc. degrees in Electrical Engineering (Unicamp). Professor, Department of Electrical Engineering, Universidade de Brasília (ENE/UnB).

ugodias@ieee.org