



Semi-automatic assessment of ICT Master Plan contents with Natural Language Processing technologies

Ethel Airton Capuano

PhD in Information Science from the Universidade of Brasília (UnB), Brasília, DF, Brazil.

University of Brasília (UnB), Brasília, Brazil.

Civil servant, Ministry of Integration and Regional Development, Brasília, Brazil.

<http://lattes.cnpq.br/1075438704907683>



Submitted on: 04/03/2022. Approved on: 01/25/2024. Published on: 06/21/2024.

ABSTRACT

This paper presents the results of a research on searching and semi-automatic assessment of interesting textual content on the Information and Communication Technology Master Plans published by a significant sample of Union's government agencies. Using natural language processing technologies, it was retrieved useful information and made interesting discoveries from the documents of the plans for assessment of the state-of-the-art on ICT in the agencies, some of which having caused surprise (due to serendipity), such as the demand management model in some units based on estimates of the teams' delivery capacity. It is concluded that the plans are in compliance with the official standards, including the observed interest of most agencies in current management issues, such as governance and risk management, and maturity in the assessment of market fads, but also is evident that its contents still lack approaches on some important ICT management practices to guide its technical evolution.

Keywords: natural language processing; text mining; semantic evaluation of ICT master plans.

INTRODUCTION

This paper presents the results of a research on textual search and semi-automatic assessment of technological contents of the Information and Communication Technology Master Plans (ICTMP) of Executive, Legislative and Judiciary Branch agencies, in addition to the Federal Prosecutor's Office, in force between 2013 and 2019, using Natural Language Processing (NLP) technologies.

What motivated this research was the idea of testing the possibility of automating processes for assessing the textual content of public sector documents with NLP technologies and a scientific curiosity concerning the technological contents addressed by agencies in their ICTMP, in a context of multidimensional and multidisciplinary assessment. Considering, as a research premise, that ICTMPs reflect, albeit with some limitations, the state-of-the-art of ICT in federal agencies, relevant questions that can be best understood based on the research results are: "How is the 'state-of-the-art' of ICT doing in the Union's agencies, in terms of internal organization (architecture)?"; "How are current ICT issues being addressed?"; "How are the most technical and structuring topics of ICT management being addressed?"; "Which agencies have the largest ICT budgets, which therefore require greater attention in terms of governance, risk management and controls?"; "How are the teams in the ICT units of the agencies and entities doing, as workforces?"

Although the state-of-the-art ICT for federal agencies has evolved since the advent (in Brazil) of the *World Wide Web* and the *Internet* in the second half of the 1990s, there is a certain consensus that this evolution has not occurred in the necessary pace for the desired advancement of public service digitalization projects, as can be seen in the revealed bottlenecks in some important services today (Batista; Azevedo; Hessel, 2020). While it is recognized, in some sectors, that continuous investments in digital services that are supported by robust ICT architectures have shown unquestionable success, as in the case of income tax declaration systems (which serve tens of millions of taxpayers by simultaneously accessing the automatic service channels) and in the case of courts of justice, which have digitized almost all of their procedural services, the investments of recent decades do not seem to show satisfactory results in other public services.

As examples of the pioneering spirit of courts of justice in innovations with intensive use of Information and Communication Technology (ICT), the Brazilian Federal Supreme Court (STF), in the second half of the 1990s, innovated the Brazilian public sector by implementing a *push system* (automatically "pushing" information to users on the world wide web) so that legal professionals could monitor the progress of their clients' cases remotely, via the *Internet*. Judiciary Branch agencies also innovated, even in the first decade of this millennium, by establishing entirely digital channels for exchanging procedural documents with the legal department of the Brazilian National Social Security Institute (INSS), given the materiality (in quantity and values) of the suits against that defendant autarchy. And, recently, the Regional Federal Appellate Court of the 4th Region (TRF-4) innovated by implementing,

with remarkable success, the first Electronic Document Management (EDM) system modeled, at source, for the Brazilian public sector – the Electronic Information System (EIS), which is also being progressively adopted by Federal Executive agencies.

The general objective of the research, therefore, focused on the practical testing of NLP technologies for a more substantive and less formal assessment of ICT planning in Union agencies and entities. In summary, this general objective consisted of verifying, with tools for automating text search processes, whether this government planning model could effectively contribute to the development of appropriate ICT solutions in accordance with market standards and best practices and the jurisprudence of control agencies. As specific objectives, evidence of relevant and current technical topics addressed by agencies and entities was sought in the texts of the selected plans.

RESEARCH METHODS, PROCESSES AND TECHNOLOGIES

The multidisciplinary research method was used, with qualitative (semantic) approaches related to the search, retrieval and processing of textual content (Coneglian; Segundo, 2023; Kao; Poteet, 2007; Meadow *et al.*, 2007; Konchady, 2006) , but also with quantitative approaches, employing textual statistics (Scott, 2019), with the aim of assessing the levels of interest of agencies in each research topic shown in the ICTMPs.

As areas of knowledge, research methodology can be framed, in theoretical terms (albeit partially), within Linguistics (Scott, 2019) and Information Science (Falcão; Lopes; Souza, 2022; Meadow *et al.*, 2007), the latter in a subject referred to as “Information Retrieval” These two areas of knowledge methodologically support the two main instrumental subjects used in this research: Natural Language Processing (*NLP*) and Textual Analysis (*Analytics*). This set of technologies also contributes to the composition of solutions in market concepts known as Text Mining and Data Science.

The textual information retrieved from the ICTMP texts refers both to attributes selected *a priori*, in the planning stage of the research project, and to *clusters* (compositions) of substantive and interesting words based on the most frequent keywords. In the search processes, both simpler technologies were used, such as electronic text editing software , as well as a more sophisticated market application software, for Natural Language Processing (NLP), which implements more advanced resources for retrieving the information, such as clusters, and statistics on texts in Portuguese (Scott, 2019). Discourse analysis was also used as a method of selecting relevant information about some attributes of interest in the ICTMPs collected.

The documents were retrieved from the agencies’ digital portals by searching the *World Wide Web* (Web), on the *Internet*. Excluding, in accordance with project decision, Federal Education Institutions (FEI) and government-controlled companies, in the Federal Executive Branch, the sample of ICTMPs surveyed, with 72 agencies in total, covers 17 of the 23 ministries in force as of Law No. 13.502, of November 1, 2017 (being it employed,

however, the ICTMP 2017-2019 of the new Ministry of Justice and Public Security and not the ICTMP of the Ministry of Justice and the Ministry of Public Security provided for in the last amendment to this Law), 42 of the 80 autarchies and foundations remaining after the exclusion of the FEI, 2 agencies of the Federal Legislative Branch, 10 agencies of the Federal Judiciary Branch and the Federal Prosecutor's Office.

The set of textual information generated from the collection of documents and the textual analysis of the ICTMPs was entered in an electronic spreadsheet for proper statistical analysis, and the contents were included according to attributes to be observed in the documents in columns and the corresponding agencies and separate entities according to a political-administrative categorization of Union agencies in lines: 3 Branches (sample with 72 agencies and entities), Executive Branch (59 agencies), Executive Branch – Direct Administration (17 agencies), Executive Branch – Indirect Administration (42 agencies and foundations), Legislative Branch (2 agencies), Judiciary Branch (10 agencies), Federal Prosecutor's Office (one agency).

In short, textual content on 37 research attributes was retrieved from each ICTMP, arranged in the spreadsheet according to the following analysis blocks and data formats:

- I. Basic Document Analysis (with 7 attributes): file volume (in MB); font and size of ICTMP letters (text); spacing between lines (actual number); number of pages (whole number); number of attachments (whole number); whether the version of the plan was new or an extension of a previous plan (binary data: "yes" or "no"); and planning time frame in years (whole number);
- II. Existence of Thematic Chapters (with 7 attributes) on: business and ICT strategies articulated in the ICTMP ("yes" or "no" for each strategy); business, information and technical (or systems) architecture ("yes" or "no" for each architecture); ICT project management methodology ("yes" or "no"); risk management ("yes" or "no"); governance ("yes" or "no");
- III. ICT needs (with 9 attributes): number of pages expressing ICT needs (or demands) (whole number); ICT needs items (whole number); occasional ICT project items (whole number); clarity in the presentation of demands ("yes" or "no"); technical analysis of ICT demands (before inclusion in ICTMP) ("yes" or "no"); records of redundant demands ("yes" or "no"); mention of Systems Development Methodology (SDM) ("yes" or "no"); mention of Business Intelligence (BI) solution ("yes" or "no"); use of methodology to prioritize meeting needs ("yes" or "no");
- IV. Previous ICTMP Performance (one attribute): percentage, revealing the executed portion of the planned needs and/or projects;
- V. ICTMP Estimated Budget (one attribute): monetary data relating to the forecast of expenses for implementing the plan within the expected time frame;
- VI. Declaration of Human Resources Gaps in the ICT Team (one attribute): binary data ("yes" or "no"), based on a declaration from the management unit; and

VII. Frequency of Words and Thematic Expressions (11 attributes): strategy; georeferencing; Information management; Knowledge management; Information security; GIS/SIG; Satellite Image; Drone/UAV; Cloud Computing ; Big Data; Mobile Application (App).

The statistical parameters calculated in the spreadsheet, for each political-administrative category of Union agencies and entities, were:

- a) average: for all attributes, but using numerical data adding the presence of “yes” in each ICTMP analyzed in the case of binary data;
- b) standard deviation: same; and
- c) variation coefficient (ratio between the standard deviation and the average): same.

Regarding their documentary structures, the researched plans of Executive Branch agencies follow, in most of their contents, the “Guia de Elaboração de PDTI do SISP” (Brasil, 2012), the SISP ICTMP Preparation Guide, in Portuguese, in a logical reasoning based on the principles and guidelines, strategies of ICT, work plan for the preparation of the ICTMP, SWOT Matrix, human resources assessment of the ICT team, prioritization method adopted, inventory of needs and *portfolio* of projects and actions, risk assessment, goal and action plan, and estimated budget.

LITERATURE REVIEW

Although the public familiar with ICT management in agencies do not find most of the concepts covered in this paper a novelty, some of them require definition, in order to avoid misinterpretations regarding the research context, and considering the institutional context of these organizations, compliance standards should also be highlighted.

Technical concepts

The first important concept is the “Information and Communication Technology Master Plan”, which is expressly defined in Decree 8.638/2016 as an “instrument for diagnosis, planning and management of information and communication technology resources and processes, with the purpose of meeting the final and informational needs of an agency or entity for a given period.”

The concept of “diagnosis” adopted in the reference guide (Brasil, 2012), however, deserves special attention, as it may not coincide exactly with what is understood by diagnosis in other contexts, as shown in the following text:

Após a fase de preparação, na qual foram organizadas as principais atividades do projeto de elaboração do PDTI, inicia-se a 2ª fase do processo: a fase de diagnóstico. Essa fase caracteriza-se por buscar compreender a situação atual da TI na organização para, em consonância com esse quadro, identificar as necessidades (problemas ou oportunidades) que se espera resolver. Para isso, contempla processos relacionados a análise do planejamento anterior, análise estratégica e levantamento de necessidades. (...) O levantamento de necessidades parte daquelas relacionadas à

informação e se desdobra em todas as outras associadas à TI: serviços, infraestrutura, contratações e pessoal de TI. (...) Um dos principais artefatos produzidos nessa fase, e muito importante para todo o processo de elaboração do PDTI, é o Inventário de Necessidades (Brasil, 2012, p. 33)¹.

Another primary concept, adopted as a motivation for the research, is the “technical adequacy” of the ICTMP, which should not necessarily be confused with the adequacy formally required in the governing standards, as these do not address in depth all topics relevant to ICT management in agencies, such as “enterprise architecture” (also referred to as “business architecture”), “information architecture” and “technical architecture” (or systems architecture).

The concepts related to “architecture”, on the other hand, when applied to the context of ICT management in organizations, are addressed in several reports and technical articles produced, for example, by company specialized in market research *Gartner Group*, well known by ICT managers, and other authors (Gartner Group, 2020a; Costelo, 2019; Hermans, 2015; Ross; Weill; Robertson, 2008). However, originally, the concept of “architecture”, in a broader context of business management with intensive use of computer information systems, is due to the pioneering research initiative known as *Business Systems Planning (BSP)*, implemented in the 1960s by several researchers working for company IBM (Kotusev, 2016, 2017), highlighting, subsequently, the business descriptive construct developed by Zachman (1987), better known as “Zachman Framework”, which is intended to describe and integrate, effectively and efficiently, the most relevant ontological elements that constitute a company (Nicolaidou *et al.*, 2009; Zachman, 2008).

The concept of “Enterprise Architecture” (*EA*), one of the management items researched in ICTMP, which in this context can be understood as “business architecture”, is defined as such by ICT market research company *Gartner Group*:

[...] uma disciplina para as empresas proativamente e holisticamente líderes responderem a forças disruptivas mediante a identificação e análise da execução de mudanças conforme a visão de negócios e os resultados desejados. A arquitetura empresarial entrega valor apresentando aos líderes de negócios e de TIC recomendações para o ajuste de políticas e projetos de modo a alcançar os resultados de negócios visados, que consolidam rupturas de negócios relevantes (Gartner Group, 2020a, *online*, tradução nossa)².

1 Translation: “After the preparation stage, in which the main activities of the ICTMP preparation project were organized, the 2nd stage of the process begins: diagnosis. This stage is characterized by seeking to understand the current IT situation in the organization in order to, in line with this framework, identify the needs (issues or opportunities) that are expected to be resolved. To this end, it includes processes related to the analysis of previous planning, strategic analysis and needs assessment. (...) Needs assessment begins with needs related to information and unfolds into all other needs associated with IT: services, infrastructure, hiring and IT personnel. (...) One of the main artifacts produced in this stage, and very important for the entire process of preparing the ICTMP, is the Inventory of Needs” (Brasil, 2012, p. 33, editorial translation).

2 Original: “[...] is a discipline for proactively and holistically leading enterprise responses to disruptive forces by identifying and analyzing the execution of change toward desired business vision and outcomes. EA delivers value by presenting business and IT leaders with signature-ready recommendations for adjusting policies and projects to achieve targeted business outcomes that capitalize on relevant business disruptions” (Gartner Group, 2020a, *online*).

In summary, in the research context, “business architecture” would address the identification, description, assessment and functional integration of the main resources of an agency, in a systemic and synergistic manner, to achieve its institutional goals. These resources can be classified as: human, material, heritage, technological, information and knowledge.

The more recent concepts of “information architecture” and “systems architecture”, which also refer to two thematic objects researched in the ICTMP, are defined by a larger group of other authors, with sometimes different meanings between the authors concerning the same topic. The pioneering author of the idea of “information architecture” is Wurman (1997), who defined it, in general terms, as an expansion of the occupation of the traditional architect applied to the organization of information. Lima-Marques and Macedo (2006), along the same lines, but with a *sui generis* philosophical approach, define this discipline as *a methodology for structuring information systems applied to any informational environment, which is understood as the space that integrates context, content and users*.

However, the most appropriate concept of “information architecture”, in the current research context, seems to be that of Hagedorn, according to Lima-Marques and Macedo (2006, p. 245): *art and science of organizing information to effectively help people to satisfy their information needs. It involves research, analysis, design and implementation*.

With regard to the concept of “technical architecture”, also referred to as “systems architecture”, it refers to the smart organization of information systems and the technologies that support them in organizations. This type of architecture would be the cognitive basis for using the organizations’ technological support components, integrating hardware components (workstations, servers, massive storage units, segmentation control devices in telecommunications networks, etc.) and basic software (such as operating systems and application servers), in addition to other fundamental software for computational operations, such as security applications, backups and load balancing.

This concept, didactically presented by Crawley (2007), is important to guide the processes of acquisition and/or development of new systems, integration between legacy systems, interoperability between components and systems, portability of legacy systems from one obsolete technological environment to a more current one, combination of legacy and new technologies, etc. The most promising model of “technical architecture”, also for federal agencies today, is that known as “Service-Oriented Architecture” (SOA), which incorporates logical and physical aspects, as defined by Bell (2008, p. 309):

A disciplina de arquitetura lógica orientada a serviço é principalmente voltada para o reuso, utilização e consumo. (...) essa disciplina tem como foco os três maiores aspectos funcionais de uma arquitetura lógica: (1) o relacionamento entre pacotes de softwares implementados, (2) as preocupações estruturais das entidades dos pacotes orientados a serviços, e (3) o comportamento e a colaboração das soluções empacotadas. A arquitetura física orientada a serviço se refere a tudo que é fisicamente

endereçável aos componentes orientados a serviço. Esse endereçamento consiste no mapeamento das entidades de *software* em componentes de hardware e nos seus locais endereçáveis numa rede de comunicação de dados³.

This construct of ICT organization is also addressed by Abrams and Schulte (2008), from company *Gartner Group*, the technical glossary of which defines it as follows:

[...] é um paradigma e uma disciplina de desenho que auxilia a TI no atendimento das demandas do negócio [...] SOA reduz a redundância e incrementa a usabilidade, manutenibilidade e valor. Isto resulta em sistemas interoperáveis e modulares, que são mais fáceis de usar e de manter. SOA cria sistemas mais simples e velozes, que incrementam a agilidade e reduzem o Custo Total de Propriedade (TCO) (Gartner Group, 2020b, *online*, tradução nossa)⁴.

Incidentally, it should be noted that this strategy based on information and ICT architectures seems to coincide with the current proposal for structural change in the service provision model of Federal Executive Branch agencies, with greater digital interaction between agencies and citizens. (Batista; Azevedo; Hessel, 2020)

However, in practice, one questions: “How should technical architecture be designed and implemented in organizations?” In response, it is argued that this architecture can be developed based on a primary design conceived according to the best technical thinking in ICT, editing internal standards in organizations for their compulsory use, which provides for certain standards and best practices, such as standards technologically defined for operating systems and relational database management systems (RDBMS).

As an example of a recurring situation where the advantages of a standardized technical architecture could avoid expenses to the treasury, hypothetically, is the acquisition or development of a new application outside the walls (by partner agencies, often), that requires, for it to be installed in the technological environment of the requesting agency, a database product not yet used or mastered by the ICT support team at that agency. As a result, there will be a need for the acquisition and technical mastery (of the database support team), in the short term, of a new software that is quite complex, thus contributing to making ICT environments even more costly, resulting in the overload of ICT management units with more tasks that could be avoided.

3 Translation: “The discipline of service-oriented logical architecture is mainly focused on reuse, use and consumption. (...) this discipline focuses on the three major functional aspects of a logical architecture: (1) the relationship between implemented software packages, (2) the structural concerns of service-oriented package entities, and (3) the behavior and collaboration of packaged solutions. Service-oriented physical architecture refers to everything that is physically addressable to service-oriented components. This addressing consists of mapping software entities into hardware components and their addressable locations in a data communication network” (Bell, 2008, p. 309, editorial translation).

4 Original: “[...] is a design paradigm and discipline that helps IT meet business demands [...] SOA reduces redundancy and increases usability, maintainability and value. This produces interoperable, modular systems that are easier to use and maintain. SOA creates simpler and faster systems that increase agility and reduce total cost of ownership (TCO)” (Gartner Group, 2020b, *online*).

Compliance standards

Federal Executive Branch agencies must carry out their ICT management activities in accordance with technical standards and best market practices, but also in accordance with the standards approved by the central agency of the Information Technology Resources Administration System (SISP, in Portuguese) and the extensive jurisprudence established by internal and external control agencies over the last two decades. The government management model by means of structuring systems, in the Federal Executive Branch, dates back to Decree-Law No. 200, of February 25, 1967, which established:

Art. 30. Serão organizadas sob a forma de sistema as atividades de pessoal, orçamento, estatística, administração financeira, contabilidade e auditoria, e serviços gerais, além de outras atividades auxiliares comuns a todos os órgãos da Administração que, a critério do Poder Executivo, necessitem de coordenação central. § 1º Os serviços incumbidos do exercício das atividades de que trata este artigo consideram-se integrados no sistema respectivo e ficam, conseqüentemente, sujeitos à orientação normativa, à supervisão técnica e à fiscalização específica do órgão central do sistema, sem prejuízo da subordinação ao órgão em cuja estrutura administrativa estiverem integrados. (...)

Art. 31. A estruturação dos sistemas de que trata o artigo 30 e a subordinação dos respectivos Órgãos Centrais serão estabelecidas em decreto (Brasil, 1967, *online*)⁵.

Originally, the SISP was established by Decree No. 1.048, of January 21, 1994, which was then referred to as the Information and Computer Science Resources Administration System. In its current version, Decree No. 7.579, of October 11, 2011 (which also refers to Decree-Law 200/1967), defined the way in which SISP is organized and, in its art. 2, established the purpose of systemic management, with the basic substantive functions that are most relevant to the assessment of ICTMP carried out in the research in question (*in litteris*):

IV – estimular o uso racional dos recursos de tecnologia da informação, no âmbito do Poder Executivo federal, visando à melhoria da qualidade e da produtividade do ciclo da informação; V – estimular o desenvolvimento, a padronização, a integração, a interoperabilidade, a normalização dos serviços de produção e disseminação de informações, de forma desconcentrada e descentralizada (Brasil, 2011, *online*)⁶.

5 Translation: “Art. 30. Personnel, budget, statistics, financial administration, accounting and auditing, and general services activities will be organized in a system form, in addition to other auxiliary activities common to all Administration agencies that, at the discretion of the Executive Branch, require central coordination. Paragraph 1 The services responsible for carrying out the activities referred to in this article are considered to be integrated into the respective system and are, consequently, subject to regulatory guidance, technical supervision and specific inspection by the central agency of the system, without prejudice to subordination to the agency into the administrative structure to which they are integrated.

(...) Art. 31. The structuring of the systems referred to in article 30 and the subordination of the corresponding Central Agencies will be established by decree” (Brasil, 1967, online, editorial translation)

6 Translation: “IV – encourage the rational use of information technology resources, within the scope of the federal Executive Branch, aiming to improve the quality and productivity of the information cycle; V – encourage the development, standardization, integration, interoperability, normalization of information production and dissemination services, in a deconcentrated and decentralized manner” (Brasil, 2011, online, editorial translation).

ANALYSIS AND DISCUSSION OF ICTMP CONTENT

Based on the electronic spreadsheet completed with the information collected from the documents and the calculated textual statistical parameters, the content analysis stage highlighted the aspects of the ICTMP presented below.

Execution performance

Performance data in the execution of previous plans, one of the information requirements for the diagnosis provided for in the SISP guide, is still a challenge for transparency in the ICT management of the agencies, noting that only 11 of the 17 agencies (64.7%) of the Direct Administration and 26 of the 42 agencies (61.9%) of the Indirect Administration of the Federal Executive Branch included these assessments in their ICTMPs. Although several agencies mention the existence of attached documents with this assessment item, they were often not found on the digital portals of the agencies.

According to **TABLE 1**, the average performance of Federal Executive Branch agencies in executing ICTMPs is only 44.3%, with a 0.4 variation coefficient, showing that more than half of the plan contents do not materialize or are only partially executed (the average of the Judiciary should not be considered because it represents data from only three non-random agencies). It is hoped that further research will be able to show the causes of this level of performance: whether the delivery expectations of the agencies' ICT unit are excessive, given their resources (thus requiring a greater sense of reality in the process of prioritizing needs); or the assessment of the time for implementation, due to the lack of detail of ICT needs in the planning stage, has been too optimistic; or a combination of these two causes.

TABLE 1 – Estimated delivery capacity calculation (hypothetical)

Available Human Resources				Time Consumption of Profiles per Activity (A_i) and Project (P_i) (man-hours)					
Profiles	Working Time (man hours))			A_1	A_2	A_3	A_4	P_1	P_2
	Day	Year	PDTIC						
Perfil A	8	2.000	6.000	750	1.500	750	750	1.500	750
Perfil B	16	4.000	12.000	3.000	5.500	1.500	250	1.250	500
Perfil C	16	4.000	12.000	1.000	5.500	1.500	2.000	1.000	1.000
Perfil D	24	6.000	18.000	0	3.000	0	0	8.000	7.000
Perfil E	24	6.000	18.000	1.000	3.000	500	500	7.000	6.000
Perfil F	6	1.500	4.500	1.000	2.000	250	500	500	250
Perfil G	12	3.000	9.000	750	2.000	3.000	750	2.000	500
Time of Team (mh)	106	26.500	79.500	7.500	22.500	7.500	4.750	21.250	16.000

Source: Prepared by the author (2022).

Delivery capacity as a prioritization criterion

In analyzing the discourse expressed by the authors of the texts, passages in which human resources gaps in ICT management units are mentioned were identified in 46 of the 72 (63.9%) ICTMPs researched, 45 of which were in Federal Executive Branch agencies (in 13 Direct Administration and 32 Indirect Administration agencies). In other words, gaps in human resources for ICT management are mentioned in more than 76.0% of Federal Executive Branch agencies, both in Direct Administration and Indirect Administration.

In particular, we highlight the initiative of some ICT management units to use a more detailed method of assessing their delivery capacity over the plan's time frame, based on a prior assessment of their human resources. The SISP Guide, despite providing for this assessment, does not propose any specific "end-to-end" methodology for the process, resulting in the assessments found in the ICTMP focusing, in general, only on available human resources, without human resource estimates necessary to implement the plans.

With this somewhat innovative (and courageous, one might add) approach, a greater sense of realism is adopted to prioritize meeting the needs of those who seek solutions. The construct (reasoning structure) presented in **TABLE 1**, with hypothetical data from any agency, supports the algorithm for calculating the delivery capacity of the ICT unit, with data on its workforce (in man-hours), and the consumption of this workforce to implement ICTMP in the agency. This calculation must counter a sum of the daily and annual net working hours

(man-hours) available for each team member (excluding hours committed to other activities in the unit) with a distribution of these hours in the various actions planned for carrying out the ICTMP in its time frame.

With implementation periods (horizons) of three years, for example, note that in the fourth column from left to right the sum of man-hours in the “ICTMP” column is equal to three times the annual sum of net hours available for each professional profile involved in the execution of the ICTMP ($79,500 / 26,500 = 3$).

With this calculation algorithm, detailed planning of the time consumption of each professional profile that makes up the ICT management team in the organization is carried out. It forces the ICT team to estimate, in advance, the total time consumption for each activity or project provided for in the ICTMP, within the planning horizon, and distribute this time with the corresponding available time of the professionals who will be involved in the execution. The most developed ICTMPs in this item, in a preliminary assessment, are those of the following agencies: Federal Prosecutor’s Office (MPF), National Council of Justice (CNJ), Brazilian Space Agency (AEB) and Private Insurance Superintendence (SUSEP).

This management innovation makes ICTMP more realistic, based on evidence, as recommended in Appellate Decision No. 2608/2018-TCU-Plenário (Brasil, 2018) in relation to public policies, plans and programs. Although this ICT team delivery capacity assessment model also presents risks of erroneous estimates, it will certainly be closer to the unit’s effective delivery capacity, as one of the applicable prioritization criteria.

Technical chapters of interest

The research project also provided for the textual verification of the possible presence of specific technical chapters of interest that would highlight the technical level of the ICTMP, whether or not provided for in the SISP Guide. The technical research topics were presented in **TABLE 2**: “Strategies”, “Architectures”, “PMM” (ICT Project Management Methodology), “Risk Management” and “Governance”.

Using the word search features of a word processing software, the presence or absence of chapters with certain contents of interest for the research was identified, with the titles indicated in **TABLE 2**. As a precaution, a search for these chapters in the bodies of the ICTMP texts was also carried out, considering that sometimes the summaries of the plans are not detailed in relation to the chapter contents.

TABLE 2 – Specific technical chapters

Union Branches and Entities	ICTMP Qty	Quantity of ICTMP with Technical Chapters				
		Strategies	Architectures	PMM (*)	Risk Management	Governance
Executive Branch	59	4	15	2	51	29
Direct Administration	17	2	2	1	15	7
Indirect Administration	42	2	13	1	36	22
Legislative Branch	2	1	1	0	0	1
Judiciary Branch	10	0	1	0	0	5
Federal Prosecutor's Office	1	0	0	0	0	1
Total	72	5	17	2	51	36

Source: Prepared by the author (2022).

(*) PMM: ICT Project Management Methodology

Topic “strategies” refers to express and detailed demonstrations of articulation between business and ICT strategies in the ICTMP itself, in order to highlight, technically, the integration and synergy between these strategies (this is not, therefore, just a mention to this arrangement). With the results presented in **TABLE 2**, one concludes that only 5 of the 72 agencies (6.9%) minimally demonstrate this synergy, being it noted that several agencies refer this vision of arrangement between strategies to other agency documents, which do not make up the ICTMP, such as strategic planning documents.

The data in the “architectures” column in **TABLE 2** consolidates the three architectures that are the subject of research in the ICTMP: corporate or business architecture, information architecture and technical or systems architecture. It is observed that only 17 of the 72 agencies (23.6%) elaborate on this topic in chapters in their ICTMP, given that the agencies of the Indirect Administration of the Executive Branch have a higher rate: 13/42, or 31.0%.

Perhaps the most frustrating result is that related to the presence of thematic chapters on ICT project development methodologies (column “PMM”): only 2 of the 72 agencies surveyed elaborated on the topic in a specific chapter in the ICTMP itself, even though several of them mention the existence of a methodology or the intention to adopt a methodology. With this result, an important methodological issue does not become transparent for technical scrutiny external to the agency, or for the exchange of experiences and learning between ICT management units and outside-the-wall specialists.

However, it should be considered a positive result that most ICTMPs present thematic chapters on two very current topics (under the terms of Decree No. 9.203, of November 22, 2017): “risk management” and “governance”. The Federal Executive Branch agencies include, in their ICTMP, a chapter on “risk management” in 51 of the cases (70.8%), with

better results for Direct Administration agencies, in 15 of the 17 agencies (88.2%). These rates decrease for topic “governance”, although they are significant and quite similar in all categories of agencies; in the Federal Executive Branch agencies, 29 out of 51 (or 56.9%) include chapters on “governance” in their ICTMPs.

Current topic quotes

Another block of semantic research objects focused on the search for textual quotes, not necessarily in specific chapters, on the following topics: “strategy”; “information security”; “GIS” (*Geographical Information System*); “SIG” (*Sistema de Informações Geográficas*); “georeferencing”; “satellite image”; “*Drone*”; “UAV” (*Unmanned Aerial Vehicle*); “mobile device application”, “*mobile*” (*mobile device application*); “*app*” (*software application*); “information management”; “knowledge management”; “*cloud computing*” and “*big data*”. **TABLE 3** presents the combined results of this search in each ICTMP, highlighting that topics related to georeferencing were consolidated into a single item for data analysis (also adding the frequencies of quotes “GIS”, “SIG”, “satellite image”, “drone” and “UAV”), as well as search results for “application”, “*mobile*” and “*app*” (consolidated in “Mobile Device Application”), given that quotes “information management” and “knowledge management” were combined on account of the natural connections between them.

TABLE 3 – Current topic quotes

Research Topics		Branches and Entities (*) and Observed Frequencies of Thematic Quotes						
#	Thematic Expression	EB	EB-DA	EB-IA	LB	JB	FPO	Subtotal
1	Strategy	1.629	514	1.115	13	137	36	1.815
2	Information security	1.240	371	869	5	77	15	1.337
3	GIS/SIG, Georeferencing, Satellite Image, Drone/UAV (**)	628	98	530	3	23	0	654
4	Mobile Device Application	234	162	72	6	18	3	261
5	Information and Knowledge Management	209	55	154	5	2	2	218
6	Cloud Computing	180	49	131	9	27	11	227
7	Big Data	16	7	9	0	4	2	22
Quantity of ICTMP (A)		59	17	42	2	10	1	72
Total Quotes Found (B)		4.136	1.256	2.880	41	288	69	4.534
Total Quotes / ICTMP (C=B/A)		70,1	73,9	68,6	20,5	28,8	69,0	63,0

Source: Prepared by the author (2022).

(*) EB: Executive Branch; EB-DA: Direct Administration of the Executive Branch; EB-IA: Indirect Administration of the Executive Branch; LB: Legislative Branch; JB: Judiciary Branch; FPO: Federal Prosecutor's Office.

(**) *Drone* (in English) or *UAV* (VANT in Portuguese): Unmanned aerial vehicle.

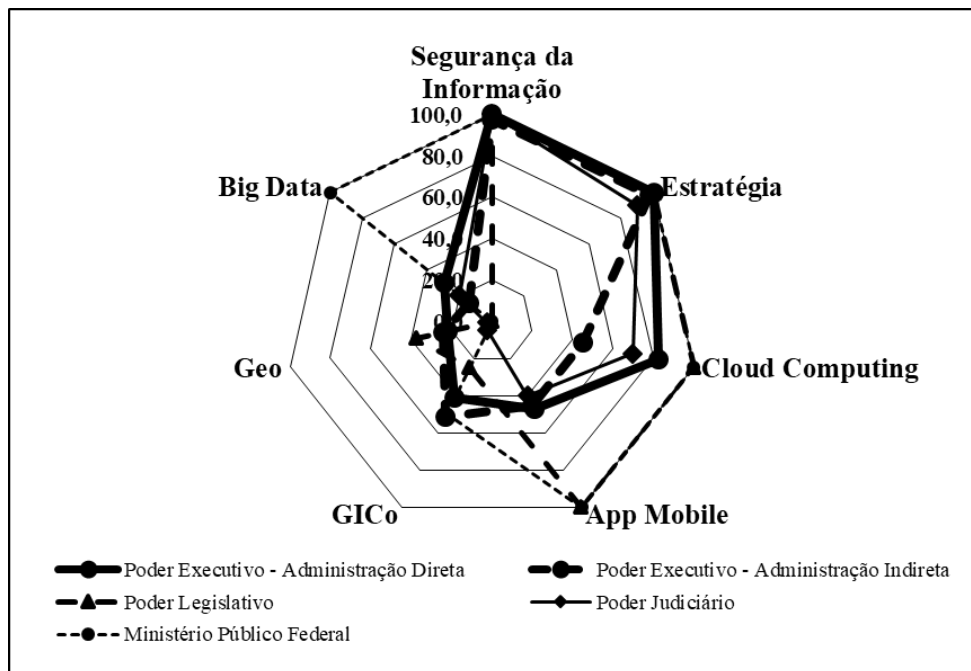
The most relevant aspects of the searches, observing the results in **TABLE 3**, are:

- the two most cited topics are, by far, “strategy” (with 1,815 quotes found) and “information security” (with 1,337 quotes);
- As for topics related to “georeferencing” and “mobile device application”, the first topic seems to be of more interest to Indirect Administration agencies than to Direct Administration Federal Executive Branch agencies, with frequency ratios of around 12.6 quotes per agency (530/42) and 5.8 quotes per agency (98/17), respectively, while the second topic seems to be of much greater interest to Direct Administration agencies, at a rate of 9.5 quotes per agency (162/17), against 1.7 quotes per agency (72/42) in Indirect Administration;
- although widely discussed, in recent years, topic “cloud computing” (does not appear very frequently in the plans of agencies in general and in the Federal Executive Branch, on average, only 3.1 quotes were found per ICTMP (180/59); even less so is topic “big data”, with only 16 quotes in total (16/59, or just one quote for almost four ICTMPs).

Considering the relevance of digital georeferencing in its context, the National Department of Mineral Production (DNPM) calls its plan the “Plano Diretor de Tecnologia da Informação, Comunicações e Geoprocessamento 2017-2019”(2017-2019 Master Plan for Information Technology, Communications and Geoprocessing).

In terms of the number of agencies in which quotes on this topic appear, in relation to the number of agencies with ICTMP researched (percentages of agencies with quotes), the “radar” style graph in **FIGURE 1** highlights the concentrations, with percentages ranging from 0 to 100, of ICTMP with thematic research quotes by Union Branch and Entity. There is a greater concentration of quotes on topics “strategy” and “information security” in Executive Branch Direct Administration agencies (thicker solid line) and lower in Indirect Administration agencies (thicker dotted line).

FIGURE 1 – Percentages of ICTMP with quotes on current topics



Source: Prepared by the author (2022).

Analysis of ICT needs

This block of research data refers to the demands for ICT solutions expressed in the ICTMP, which correspond to the item diagnosis referred to as “ICT needs” in the SISP Guide. In summary, in this cognitive logic, needs represent the most important indicator according to which the ICT management unit must direct its resources and, thus, contribute to the technological evolution of the organization.

TABLE 4 presents the quantities of items that appear as “ICT needs” in the plans of the surveyed agencies, grouped by political-administrative category. As provided for in the

SISP Guide, some agencies define items for the acquisition or development of goods and services as “ICT needs”, without a more detailed breakdown, while others detail their needs in projects and actions and, sometimes, in goals.

TABLE 4 – ICT needs

Union Branches and Entities	ICTMP Qty. (A)	Quantities of Items defined as “Needs”	
		Subtotal (B)	Average (C=B/A)
Executive Branch	59	3.982	67,5
Direct Administration	17	832	48,9
Indirect Administration	42	3.150	75,0
Legislative Branch (*)	2	133	66,5
Judiciary Branch (*)	10	863	86,3
Federal Prosecutor’s Office (**)	1	-	-
Total	72	4.978	69,1

Source: Prepared by the author (2022).

(*) Recorded as needs or projects.

(**) Demands are not consolidated in the same format as that of the other agencies.

In general, the Federal Executive Branch agencies express their ICT needs on 24 pages of the ICTMP, on average (with a 1.2 variation coefficient), while the agencies of the Judiciary Branch have a slightly lower average, around 19 pages, with a 0.9 variation coefficient. The total documentary volume recording the ICT demands of the ICTMP surveyed totaled 1,693 pages with 4,978 “needs” and 6,106 projects or actions with some detail.

It is important to note that the agencies of the Legislative and Judiciary Branches and the Federal Prosecutor’s Office (FPO) are not subordinate to the SISP, reason for which their ICTMP do not generally have the same structure and data presentation formats that those of the Federal Executive Branch agencies. In relation to the ICT demands observed, several agencies of the Legislative and Judiciary Branches and the Federal Prosecutor’s Office defined their demands directly as “projects”, not as “needs”.

With discourse analysis of the chapter texts on ICT needs, it was observed that the descriptions of the items are generally at very high levels of abstraction, with little clarity, from a technical point of view, about what is being demanded, or needs are described with objects so broad that they should certainly have been broken down into a number of smaller and more specific demands (needs or projects/actions). As examples of unclear inventoried needs, items were found described (literally) as: “needs assessment and deployment of resources to improve the availability of ICT infrastructure”, “improve the strategic decision-making process”, “implement knowledge management technologies in support of strategic

issues”, “develop computer-based information systems using standardized platforms, using defined and mature processes”, “increased productivity and delivery capacity”, “development of mobile applications”, “develop new systems”, “acquisition of products/systems to meet specific needs of business areas”, “integrated monitoring center”, “solution for human resources management”, “implement/acquire/customize computerized systems according to the stated needs after feasibility studies”, “improved quality of IT services provided by the agency to its users”, “improved building networks”, “increase use of market solutions”.

Another point of concern is that in the overwhelming majority of Federal Executive Branch agencies surveyed, apparently redundant needs are found in the same ICTMP, which could be consolidated by employing a more corporate and integrated approach to service, perhaps resulting from the lack of more in-depth technical analyzes of ICT needs presented by the various internal units of the agencies.

The data in **TABLE 5** shows the list of the 10 nominative words (nouns that generate denominations of objects, actions or ideas) most frequently found in the “ICT Needs” chapters of the ICTMPs researched, having excluded from this table words (acronyms) “IT” and “ICT”, which, due to its direct relationship with the general topic of the texts, is repeated thousands of times without adding anything new (relevant information). The third column from left to right shows the corresponding frequencies with which these words, individually, are found in texts, in absolute numbers (total count), and the fourth and last column, from left to right, shows the quantities of noun clusters (noun phrases) formed by combining each of the most frequent words observed (in the singular and plural).

TABLE 5 – 10 Most frequent words and formed clusters

#	Word	Word frequency	Cluster frequency
1	system/systems	3.360	2.049
2	service/services	1.856	1.824
3	information	1.840	1.482
4	Management	1.580	2.118
5	data	1.331	1.011
6	project/projects	1.330	680
7	process/processes	1.316	963
8	solution/solutions	1.241	953
9	acquisition/acquisitions	1.001	763
10	software	908	1.021

Source: Prepared by the author (2022).

The word clusters are identified in the texts and computed, in terms of frequency of occurrences, only using specific NLP software, such as that related to the work of Scott

(2019), with several other search textual statistics functions not found in common word processors. The linguistic structures in word clusters, in this case compositions of nouns with other nouns, are semantically richer than isolated nouns, as in the following example: “management” does not clarify much about the context of the management application, but “information management” identifies the management application area, presenting a more specific context, with greater explanatory power; the same occurs with noun “security”, which when added to “information” becomes “information security”.

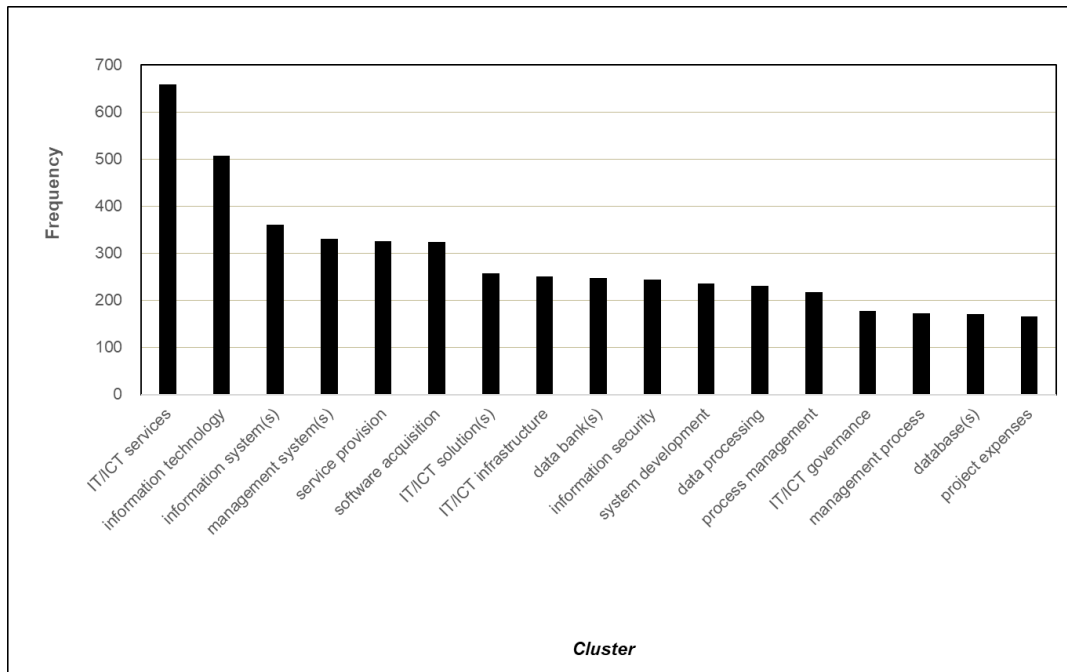
However, as a measure of saving computational research resources (text processing time), only the clusters with frequency equal to or greater than 10 were added, thus inferring that the frequencies of noun phrases formed with the 10 most frequent words (top ten) will be even more numerous. The most frequent *clusters* are composed with word “management”, totaling 2,118 appearances, which generally presupposes the management of something, therefore requiring a nominal complement, appearing in technical texts as “management of...”; otherwise, the composition of a cluster can also appear as “management process(/es)”, where “management” is the nominal complement of another noun (this composition of clusters appeared 173 times in the set of the ICTMPs surveyed).

The second structure of most frequent clusters is composed of word “system(s)”, with 2,049 appearances, and the third is composed of word “service(s)”, totaling 1,824 appearances. This number of clusters observed (with frequency equal to or greater than 10) related to the concept of “service(/s)” denotes a concern on the part of agencies with the provision of ICT-based services, a positive sign of the larger objectives of ICTMPs, even if this evolution does not occur at the desired pace.

The results of the search for the most frequent clusters in the ICTMP researched are shown in **CHART 1**. Nominative expression “ICT services” appears as the most frequent cluster, with 658 appearances, followed by “information technology”, with 506 occurrences, and “information system(/s)”, with 361 occurrences. It should be noted that of the 10 most frequent words presented in **TABLE 5**, only 2 do not make up clusters among the 10 most frequent: process(/es) and project(/s).

Considering that the term “solution”, from Normative Instruction No. 4/2010-SLTI/MPOG, has become a standard expression to refer to a set of ICT resources combined and intended to solve a business problem, we sought, in the ICTMPs, in the chapters on “ICT needs”, structured noun phrases (compound nominative expressions), such as “solution/solutions of...” and “solution/solutions for ...”, observing in the resulting phrases, the complementary nominative term in the sequence, which qualifies word “solution(s)”. With this textual search operation, the results presented in **CHART 2** were obtained, where “IT/ICT solution(s)”, “management solution(s)” and “security solution(s)” appear as the three most frequent clusters.

CHART 1 – Most frequent clusters among “ICT needs”

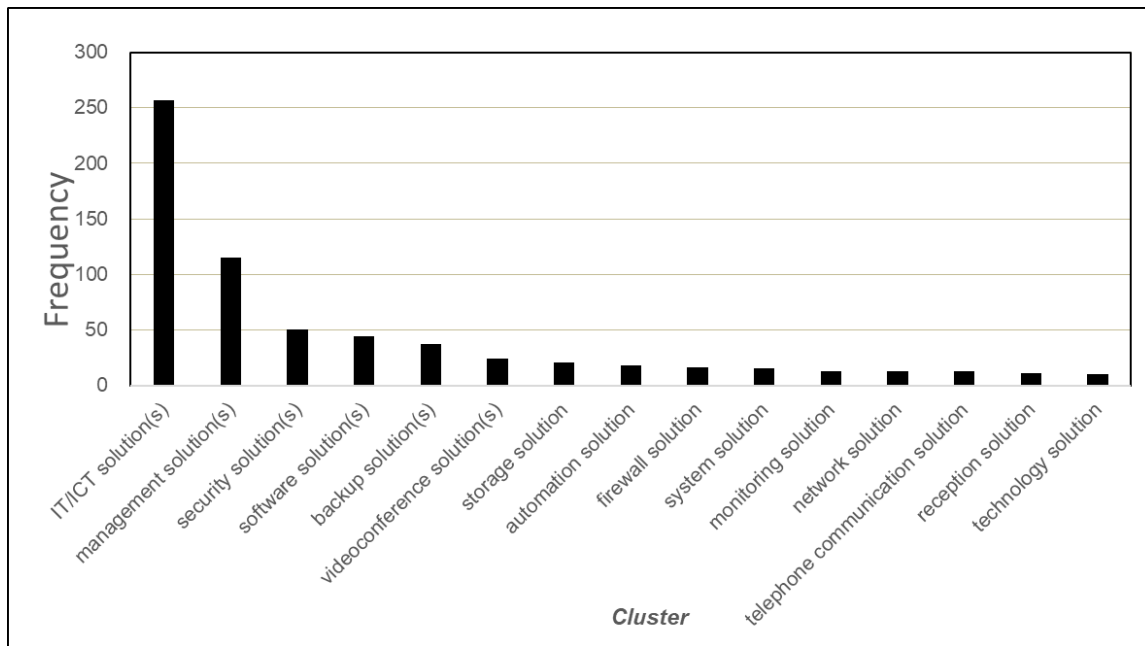


Source: Prepared by the author (2022).

Considering that the term “solution”, from Normative Instruction No. 4/2010-SLTI/MPOG, has become a standard expression to refer to a set of ICT resources combined and intended to solve a business problem, we sought, in the ICTMPs, in the chapters on “ICT needs”, structured noun phrases (compound nominative expressions), such as “solution/solutions of...” and “solution/solutions for ...”, observing in the resulting phrases, the complementary nominative term in the sequence, which qualifies word “solution(s)”. With this textual search operation, the results presented in **CHART 2** were obtained, where “IT/ICT solution(s)”, “management solution(s)” and “security solution(s)” appear as the three most frequent clusters .

These results corroborate the results presented in **FIGURE 1** regarding the high frequency of needs for “information security” solutions expressed in the ICTMPs researched, and it is also interesting to note the significant frequency of a need for a solution that is rarely commented on: “videoconference solution”. With regard to the other needs mentioned, such as “backup solution/solutions”, “network solution”, “storage solution” and “telephone communication solution”, for example, these are not new in the ICTMPs as they constitute common infrastructure needs to the agencies.

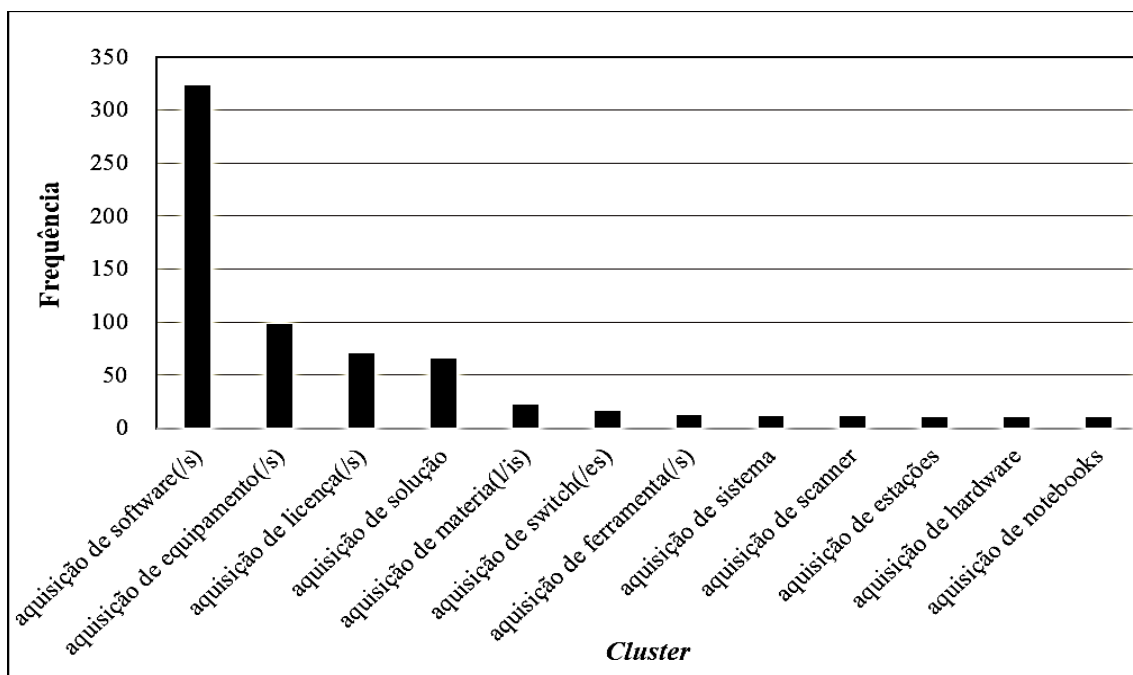
CHART 2 – Most frequent clusters with “solutions)”



Source: Prepared by the author (2022).

With **CHART 3**, the power of PLN technologies is again evident, including, among the 10 most frequent clusters (with frequencies equal to or greater than 10) observed among the “ICT Needs”, a series of interesting clusters composed of the word “acquisition”. These expressions reveal purchasing intentions for ICT components.

CHART 3 – Most frequent clusters with “acquisition”



Source: Prepared by the author (2022).

Another interesting aspect of this syntagmatic analysis refers to the clusters composed of word “management”: when adding the frequencies of the clusters with structures “management system(s)” and “management solutions”, more than 400 clusters contain this idea – that of solutions or computer information systems for managing agencies. It is concluded, given the textual evidence present in the ICTMPs researched, that there is at least a genuine concern on the part of Union agencies with the automation of their management processes.

With regard to ICT budgets, the plans researched revealed that the largest budgets belong to larger agencies in every sense, with decentralized regional units in the country and higher number of public servants, such as the Ministry of Health, Ministry of Finance, National Department of Transport Infrastructure, Ministry of Justice and Public Security and Ministry of Agriculture, Livestock and Supply. It was also observed that the budgetary issue is not addressed in all agencies of the Executive Branch, where only 49 of the 59 agencies surveyed (83.1%) had estimated their expenses.

As for values, the average budget amount for each ICTMP surveyed is R\$ 115.8 million over the average planning time frame (which is 2.8 years), resulting in an average annual expenditure per agency of approximately R\$ 41.3 million.

CONCLUSION

It is concluded, based on the sample size and the research results brought in this article, firstly that Natural Language Processing (NLP) technologies can effectively contribute to semi-automatic assessment processes of textual content of documents in the public sector, such as in the case of plans; and, second, that the ICTMPs published between 2013 and 2019 appear to be prepared in accordance with the SISP Guide and thus contribute to the technological evolution of the agencies, but they still lack significant technical improvements to become more effective instruments in this sense.

The research results show positive developments in current ICT management issues, with frequent mentions of governance and risk management topics, and significant interest in the adoption of disruptive technologies by some Federal Executive Branch agencies, such as digital georeferencing technologies and applications for mobile devices, although perhaps excessively cautious given the cost/benefit ratio of these technologies. Another positive aspect is that some more controversial ICT management models, even if adopted by companies in the market, seem to be considered within the scope of SISP in view of their complexity in an eventual scale implementation in the federal public sector, as in the case of cloud computing.

Another relevant aspect, which refers to a fortunate finding of this research, is that some ICT management units are daring to innovate in the demand management model, opposing, technically (with data), on the one hand the ICT needs and, on the other hand, estimates of the actual delivery capacity of their teams. With this, ICT management in agencies could become a more data and evidence-based activity, therefore more realistic, in accordance with the recommendations of the Federal Audit Court in Ruling 1603/2008-TCU-Plenário

(Brasil, 2008), on the needs for human resources in the agencies' ICT units, and in Ruling 2608/2018-TCU-Plenário (Brasil, 2018), on technical criteria for the development of public policies, plans and programs.

Statistical data on textual content also reveals, on the other hand, according to the perception of a recently published government authority, that in general, the Federal Executive Branch agencies still lack an evolution in ICT “from the door in” (Batista ; Azevedo; Hessel, 2020), to resolve bottlenecks in technological infrastructure and integration of computer information systems that prevent full service to citizens. The level of detail of the demands included in the ICTMP, generally superficial (and with redundancies), also does not contribute to the assessment of service efforts and better management of ICT needs in the agencies.

The little interest revealed in the elaboration of information architecture and technical architecture may be contributing to the slow progress of SISP agencies in meeting the technical requirements for integration, interoperability and portability of systems provided for in Ordinance No. 92/SLTI/MPOG, of December 24, 2014 (which established the “ePING” interoperability architecture), as well as the institutional objectives proposed in Decree No. 10.046, of October 9, 2019 (on governance in data sharing processes). Considering a natural analogy between ICT architecture and traditional architecture (in civil construction), developing a computing environment without information architecture and technical architecture is equivalent to constructing a building without an architectural design, assuming the risks of inadequacy of spaces and inappropriate use of costly ICT resources.

Obviously, one cannot ignore adverse aspects that go beyond the governability of the ICT units in the agencies, such as issues with inadequate corporate business strategies and processes and inadequate or absent governance models. The most obvious example refers to the hierarchical level at which the holders of ICT management units are located in the agencies, the overwhelming majority of whom are located in intermediate positions, which demonstrates, in practice, the level of priority for ICT management.

REFERENCES

ABRAMS, C.; SCHULTE, R. W. **Service-oriented architecture overview and guide to SOA research**. [s. l.]: Gartner Group, 2008.

BATISTA, V.; AZEVEDO, A.; HESSEL, R. Força-tarefa contra filas no INSS ganha reforço. **Correio Braziliense**, Brasília, 17 jan. 2020. Economia, p. 7.

BELL, M. **Service-oriented modeling: service analysis, design, and architecture**. New Jersey: Wiley, 2008. ISBN 0470141115.

BRASIL. **Decreto-lei nº 200**, de 25 de fevereiro de 1967. Dispõe sobre a organização da Administração Federal, estabelece diretrizes para a Reforma Administrativa, e dá outras providências. Presidência da República, Brasília, 1967.

BRASIL. **Decreto-lei nº 7.579**, de 11 de outubro de 2011. Dispõe sobre o Sistema de Administração dos Recursos de Tecnologia da Informação–SISP, do Poder Executivo federal. Presidência da República, Brasília, 2011.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. **Guia de elaboração de PDTI do SISP: Versão 1.0**. Brasília: MPOG, 2012. Disponível em: http://www.sisp.gov.br/.../Guia_de_Elaboração_de_PDTI_v1.0...pdf. Acesso em: 2 jan. 2020.

BRASIL. Tribunal de Contas da União. **Acórdão nº 2608/2018-TCU-Plenário**. Lei 13.707/2018, art. 124. Relatório de políticas públicas elaborado com objetivo de subsidiar o Congresso Nacional na elaboração da Lei Orçamentária. Acórdão 2127/2017-Plenário. Monitoramento. Adoção de diversas medidas, por parte do Poder Executivo, para atender às recomendações. Oportunidade de melhorias. Recomendações aos órgãos mencionados e a unidades deste Tribunal. Brasília: Tribunal de Contas da União, 2018. Disponível em: <https://pesquisa.apps.tcu.gov.br/#/documento/acordao->. Acesso em: 25 out. 2020.

BRASIL. Tribunal de Contas da União. **Acórdão nº 1603/2008-TCU-Plenário**. Levantamento de auditoria. Situação da governança de Tecnologia da Informação – TI na Administração Pública Federal. Ausência de Planejamento Estratégico Institucional. Deficiência na estrutura de pessoal. Tratamento inadequado à confidencialidade, integridade e disponibilidade das informações. Recomendações. Brasília: Tribunal de Contas da União, 2008. Disponível em: <https://pesquisa.apps.tcu.gov.br/#/resultado/acordao-completo/>. Acesso em: 25 out. 2020.

CONGLIAN, C. S.; SEGUNDO, J. E. S. Inteligência artificial e ferramentas da *web* semântica aplicadas a recuperação da informação: um modelo conceitual com foco na linguagem natural. **Informação e Informação**, Londrina, v. 27, n. 1, p. 625–651, jan./mar. 2022. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/44729>. Acesso em: 9 dez. 2023.

COSTELO, K. **The Evolution of Enterprise Architecture**. Gartner Group, 2019. Disponível em: <https://www.gartner.com/smarterwithgartner/the-evolution-of-enterprise-architecture/>. Acesso em: 2 jan. 2020.

CRAWLEY, E. **Introduction to system architecture**: architecture to value. Lecture Notes. Massachusetts: MIT, 2007. Disponível em: <https://ocw.mit.edu/courses/engineering-systems-division/esd-34-system-architecture-january-iap-2007/lecture-notes/lec1.pdf>. Acesso em: 21 jan. 2020.

FALCÃO, L. C. J.; LOPES, B.; SOUZA, R. R. Absorção das tarefas de processamento de Linguagem Natural (NLP) pela Ciência da Informação (CI): uma revisão da literatura para tangibilização do uso de NLP pela CI. **Em Questão**, Porto Alegre, v. 28, n. 1, p. 13-34, jan./mar. 2022. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/111323/64623>. Acesso em: 9 dez. 2023.

GARTNER GROUP. **Gartner glossary**: enterprise architecture. [s. l.]: Gartner Group, 2020a. Disponível em: <https://www.gartner.com/en/information-technology/glossary/enterprise-architecture-ea>. Acesso em: 2 jan. 2020.

GARTNER GROUP. **Gartner Glossary**: Service-oriented Architecture (SOA). [s. l.]: Gartner Group, 2020b. Disponível em: <https://www.gartner.com/en/information-technology/glossary/service-oriented-architecture-soa>. Acesso em: 22 jan. 2020.

HERMANS, P. **The Zachman Framework for architecture revisited**: on conceiving the informational enterprise. *Semiosis*, [s. l.], p. 1-18, 2015.

KAO, A.; POTEET, S. R. (ed.). **Natural language processing and text mining**. London: Springer, 2007.

KONCHADY, M. **Text mining application programming**. Massachusetts: Charles River Media, 2006.

KOTUSEV, S. Enterprise architecture: what did we study? **International Journal of Cooperative Information Systems**, [s. l.], v. 26, n. 4, 2017. Disponível em: https://www.academia.edu/36770941/Enterprise_Architecture_What_Did_We_Study. Acesso em: 24 out. 2020.

KOTUSEV, S. The history of enterprise architecture: an evidence-based review. **Journal of Enterprise Architecture**, [s. l.], v. 12, n. 1, p. 29-27. 2016. Disponível em: https://www.researchgate.net/publication/308936998_The_History_of_Enterprise_Architecture_An_Evidence-Based_Review/link/.../download. Acesso em: 24 out. 2020.

LIMA-MARQUES, M.; MACEDO, F. L. O. Arquitetura da Informação: base para a gestão do conhecimento. *In*: TARAPANOFF, K. (org.). **Inteligência, Informação e Conhecimento**. Brasília: UNESCO: IBICT, 2006, p. 241-255.

MEADOW, C. T.; BOYCE, B. R.; KRAFT, D. H.; BARRY, C. **Text information retrieval systems**. 3. ed. [s. l.]: Elsevier, 2007.

NICOLAIDOU, M.; TSADIMAS, A.; ALEXOPOULOU, N.; ANAGNOSTOPOULOS, D. Employing Zachman Enterprise Architecture Framework to Systematically Perform Model-Based System Engineering Activities. *In: PROCEEDINGS OF THE 42ND HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES 2009, 42.*, 2009, Havaí. **Anais** [...] Hawaii: HICSS, 2009.

ROSS, J. W.; WEILL, P.; ROBERTSON, D. C. **Arquitetura de TI como estratégia empresarial**. Tradução Roger Maioli dos Santos. São Paulo: M. Books, 2008.

SCOTT, M. **WordSmith Tools Manual**. Version 7.0. Stroud: Lexical Analysis Software Ltd., 2019.

WURMAN, R. S. **Information architects**. New York: Graphis, 1997.

ZACHMAN, J. A. **The concise definition of the Zachman Framework**. [s. l.], Zachman International, 2008.

ZACHMAN, J. A. A framework for information systems architecture. **IBM Systems Journal**, [s. l.], v. 26, n. 3, 1987, p. 276-292.