

CIÊNCIA DE DADOS PARA TOMADA DE DECISÃO NO CONTEXTO DA ORGANIZAÇÃO PÚBLICA: CASO EMPÍRICO NA CONTROLADORIA GERAL DO ESTADO DE SANTA CATARINA

DATA SCIENCE FOR DECISION MAKING IN THE CONTEXT OF PUBLIC ORGANIZATION: AN EMPIRICAL CASE IN THE COMPTROLLER GENERAL'S OFFICE OF THE STATE OF SANTA CATARINA

*Vander de Oliveira Veras¹
Roberto Fabiano Fernandes²
Viviane Brandão Miguez³
Edson Rosa Gomes da Silva⁴
Helio Aisenberg Ferenhof⁵
Alex Croisfelt Lautenschlager⁶
Andrei Francisco Fernandes⁷*

¹ Doutorando em Administração (UDESC), Graduado em Administração (UDESC) – vander_veras@yahoo.com.br

² Doutor em Engenharia e Gestão do Conhecimento (UFSC), Graduado em Ciência da Computação (FURB) – robertofabiano.fernandes@gmail.com

³ Doutora em Engenharia da Produção (UFSC), Graduada em Sistemas da Informação (UNISUL) – vivinemiguez@gmail.com

⁴ Doutor em Engenharia e Gestão do Conhecimento (UFSC), Graduado em Economia (UFSC) e Direito (IESGF) – edsoneconomia@gmail.com

⁵ Doutor em Engenharia da Produção (UFSC), Graduado em Ciência da Computação (Estácio de Sá) – hferenhof@gmail.com

⁶ Graduado em Tecnologia da Informação (UNICESUMAR) – alexciviluem@gmail.com

⁷ Mestre em Administração (UDESC), graduado em Administração (UFSC) – andreipmsc@gmail.com

RESUMO:

Para que uma organização tome decisões é fundamental ter acesso a dados e informações valiosas. Perante o desafio que se apresenta, questiona-se, a Ciência de Dados pode ajudar a aumentar e qualificar o processo com base no conhecimento de uma organização, visando apoiar a tomada de decisão? Para responder a essa questão, uma pesquisa-ação foi realizada na Controladoria-Geral do Estado de Santa Catarina (CGE-SC), onde técnicas de Ciência de Dados e Automação Robótica de Processos (RPA) foram aplicadas para alimentar os bancos de dados de forma qualificada. Considera-se, portanto, que o conjunto de métodos, práticas e automação de processos utilizados auxiliam na extração de conhecimento de um grande volume de dados para a tomada de decisão.

Palavras-chave: Ciência de Dados; Mapeamento de Processos Intensivos em Conhecimento; Web Scraping; Business Intelligence na Gestão Pública; Automatização de Processos.

ABSTRACT:

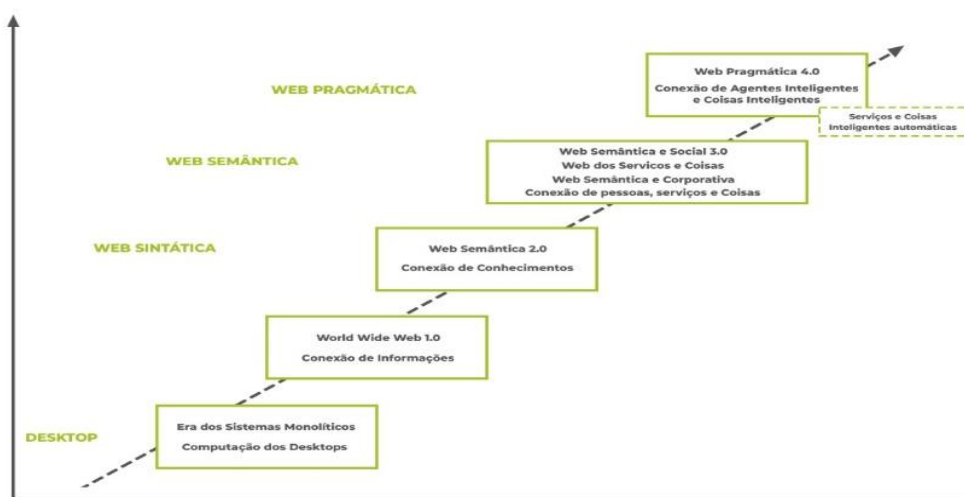
For an organization to make decisions, it is essential to have access to valuable data and information. Given this challenge, the question is, can Data Science help to increase and qualify the process based on the knowledge of an organization, aiming to support decision making? To answer this question, an action-research was conducted at the Office of the Comptroller General of the State of Santa Catarina (CGE-SC), where Data Science and Robotic Process Automation (RPA) techniques were applied to feed the databases in a qualified way. It is considered, therefore, that the set of methods, practices and process automation used assist in extracting knowledge from a large volume of data for decision making.

Keywords: Data Science; Knowledge Intensive Process Mapping; Web scraping; Business Intelligence in Public Management; Process Automation.

1 INTRODUÇÃO

A complexidade advinda, principalmente, pelo uso massivo das tecnologias da informação viabilizou o crescimento expressivo da produção de dados e informações diversificadas na publicação de vídeos, fotos, comentários, rotinas, fatos e opiniões, por parte dos usuários. Além disso, o próprio avanço da web (vide “Figura 1”) fez com que houvesse a produção automática de conteúdos realizados por algoritmos que efetuam a associação de elementos e pelo rastreamento de tudo que as pessoas postam na internet, até mesmo assuntos que a princípio não estariam relacionados (Souza e Alvarenga, 2004).

Figura 1. Evolução da Web



Fonte: Adaptado de Ferreira (2019).

Segundo Wang, Chen e Liu (2015), em um ambiente envolto pela geração de uma grande massa de dados e informações, também se encontram os agentes públicos que vêm sendo constantemente convocados na tomada de decisões, e para isso se amparam em dados ou informações oriundas de repositórios de dados, na maioria das vezes, estruturados.

Contudo, para que uma organização pública ou privada tome decisões rápidas e corretas, é relevante o acesso a dados e informações fidedignas e mais confiáveis, ou seja informações úteis. Obter informações consideradas úteis, que agreguem valor para a organização ainda é um desafio. Com esse entendimento, percebe-se a necessidade latente de se ter formas de enriquecer as bases de dados existentes, com dados e informações complementares advindas de diferentes fontes de dados, além da web, de forma que proporcionem maior detalhamento e valor estratégico para auxiliar a tomada de decisão, o que cume no aumento de uma base de conhecimento qualificada em uma organização.

Diante do desafio apresentado se indaga, a Ciência de Dados pode ajudar no processo aumentar e qualificar a base de conhecimento de uma organização, visando suportar a tomada de decisão? Objetivando responder tal questionamento, foi conduzida uma pesquisa-ação junto a Controladoria Geral do Estado de Santa Catarina (CGE-SC), onde técnicas de Ciência de Dados e automação robótica de processos (RPA) foram aplicadas com o intuito de alimentar as bases de dados de forma qualificada e que esta possa auxiliar no processo de tomada de decisão.

Como parte do cenário desta pesquisa, se destaca que a Coordenadoria de Informações Estratégicas – CIES, da Controladoria Geral do Estado de Santa Catarina – CGE encontrava-se em um

momento oportuno pois estava executando uma iniciativa de mapeamento dos processos da CIES, tendo como intuito a padronização, monitoramento e melhoria nos processos de negócio para um gerenciamento consciente dos seus processos e recursos e, ao aplicar técnicas de modelagem de processos, processos intensivos em conhecimento que auxiliaram na compreensão e documentação dos processos de desenvolvimento da base de dados governamental (*Business Intelligence* da CGE), cuja arquitetura está representada na “Figura 2”, também foram mapeados e serviram de base e guia para a necessidade de quais dados e informações precisavam de enriquecimento embasando assim, a construção de relatórios técnicos e gerenciais e de painéis que apresentam os principais indicadores da CGE-SC. As demais sessões apresentam a fundamentação teórica para sustentar essa pesquisa, a pesquisa-ação em si, os resultados e discussões e na sequência as considerações acerca dela.

Figura 2. Arquitetura do *Business Intelligence* da Controladoria Geral do Estado de SC



Fonte: Dados da Pesquisa

2 FUNDAMENTAÇÃO TEÓRICA

Serão apresentados os assuntos basilares que tratam das explicitações sobre as organizações intensivas em conhecimento e da automatização dos processos intensivos em conhecimento, nesse último destacando o papel das fases do ciclo de vida da Ciência de Dados como auxílio aos gestores em suas Tarefas Intensivas em Conhecimento e o uso da Automação Robótica de Processos (RPA).

2.1 Organização Intensiva em Conhecimento

Inicia-se a fundamentação teórica tratando das Organizações Intensivas em Conhecimento descrevendo-as como empresas em que o conhecimento se estabelece como um elemento imprescindível para o desenvolvimento das atividades da organização (Neves, et al., 2014), que empregam o conhecimento das pessoas para se desenvolver e apresentam soluções imateriais para os clientes (Jemielniak e Kociatkiewicz, 2009). Destacam-se como principais características para este tipo de Organização, primeiramente, o conhecimento aplicado à gestão da cadeia produtiva, os colaboradores com o seu conhecimento e as práticas de gestão do conhecimento (Bortoluzzi, Genari e Macke, 2018).

Segundo Mohanta (2013), as Organizações Intensivas em Conhecimento são empresas onde a maior parte da força de trabalho é formada por colaboradores bem formados e qualificados, e onde a maioria dos trabalhos têm uma natureza intelectual. Nesse sentido, percebe-se que os trabalhadores destas organizações possuem habilidades cognitivas elevadas e são altamente responsáveis por suas decisões.

Estas organizações são também qualificadas como corporações proativas, estratégicas e, apesar de possuir processos de grande complexidade e intensivos em conhecimento, são de fácil adaptação à mudança.

Neste sentido, entende-se que as organizações intensivas em conhecimento são as que utilizam fortemente ativos intangíveis, tal como o conhecimento, na execução de suas atividades. Estas atividades são centradas em pessoas e dependentes da experiência e habilidades dos seus executores (os trabalhadores do conhecimento), desta forma, são processos de difícil automatização. Apesar da dificuldade de automatização destas atividades intensivas em conhecimento, ela é necessária, pois possibilita alocar esses profissionais em setores mais estratégicos, favorecendo uma gestão mais eficiente e segura.

2.2 Automatização de Processos Intensivos em Conhecimento

Como já definido, uma Organização Intensiva em Conhecimento possui processos de alta complexidade, com isso, entende-se processo como “uma agregação de atividades e comportamentos executados por humanos ou máquinas para alcançar um ou mais resultados” (ABPMP, 2013, p.35), e que atividades intensivas em conhecimento são aquelas em que o conhecimento exerce um papel-chave (Davis, 1991). Assim, infere-se que os processos intensivos em conhecimento podem ser um encadeamento de atividades intensivas em conhecimento as quais são geralmente complexas, difíceis de

automatizar, envolvem dados estruturados, semiestruturados e não estruturados.

De forma complementar, no CBOOK v.3 (2013), há a citação que “é importante que os processos intensivos em conhecimento sejam corretamente identificados e tratados com técnicas adequadas para que a transformação não resulte mais em danos do que benefícios”. Com esse entendimento, a sua representação gráfica facilita a compreensão das atividades complexas e, assim, a identificação de tarefas que podem ser automatizadas por técnicas adequadas.

Segundo Freitas Junior e Gonçalves (2014) e Koenderink (2010), as tarefas intensivas em conhecimento necessitam de conhecimento especializado para a sua execução. Para isso, elas devem ser executadas por especialistas devido a sua ampla experiência na execução deste tipo de tarefa.

As tarefas intensivas em conhecimento são integrantes dos processos intensivos em conhecimento, os quais são definidos por Di Ciccio, Marella e Russo (2015) como aqueles processos cuja conduta e execução dependem fortemente dos trabalhadores do conhecimento que executam várias tarefas de tomada de decisão interligadas e intensivas em conhecimento. Esses processos são genuinamente centrados no conhecimento, informação e dados, os quais requerem uma flexibilidade substancial no tempo de concepção e execução.

Nesse contexto, a Ciência de Dados auxilia os gestores em suas Tarefas Intensivas em Conhecimento (Schreiber et al., 2000), pela execução das fases do ciclo de vida da Ciência de Dados.

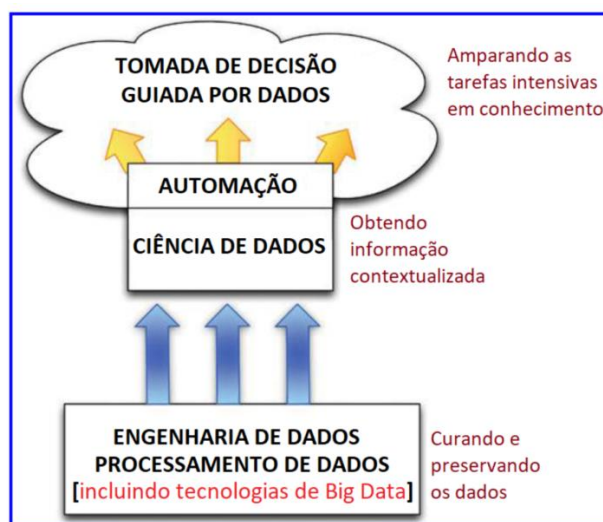
Figura 3 – Ciclo de Vida da Ciência de Dados



Dentre as fases apresentadas na “Figura 3”, destaca-se a fase de obtenção de dados resultantes de várias fontes, como do processamento de arquivos de texto; do monitoramento de uma rede de sensores; de consultas a bases de dados de sistemas legados e, principalmente, de dados oriundos da web (Rautenberg e Do Carmo, 2019).

No tocante à obtenção de dados, ressalta-se a sua necessidade como complemento para a tomada de decisão. O emprego de técnicas da Engenharia de Dados, como a Ciência de Dados, auxilia na obtenção da informação contextualizada e ampara a tomada de decisão que é uma das atividades intensivas em conhecimento.

Figura 4 – Ciência de Dados amparando o Processo de Tomada de Decisão



Fonte: Provost e Fawcett (2013) - adaptado por Rautenberg e Do Carmo (2019)

No contexto de obtenção de informações contextualizadas, o processo de extração de dados da web - Web Scraping - se torna relevante, pois utilizam da inteligência computacional para coletar um número massivo de dados. Destaca-se como aplicação de tecnologia para atender essa demanda o uso de RPA, destinada à automatização dos processos de negócios.

2.2.1 – Automação Robótica de Processos (RPA)

Segundo a ABPMP (2020, p. 262), a automação robótica de processos (Robotic Process Automation – RPA) refere-se ao uso de robôs de software que imitam tarefas geralmente executadas por seres humanos”. É uma aplicação tecnológica que se utiliza da programação para economia de tempo para os trabalhadores do conhecimento e otimização de entregas e resultados pois aprimora a

auditabilidade dos processos manuais, simplificando e otimizando os processos.

Além das propostas de valor apresentadas, pode-se incluir os seguintes benefícios adicionais:

- Muitos processos podem ser facilmente automatizados;
- A automação de tarefas repetitivas economiza tempo e recursos preciosos;
- Não é necessário treinamento quando humanos não realizam o processo;
- Os robôs de software não se cansam;
- Automatiza a extração de dados para gerar relatórios precisos, eficazes e oportunos.
- A produtividade aumenta, o que ajuda a aumentar a escalabilidade.

Com esse entendimento, utilizou-se recursos de RPA na automatização de processos mapeados, pois foi identificada a necessidade de automatização, devido as atividades repetitivas e normalmente exercidas por profissionais de nível mais alto. Nesse caso específico, os scripts desenvolvidos se encarregam das tarefas repetitivas, proporcionando aos profissionais que detêm o conhecimento a dedicação do seu tempo em outras tarefas.

De forma conclusiva à Revisão da Literatura, destaca-se que ao analisar as contribuições dos estudos anteriores por meio da busca em bases de dados científicas, é possível perceber que o presente artigo se distingue por investigar como a Ciência de dados pode auxiliar no processo de tomada de decisão nas organizações públicas, amparadas pelo mapeamento de processo intensivos em conhecimento. Vale ressaltar que não foram encontrados estudos com o mesmo cruzamento, viés e natureza nos periódicos pesquisados, o que justifica também a realização deste estudo.

Quadro 1 - Estudos empíricos sobre Ciência de dados, Tomada de Decisão e Processos Intensivos em Conhecimento

Autor/Ano	Objetivo	Aspectos metodológicos	Conclusões
Rautenberg e Carmo (2019)	Evidenciar a diferença e a complementariedade dos conceitos Big Data e Ciência de Dados	Pesquisa descritiva com abordagem qualitativa.	Aponta-se que o Big Data é relacionado à camada basilar de materiais, privilegiando os 6Vs atribuídos aos dados. Já a Ciência de dados é conexo à camada dos softwares, a qual metodologicamente transforma os dados em informação para o Apoio à Tomada de Decisão.
Bertoluzzi, Genari e Macke (2018)	Caracterizar os profissionais de TI, inseridos em organizações intensivas em conhecimento, e verificar os aspectos relacionados à aquisição de capital humano pelos referidos profissionais.	Pesquisa qualitativa com realização de entrevistas em profundidade. Os dados foram explorados com o emprego da análise de conteúdo.	Identificação dos principais elementos vinculados ao perfil dos profissionais de TI e aquisição de capital humano, através da aprendizagem formal.
Di Ciccio, Marrella e	Objetivo deste trabalho é preencher uma lacuna em caracterizar precisamente	Pesquisa com abordagem qualitativa e quantitativa (mista).	Forneceu uma caracterização precisa de Processos Intensivos em Conhecimento (KIPS) e, a partir de três aplicações no

Russo (2015)	o que são Processos Intensivos em Conhecimento, e (ii) identificar um conjunto de requisitos para sistemas orientados a processos visando seu suporte;		mundo real, foram elaborados alguns requisitos gerais para suportar o ciclo de vida de um KiPs. Finalmente, foi apresentado uma análise crítica sobre uma série de abordagens utilizadas para apoiar os KiPs, discutindo sua eficácia contra as exigências concebidas.
--------------	--	--	--

Fonte: Elaborada pelos autores

3 PROCEDIMENTOS METODOLÓGICOS

Para atingir os objetivos deste trabalho, bem como responder à questão de pesquisa, foi conduzida uma pesquisa-ação, seguindo os preceitos de Tripp (2005), junto a Controladoria Geral do Estado de Santa Catarina (CGE-SC), onde técnicas de Ciência de Dados e automação robótica de processos (RPA) foram aplicadas com o intuito de alimentar as bases de dados de forma qualificada e que esta possa auxiliar no processo de tomada de decisão.

Para fundamentação teórica fez-se uma pesquisa bibliográfica exploratória seguindo o método SSF de Ferenhof e Fernandes (2016), que versou sobre organizações intensivas em conhecimento, processos intensivos em conhecimento e automatização de processos. Estes conceitos ampararam o processo de identificação das atividades intensivas em conhecimento e, consequentemente, o seu melhoramento.

Destaca-se que durante a pesquisa-ação se aplicaram métodos e técnicas para mapeamento de processo que possibilitaram visualizar a sua execução e vislumbrar uma possibilidade de automatização. Por fim, utilizou-se técnicas pertencentes a ciência da de dados como web scraping, mineração de textos para o desenvolvimento da solução de automatização dos processos.

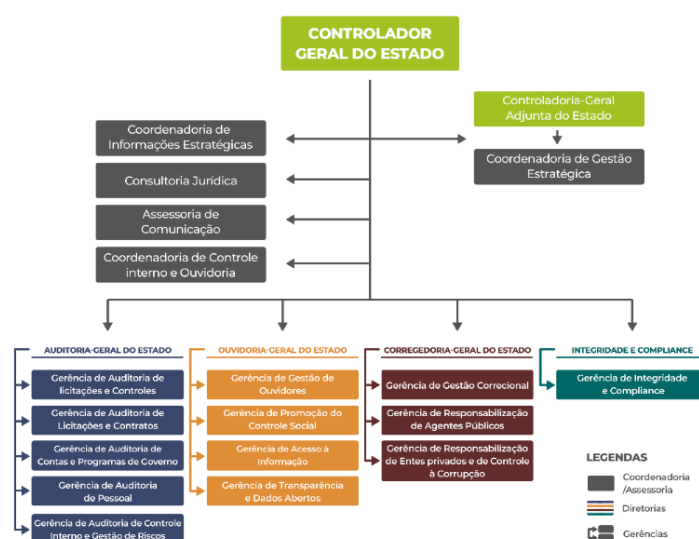
4 RELATO DA PESQUISA-AÇÃO NA CONTROLADORIA GERAL DE SANTA CATARINA

A Controladoria Geral do Estado de Santa Catarina (CGE-SC), que está estruturada conforme organograma apresentada na “Figura 5”, foi criada pela Lei Complementar nº 741, de 12 de junho de 2019. Ela é o órgão central do Sistema Administrativo de Controle Interno e de Ouvidoria do Poder Executivo. Subordinada diretamente ao Governador do Estado, tem a função de assessorá-lo na defesa do patrimônio público.

A CGE-SC possui em sua estrutura uma Coordenadoria de Informações Estratégicas (CIES) que, conforme Portaria CGE nº 08, de 15 de fevereiro de 2022, possui como competência prestar assessoria

ao Controlador-Geral e às unidades finalísticas da Controladoria-Geral do Estado por meio de coleta, busca e tratamento de informações de natureza estratégica para sua atuação, com emprego intensivo de recursos de tecnologia da informação e de atividades de investigação e inteligência, inclusive com emprego de técnicas operacionais, inspeções e análises, com vistas à coleta e busca de dados que permitam produzir informações estratégicas para subsidiar as atividades da Controladoria-Geral do Estado.

Figura 5 – Organograma da Controladoria Geral do Estado de Santa Catarina



Fonte: CGE-SC (2022)

Neste contexto, com vista a realizar o mapeamento das atividades intensivas em conhecimento, a CGE, em colaboração com a Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC), abriu um edital para selecionar recursos humanos qualificados para realizar a atividade de mapeamento.

Ao final do edital, foram selecionados dois bolsistas de pesquisa, Doutores em Engenharia e Gestão do Conhecimento, para realizar a atividade de mapeamento dos processos intensivos em conhecimento. Os bolsistas foram alocados na Coordenadoria de Informações Estratégicas (CIES) e nesse setor foram iniciados o mapeamento dos processos intensivos em conhecimento. Destaca-se que os bolsistas, bem como o coordenador da CIES formam o quadro de pesquisadores que executaram esta pesquisa-ação.

Como a CIES é responsável por municiar os auditores internos da CGE com informações

estratégicas advindas do sistema de *Business Intelligence* (BI) para a construção de relatórios técnicos, identificou-se que há nesse processo atividades intensivas de conhecimento e que por hora são executados manualmente e, por conseguinte, tomam uma parcela significativa do tempo de trabalho. O processo está relacionado a um dos módulos do BI/CGE desenvolvido pela CIES, denominado COPAS (Controle Organizacional Prévio para Análises Sistêmicas) e que apresenta informações executadas pelo Grupo Gestor de Governo (GGG) no Sistema Integrado de Planejamento e Gestão Fiscal do Estado (SIGEF). Tais informações são referentes aos pedidos de aquisição, contratação e alteração de contratos e instrumentos congêneres, incluindo aditivos de obras e serviços, realizados pelos órgãos e entidades da Administração Direta e Indireta do Poder Executivo. O Decreto nº 903/2020 definiu que o GGG é responsável pela análise prévia destes processos.

O mesmo decreto define a composição do GGG integrada pelos Secretários de Estado da Fazenda (seu Presidente), Secretário da Administração, pelo Chefe da Casa Civil e pelo Procurador (a) Geral do Estado. Cabe a este grupo a análise prévia e deliberação sobre a aquisição de materiais, contratação de serviços e obras, alteração de contratos e instrumento congêneres, no âmbito da administração direta e indireta do Poder Executivo. Além disso, cabe também analisar e autorizar previamente as alterações de ordem administrativa, financeira, orçamentária, patrimonial e organizacional, a criação de cargos de provimento em comissão, funções de confiança e empregos públicos permanentes ou comissionados, a serem realizadas pelas entidades da Administração Pública Estadual Indireta.

Entende-se que as informações fornecidas por este módulo são de suma importância, pois as decisões tomadas por este grupo têm caráter normativo ou autorizativo, e terão a forma de resolução e produzirão efeitos após serem homologadas pelo Governador do Estado e publicadas no Diário Oficial do Estado – DOE-SC.

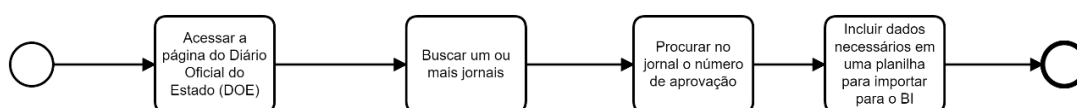
O diário oficial é o meio que a administração pública tem para divulgar todos os seus atos, tornando público tudo que o Estado faz. Ele procura atender os princípios da administração pública, definidos no Artigo 37 da Constituição Federal como: Legalidade, Impessoalidade, Moralidade, Publicidade e Eficiência. Dentre estes, o princípio da Publicidade é definido como os atos da administração pública que devem ser feitos com o conhecimento da população e, para isso, devem ser divulgados para que haja conhecimento e acesso de todos. Diante destas descrições, entende-se que as informações existentes no DOE-SC são de grande importância para o conhecimento da população, das empresas prestadoras de serviços e para os próprios órgãos de controle do Estado.

Devido a importância deste módulo para a análise e tomada de decisão, percebeu-se a

necessidade de complemento de informações relacionadas ao módulo, mas que não se encontravam disponíveis no BI da CGE, pois elas estão disponibilizadas na web, especificamente nos Diários Oficiais do Estado que se encontram na url: <https://doe.sea.sc.gov.br/index.php/buscar-jornal/>.

Percebendo-se isso, primeiramente, mapeou-se como os auditores realizavam a busca por informações complementares. Essa atividade segundo a ABPMP (2020, p. 157) chama-se situação atual do processo (AS IS) e ela serve para entendimento de como os auditores procedem para adquirir as informações complementares que não estão presentes no BI-CGE. Apresenta-se na “Figura 6”, a modelagem AS IS:

Figura 6 - Modelagem BPMn do processo de busca manual de Diários Oficiais



Fonte: Dados da pesquisa

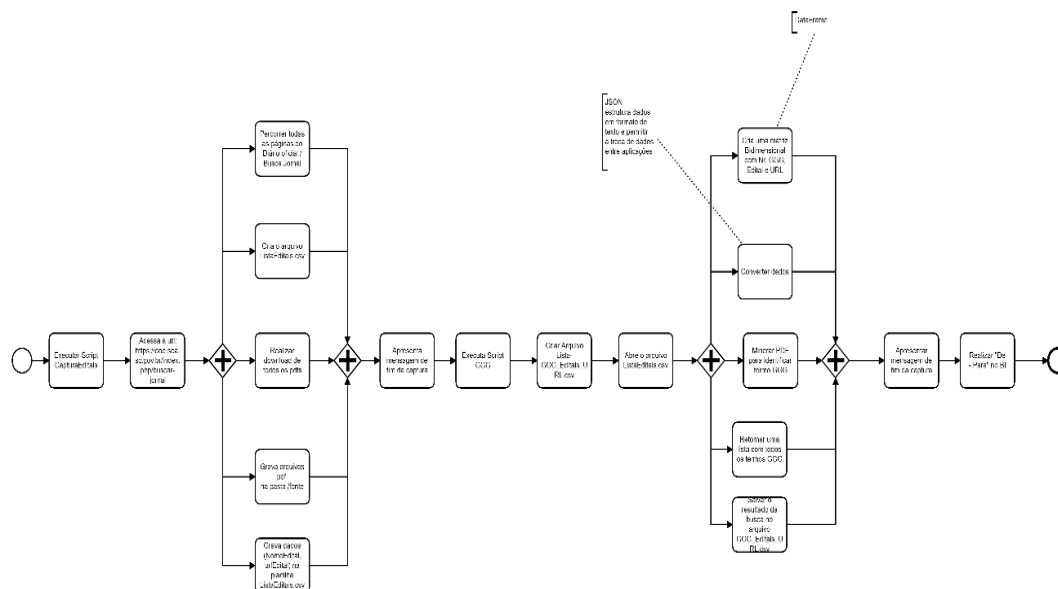
Após o mapeamento da situação inicial, identificou-se as atividades intensivas em conhecimento e que elas poderiam ser automatizadas por meio de tecnologias de RPA, como por exemplo, por um robô de busca.

A “Figura 7” apresenta o processo com vistas à automação do processo de busca dos diários, isto é, como acessar sem intervenção humana um browser, em uma url (endereço eletrônico único) especificada e realizar o download de cada um dos documentos publicados no formato PDF de forma automática. Na sequência, executar outro robô buscando por termos presentes nos diários identificando as aprovações existentes em cada diário publicado.

Na fase seguinte, usando como base o mapeamento AS-IS, partiu-se para a construção de um script na linguagem de programação Python. Concebe-se que, devido ao processo ser automatizado, na forma de robô de software, ele é usado para executar tarefas repetitivas, como um mecanismo de busca, ou simular o clique em um botão para realizar o download de um ou mais documentos.

Nesse sentido, realizou-se uma nova modelagem com perspectiva de procurar a melhor forma de realizar esse processo, que no CBOK v.4 (2020) denomina como processo otimizado (TO BE), conforme apresenta-se na “Figura 7”:

Figura 7 - Modelagem BPMn do processo otimizado (TO BE) de busca automatizada dos Diários Oficiais



Fonte: Dados da pesquisa

Na construção do robô foram utilizadas algumas técnicas computacionais para a automação do processo:

a) A primeira foi a técnica de *web scraping*, definida por Mitchell da seguinte forma:

“...a prática de coletar dados por qualquer meio que não seja um programa interagindo com uma API (ou, obviamente, por um ser humano usando um navegador web). Isso é comumente feito escrevendo um programa automatizado que consulta um servidor web, requisita dados (em geral, na forma de HTML e de outros arquivos que compõem as páginas web) e então faz parse desses dados para extrair as informações necessárias” (MITCHELL, 2019, p. 10).

Esta técnica foi utilizada para acessar automaticamente o endereço da página do Diário Oficial do Estado de Santa Catarina e realizar o clique nos botões de “DOWNLOAD” de cada edital publicado nas diversas páginas do site. Destaca-se que as páginas da url mencionadas são incrementais e, portanto, vão aumentando na medida em que haja mais publicações de Diários.

Figura 8 – Recorte do Script com indicação do link e botão DOWNLOAD

```
while paginacao.find('a', {'class': 'next page-numbers'}):
    numeroPag = numeroPag+1
    #Url da página do DOE
    url = 'https://doe.sea.sc.gov.br/index.php/buscar-jornal/?cp=' + str(numeroPag)
    response = urlopen(url)
    html = response.read()
    soup = BeautifulSoup(html, 'html.parser')
    paginacao = soup.find('ul', {'class': "pagination wpdm-pagination pagination-centered text-cent
    dadosEditais = []
    for editais in soup.find_all("div", {"class": "col-lg-12 col-md-12 col-sm-12"}):
        dadosEditais = []
        aEditais = editais.find('a')
        nomeEditais = aEditais.get_text()
        enderecoEditais = aEditais['href']
        divLinkdownload = editais.find('div', {"class": "ml-3 wpdm-dl-btn"})
        #botão DOWNLOAD
        aLinkdownload = divLinkdownload.find('a')
        hrefLinkdownload = aLinkdownload['data-downloadurl']
        #print(nomeEditais, " ", enderecoEditais, " ", hrefLinkdownload)
        dadosEditais.append(nomeEditais)
        dadosEditais.append(enderecoEditais)
        dadosEditais.append(hrefLinkdownload)
        editais.append(dadosEditais)
    #print(numeroPag)
```

Fonte: Dados da pesquisa

- b) A segunda técnica envolveu o uso de “Expressões Regulares”, definidas por Sweigart (2015) e Mitchell (2019, p. 41) como aquelas que são usadas para identificar *strings* (cadeia de caracteres) regulares. Elas são usadas também para encontrar e substituir padrões em uma sequência de caracteres ou até mesmo em um arquivo.

Figura 9 – Recorte do Script com uso das Expressões Regulares

```
#Procura a sequência de caracteres no json_file_string retorna uma lista
list_GGG = list(find_all_substring(json_file_string,"2023SO"))
# Converte em Dataframe e junta com o DF principal
df_append_out = pd.DataFrame({'AS_SO':list_GGG})
df_append_out['diario']= file
df_append_out['url_Diario']= url_Editais
df_out = df_out.append(df_append_out)

list_GGG = list(find_all_substring(json_file_string,"2023AS"))
# Converte em Dataframe e junta com o DF principal
df_append_out = pd.DataFrame({'AS_SO':list_GGG})
df_append_out['diario']= file
df_append_out['url_Diario']= url_Editais
df_out = df_out.append(df_append_out)
```

Fonte: Dados da Pesquisa

O resultado desse script é a geração de um arquivo no formato “csv” (Comma-separated values) com os dados complementares para o BI da CGE, trazendo o número da aprovação ou solicitação do Grupo Gestor de Governo, número do Diário e a Url do Diário - [‘AS_SO’, ‘diario’, ‘url_Diario’].

Essas informações alimentarão campos vazios do BI nomeados como “Solicitação”, “Aprovação” e “URL DIARIO”.

Figura 10 – Pannel do BI CGE-SC com os dados complementares capturados da WEB

CGE - Informações Estratégicas											
COPAS - Controle Organizaci...											
HOME	GERAL	DIÁRIO	SUBELEMENTO DESPESA			ITEM					
169000091 - Outros Recursos Primários - Multas de Trânsito	3 - Despesas Correntes	33 - Outras Despesas Correntes	90 - Aplicações Diretas	30 - Material de Consumo	33903028 - Material de Proteção e Segurança	2021SO002724	2021AS003947	3,00	https://doe.sea.sc.gov.br/index.php/download/11-05-2021-n-21518/ ; https://doe.sea.sc.gov.br/index.php/download/02-06-2021-n-21534/ ; https://doe.sea.sc.gov.br/index.php/download/02-06-2021-n-21534/	2021	
219000020 - Recursos de Outras Fontes - Exercício Corrente - Outras Taxas Vinculadas - Taxa de Fical	3 - Despesas Correntes	33 - Outras Despesas Correntes	90 - Aplicações Diretas	30 - Material de Consumo	33903001 - Combustíveis e Lubrificantes Automotivos	2021SO002760	2021AS003590	3,00	https://doe.sea.sc.gov.br/index.php/download/11-05-2021-n-21518/ ; https://doe.sea.sc.gov.br/index.php/download/11-05-2021-n-21518/ ; https://doe.sea.sc.gov.br/index.php/download/23-04-2021-n-21506/	2021	
					33903024 - Material para Manutenção de Bens Imóveis	2021SO002660	2021AS003592	3,00	https://doe.sea.sc.gov.br/index.php/download/11-05-2021-n-21518/ ; https://doe.sea.sc.gov.br/index.php/download/11-05-2021-n-21518/ ; https://doe.sea.sc.gov.br/index.php/download/23-04-2021-n-21506/	2021	
					33903028 - Material de Proteção e Segurança	2021SO002724	2021AS003947	3,00	https://doe.sea.sc.gov.br/index.php/download/11-05-2021-n-21518/ ; https://doe.sea.sc.gov.br/index.php/download/02-06-2021-n-21534/ ; https://doe.sea.sc.gov.br/index.php/download/02-06-2021-n-21534/	2021	

Fonte: Dados da Pesquisa

Além de parecer um mero complemento de informações no BI, esses campos possibilitam a consulta rápida de informações que antes não estaria disponível de imediato aos auditores da CGE, trazendo algumas vantagens, como:

- Diminuição de erros;
- Rastreabilidade das informações;
- Mais celeridade na execução das tarefas;
- Gestão segura e eficaz;
- Otimização do tempo, entre outras.

5 CONSIDERAÇÕES FINAIS

Com base nos relatos sobre o caso empírico, envolvendo especificamente a Controladoria Geral do Estado de Santa Catarina, entende-se a importância que o emprego das técnicas da Ciência de Dados possui na produção e amparo de informações complementares para a análise das informações estratégicas para a tomada de decisão.

Nesse mesmo contexto, ressalta-se o uso da modelagem de processos e sua automatização como uma alternativa vantajosa na melhoria da eficiência da gestão pública e na qualidade das políticas públicas. É, portanto, premissa nesse caso empírico, pois facilitou a identificação das atividades que necessitavam de automatização, bem como do que precisava ser usado no processo de automatização. Concebe-se com isso, o uso de técnicas computacionais modernas para a navegação, identificação e extração dos dados existentes em páginas web para agregar valor aos dados e/ou informações já existentes nas diversas bases governamentais para melhorar o processo de tomada de decisão

governamental.

Destaca-se que o conjunto de métodos, as práticas e sistemas científicos usados na extração de dados a partir de um grande volume de dados e, como o resultado, auxiliam na promoção do crescimento econômico por meio da grande quantidade de informações que, ao ser utilizada ou transformada possibilita a melhoria dos serviços de governo e a geração de novo conhecimento.

Com esse entendimento, destaca-se que a automatização de uma das atividades intensivas em conhecimento que era realizada por auditores na CGE-SC impactou na melhoria das informações usadas na tomada de decisão e ampliou a necessidade de outras automatizações para outros setores da administração pública catarinense. Essa percepção de que métodos, práticas e sistemas científicos possibilita a melhoria das políticas públicas governamentais se deve ao fato que os modelos governamentais tradicionais estão sendo substituídos por novos modelos, amparados pelo movimento de transformação digital das organizações e pelo desenvolvimento das Tecnologias da Informação.

Por fim, destaca-se que estar preparado para a Era dos Dados não significa apenas ter um uma vasta base de dados ou dispor de um código-fonte distinto que atenda seus usuários de maneira digital e automatizada. Será necessário redesenhar estruturas e processos decisórios, de forma que todas as decisões da organização estejam apoiadas, em menor ou maior grau, por dados.

Como trabalhos futuros recomenda-se a aplicação das técnicas em outros órgãos públicos e em empresas privadas, podendo comparar os resultados e medindo a eficácia das ações.

Agradecimentos

Os autores agradecem a Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina - FAPESC pelo apoio financeiro à pesquisa.

REFERÊNCIAS

ABPMP - Association of Business Process Management Professionals. **BPM CBOK V3.0**: Guia para o Gerenciamento de Processos de Negócio Corpo Comum de Conhecimento. Brasil: ABPMP, 2013.

ABPMP - Association of Business Process Management Professionals. **BPM CBOK V4.0**: Guia para o Gerenciamento de Processos de Negócio Corpo Comum de Conhecimento. Brasil: ABPMP, 2020.

BORTOLUZZI, Fernanda Rocha; GENARI, Denise; MACKE, Janaina. Capital humano nas organizações intensivas em conhecimento: desafios e perspectivas. **Revista Inteligência Competitiva**, v. 8, n. 1, p. 44-79, 2018.

DAVIS, G. B. et al. Conceptual model for research on knowledge work. **Management Information Systems Research Center Working Paper**, v. 91, n. 10, 1991.

DI CICCIO, Claudio; MARRELLA, Andrea; RUSSO, Alessandro. Knowledge-intensive processes: characteristics, requirements and analysis of contemporary approaches. **Journal on Data Semantics**, v. 4, p. 29-57, 2015.

FERENHOF, Helio Aisenberg; FERNANDES, Roberto Fabiano. Desmistificando a revisão de literatura como base para redação científica: método SSF. **Revista ACB**, v. 21, n. 3, p. 550-563, 2016.

FERREIRA, Marco. A Evolução da Web: o que esperar da Web 4.0. **Universidade Federal de Juiz de Fora**, v. 9, 2019. Disponível em :< <https://www.ufjf.br/conexoesexpandidas/2019/09/09/a-evolucao-da-web-o-que-esperar-da-web-4-0/>>. Acesso em 31 de mar. de 2022.

FREITAS JUNIOR, Vanderlei; GONÇALVES, Alexandre Leopoldo. Ontologias como suporte à avaliação enquanto tarefa intensiva em conhecimento: uma revisão bibliométrica. **Perspectivas em Ciência da Informação**, v. 19, p. 223-241, 2014.

KOENDERINK, Nicole. **A knowledge-intensive approach to computer vision systems**. 2010. Tese de Doutorado. Wageningen University and Research.

MARCONI, M. de A; LAKATOS, Eva Maria. Fundamentos de metodologia científica. 8. Ed. **São Paulo: Atlas**, v. 310, 2017.

MITCHELL, Ryan. **Web Scraping com Python: Coletando mais dados da web moderna**. Novatec Editora, 2019.

NEVES, Sandra Miranda et al. Risk management in software projects through knowledge management techniques: cases in Brazilian incubated technology-based firms. **International Journal of Project Management**, v. 32, n. 1, p. 125-138, 2014.

PROVOST, Foster; FAWCETT, Tom. Data science and its relationship to big data and data-driven decision making. **Big data**, v. 1, n. 1, p. 51-59, 2013.

RAUTENBERG, Sandro; DO CARMO, Paulo Ricardo Viviurka. Big data e ciência de dados: complementariedade conceitual no processo de tomada de decisão. **Brazilian Journal of Information Science: research trends**, v. 13, n. 1, p. 56-67, 2019.

SANTA CATARINA. Decreto nº 903 de 21 outubro de 2020. Dispõe sobre o Grupo Gestor de Governo (GGG) e estabelece outras providências, 2020.

SANTA CATARINA. Lei Complementar nº 741 de 12 junho de 2019. Dispõe sobre estrutura básica e o modelo de gestão da Administração Pública Estadual, no âmbito do Poder Executivo. **Diário Oficial do Estado**: n. 21.035, 12 jun., 2019.

SANTA CATARINA. Portaria CGE nº 08 de 16 fevereiro de 2022. Institui e define as competências e atribuições da Coordenação de Informações Estratégicas (CIES) na Controladoria Geral do Estado. **Diário Oficial do Estado**: n. 21.712, 16 fev. 2022.

SOUZA, Renato Rocha; ALVARENGA, Lídia. A Web Semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, v. 33, p. 132-141, 2004.

SWEIGART, Al. Automatize tarefas maçantes com Python. **São Paulo: Novatec**, 2015.

TRIPP, David. Pesquisa-ação: uma introdução metodológica. **Educação e pesquisa**, v. 31, n. 03, p. 443-466, 2005.

WANG, Y., CHEN, H., & LIU, Y. Big data and its technical challenges. **International Journal of Information Management**, 35, 1-12, 2015.