

Fiscalização a serviço da sociedade

REVISTA do TCU

Revista do Tribunal de Contas da União • Brasil • ano 48 • número 137 • Setembro/Dezembro 2016



A Evolução do Controle na Era Digital



Fiscalização a serviço da sociedade

REVISTA do TCU

Revista do Tribunal de Contas da União • Brasil • ano 48 • número 137 • Setembro/Dezembro 2016

© Copyright 2016, Tribunal de Contas da União
Impresso no Brasil / Printed in Brazil

Os conceitos e opiniões emitidas em trabalhos doutrinários assinados são de inteira responsabilidade de seus autores.

Permite-se a reprodução desta publicação, em parte ou no todo, sem alteração do conteúdo, desde que citada a fonte e sem fins comerciais.

www.tcu.gov.br

Missão

Aprimorar a Administração Pública em benefício da sociedade por meio do controle externo

Visão

Ser referência na promoção de uma Administração Pública efetiva, ética, ágil e responsável



Revista do Tribunal de Contas da União. - v.1, n.1 (1970) - . – Brasília : TCU, 1970- .

v.

De 1970 a 1972, periodicidade anual; de 1973 a 1975, quadrimestral; de 1976 a 1988, semestral; 1989, quadrimestral; 1990 a 2005, trimestral; 2006, anual; a partir de 2007, quadrimestral.

ISSN 0103-1090

1. Controle de gastos públicos – Brasil. 2. Controle externo – Brasil. I. Tribunal de Contas da União.



TRIBUNAL DE CONTAS DA UNIÃO

FUNDADOR

Ministro Iberê Gilson

SUPERVISOR

Ministro Aroldo Cedraz de Oliveira

CONSELHO EDITORIAL

Ministro-substituto Augusto

Sherman Cavalcanti

Procurador-Geral Paulo Soares Bugarin

Eduardo Monteiro de Rezende

Rainério Rodrigues Leite

Flávia Lacerda Franco Melo Oliveira

RESPONSABILIDADE EDITORIAL

Instituto Serzedello Corrêa

Centro de Documentação

COLABORADORES

Biblioteca Ministro Ruben Rosa

TRADUÇÃO

Secretaria de Relações Internacionais

Projeto Gráfico

Pablo Frioli

Imagens

iStock

Diagramação, capa e fotomontagens

Vanessa Vieira - ISC/ Seducont

Centro de Documentação

SAFS Quadra 4 Lote 1

Edifício Anexo III - Sala 21

Brasília-DF

70.042-900

revista@tcu.gov.br

Impresso pela Sesap/Segedam

Carta ao Leitor



Bruno Spada

Aroldo Cedraz de Oliveira

é ministro do Tribunal de Contas da União e supervisor do Conselho Editorial da Revista do TCU.

Prezado leitor,

A Administração Pública atravessa um período marcado por desafios importantes, provocados pela evolução tecnológica da Era Digital e pelo envolvimento crescente da sociedade na formulação de políticas públicas e em discussões sobre eficácia e transparência na aplicação de recursos do Estado. Testemunhamos uma transformação nas relações entre governo e cidadãos, à qual o controle externo não pode e não vai ficar indiferente.

Os que acompanham a *Revista do TCU* sabem que o tema não é inédito nesta publicação. Ainda assim, é certo que estamos longe de esgotar o assunto, haja vista a amplitude de possibilidades e perspectivas que essa nova era nos traz. A presente edição, em particular, dedica-se a um olhar para o futuro próximo – para os avanços que poderão ser conquistados a partir das mais modernas técnicas de mineração de dados, análise semântica de textos, geoprocessamento e realidade virtual.

Investimentos em ciência e tecnologia são essenciais ao desenvolvimento sustentável e à construção de economias produtivas, com sociedades mais justas e inclusivas. Segundo a Organização para Cooperação e Desenvolvimento Econômico (OCDE), “uma das lições mais importantes das últimas duas décadas foi o papel fundamental da inovação no desenvolvimento econômico”. Assim, ao intensificar o uso da Tecnologia da Informação no controle externo e compartilhar suas experiências, o Tribunal de Contas da União (TCU) aponta caminhos a serem seguidos pelo Estado, de modo a tornar-se mais eficiente, transparente e eficaz no atendimento aos anseios da população.

Não há dúvidas de que a atividade de auditoria, na Era Digital, passa necessariamente pelo uso intensivo do assim chamado Big Data Analytics. O processamento de volumes colossais de dados, para deles extrair conhecimentos úteis à gestão pública e à sua fiscalização, torna-se cada vez mais presente graças à evolução da aprendizagem de máquina e da inteligência artificial. Como nos lembra o entrevistado desta edição, Cezar Taurion, os avanços nessa área são tão rápidos que temos dificuldade até mesmo em imaginar os resultados que poderão ser alcançados nos próximos anos.

Ao assumir posição de vanguarda no uso desses instrumentos, o TCU já tem colhido frutos de modelos preditivos e algoritmos inteligentes produzidos por nossas próprias equipes técnicas. Como relata na coluna “Opinião” o Secretário Wesley Vaz, responsável por tais iniciativas, trata-se de incorporar novas ferramentas à rotina de trabalho dos auditores, de modo a escolher melhor as ações de controle a serem realizadas pelo Tribunal e, ato contínuo, executá-las de forma eficiente, tempestiva e precisa, conferindo assim maior efetividade à atuação desta Corte.

De maneira similar, estão em desenvolvimento novos métodos de trabalho com base na captação, no tratamento e na análise de imagens – seja por técnicas tradicionais de geoprocessamento ou com as inovações da realidade virtual – que prometem revolucionar, dentro de pouco tempo, a realização de auditorias ambientais, em obras públicas e em outros empreendimentos que envolvam grandes extensões territoriais.

Por fim, conforme ilustram diversos artigos publicados nesta edição, é importante ressaltar que a transformação digital da atividade de controle não é assunto de interesse exclusivo dos Tribunais de Contas e órgãos similares. Universidades brasileiras, como a Universidade de Brasília (UnB) e a Universidade Federal de Minas Gerais (UFMG), e centros de excelência internacionais em contabilidade e auditoria, como a Universidade Rutgers, nos Estados Unidos, possuem equipes de pesquisadores dedicados ao desenvolvimento de soluções que potencializem o uso da tecnologia e de grandes massas de dados como instrumentos para se combater, com maior eficácia, os desvios e as fraudes na aplicação dos recursos públicos.

Estamos, portanto, diante do desafio de inovar continuamente, com foco em resultados orientados pela estratégia do Tribunal, de modo a consolidar o protagonismo desta Casa perante o sistema de controle nacional e internacional. Mais ainda, trata-se de imperativo para que sejamos capazes de desempenhar a contento a missão que nos foi incumbida há 126 anos: aprimorar a administração pública em benefício da sociedade.

Boa Leitura!

Entrevista

07



Entrevista

Cezar Taurion

Sócio e líder de Digital Transformation da Kick Ventures e VP de Inovação do Instituto Smart City Business

- 07** As inovações tecnológicas na fiscalização

Opinião

13



Opinião

Wesley Vaz Silva

Secretaria de Gestão de Informações para o Controle Externo

- 13** Os pilares da estratégia de análise de dados e consumo de informações no TCU

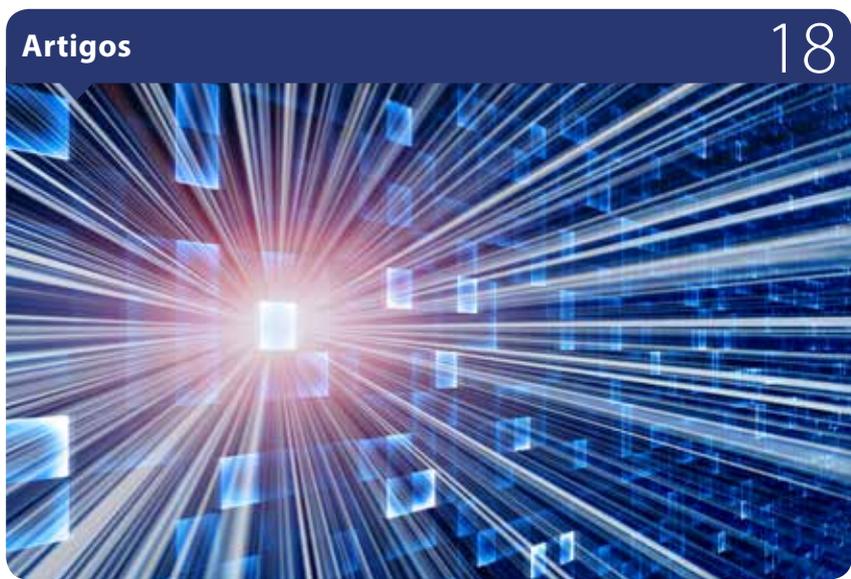
Destaque

17



Destaque

- 17** Nova Pesquisa em Jurisprudência



Artigos

18 Fundamentos de uma auditoria em realidade misturada

- *Aroldo Cedraz*
- *Francisco Osório Ramos*

24 Plataforma de serviços de dados abertos do TCU: *crowdsourcing*, nuvem cívica e aplicativos cívicos

- *Monique Louise Monteiro*
- *Remis Balaniuk*
- *Marcelo Pacote*

32 A recuperação semântica da informação no contexto do controle externo

- *Márcia Martins de Araújo Altounian*
- *Beatriz Pinheiro de Melo Gomes*

43 Geotecnologias e o monitoramento dos Objetivos de Desenvolvimento Sustentável pelas Entidades de Fiscalização Superior

- *Rherman Radicchi Teixeira Vieira*
- *André Delgado de Souza*
- *Leonardo Pereira Garcia*
- *Erick Muzart Fonseca dos Santos*

53 InfoSAS: um sistema de mineração de dados para controle da produção do SUS

- *Oswaldo Carvalho*
- *Marcos Prates*
- *Raquel Minardi*
- *Wagner Meira Jr.*
- *Renato M. Assunção*
- *José Nagib Cotrim Árabe*

61 Aplicativos de auditoria: uma ferramenta eficaz para assegurar de compras governamentais

- *Qiao Li*
- *Jun Dai*

71 Modelagem de dados geográficos para definição de corredores alternativos no rododanel da região metropolitana de Belo Horizonte: cenários comparativos

- *José Irley Ferreira Júnior*
- *Rodrigo Affonso de Albuquerque Nóbrega*
- *Leise Kelli de Oliveira*

81 O potencial de dados de sensoriamento remoto na fiscalização de obras públicas

- *Osmar Abílio de Carvalho Júnior*
- *Renato Fontes Guimarães*
- *Roberto Arnaldo Trancoso Gomes*

97 Implementação de um geocatálogo para auxílio na localização e recuperação de dados geográficos abertos

- *Drausio Gomes dos Santos*
- *Alexandre Zaghetto*

107 O porquê de governança de dados em organizações de controle

- *Ricardo Dantas Stumpf*

117 O georreferenciamento dos bens imóveis públicos no sistema geodésico brasileiro para fins de incorporação no cadastro técnico multifinalitário: construção da regularização imobiliária dos municípios

- *Davi Lopes Silva*

125 Uso de técnicas de inteligência artificial para subsidiar ações de controle

- *Luís André Dutra e Silva*

Vinícius Magalhães



As inovações tecnológicas na fiscalização

Cabe ao Tribunal de Contas da União verificar se as instituições governamentais estão utilizando recursos públicos de forma efetiva e eficiente. Como as inovações tecnológicas podem ajudar o TCU a cumprir essa missão e “fazer mais com menos”?

A evolução tecnológica avança em ritmo exponencial, surpreendendo aos mais desavisados. A capacidade computacional de um supercomputador de 10 milhões de dólares em 1985 está embutida hoje em um smartphone. A Internet está se massificando. Um rápido olhar ao passado recente mostra quão impressionante é o ritmo das mudanças. Em 1995, há apenas 21 anos, éramos 35 milhões de navegantes na Internet mundial. Hoje somos quase três bilhões, cerca de 40% da população mundial. 80 milhões usavam celulares, hoje três em cada quatro pessoas do mundo têm um, ou mais de 5,2 bilhões de pessoas. Não devemos pensar na revolução digital como uma simples automatização de processos atuais. Estamos em um novo jogo. O trabalho, os modelos de negócios, as estruturas organizacionais, produtos e serviços serão efetivamente transformados através das tecnologias digitais. Hoje, as tecnologias permitem de forma muito mais fácil e barata que

Cezar Taurion

Sócio e líder de Digital Transformation da Kick Ventures e VP de Inovação do Instituto Smart City Business

Além de profissional da Tecnologia da Informação, Cezar Taurion é um ávido estudioso dessa área desde o final da década de 70, tendo produzido nove livros sobre os mais variados ramos da TI, como software livre, grid computing, software embarcado, inovação, cloud computing, big data e transformação digital. Na sua formação acadêmica entram conhecimentos em Economia, Ciência da Computação e Marketing de Serviços. Ele foi professor do MBA em Gestão Estratégica de TI na FGV/RJ e de Empreendedorismo na Internet pelo NCE/UFRJ. Nos últimos anos, tem participado ativamente do desenvolvimento e aplicação de novas tecnologias, tanto no Brasil como no exterior, o que lhe permite acompanhar casos reais das mais diversas características e complexidades.

No desempenho dessa atividade, Taurion busca compreender e avaliar os impactos das inovações tecnológicas nas organizações e em seus processos de negócio. Para ele, o pensar fora da caixa é insuficiente porque “a caixa ainda nos prende a uma fonte de referência, limitando o pensamento inovador”. Nesta entrevista à *Revista do TCU*, o especialista em TI aborda temas como transformação digital, tecnologias exponenciais, Inteligência Artificial (IA) e o impacto de tudo isso na sociedade, nos negócios e nas atividades do Tribunal de Contas da União.

há dez anos, cruzar dados obtidos de diversas fontes, que variam de bases de dados convencionais a redes sociais, onde todos nós deixamos nossas pegadas digitais. Os cidadãos estão mais empoderados e com seus smartphones estão conectados a milhares de outros através de redes sociais, compartilhando seus pontos de vista, opiniões e comentários. Assim, além de tecnologias mais baratas e disponíveis, as ações de fiscalização podem contar com a colaboração da própria sociedade.

Quais as tecnologias que o senhor classificaria como mais importantes para o Controle Externo? Por quê?

Algumas tecnologias como a Internet e os smartphones permitem acesso rápido e fácil a qualquer momento. Todos os órgãos de governo podem e devem estar conectados, enviando informações continuamente. Por que esperar semanas e meses para analisar dados, se esses puderem ser analisados de forma contínua? As ações terão efeito muito mais imediato. As análises podem ser muito mais eficazes se usarmos algoritmos e inteligência artificial. Aliás, algoritmos já estão em toda parte, de recomendação de livros e filmes à detecção de fraudes por operadoras de cartões de crédito. Estamos imersos em um oceano de dados, e pouco uso fazemos deles. Estimativas apontam que em 2020 estaremos criando em todo o planeta 73,5 zettabytes de dados ou 73 seguido de 21 zeros! Os resultados da aplicação de algoritmos mostram que se conseguem bons resultados em praticamente qualquer situação. Por exemplo, um estudo efetuado nos EUA, pela American Psychological Association, analisando 17 casos de estudo de práticas de contratação por grandes empresas, mostrou que o uso de algoritmos vencia as melhores práticas (geralmente baseadas na intui-

“ Todos os órgãos de governo podem e devem estar conectados, enviando informações continuamente. Por que esperar semanas e meses para analisar dados, se esses puderem ser analisados de forma contínua? ”

ção) por 25%, quando considerando o sucesso da contratação, ou seja, o acerto da contratação do novo funcionário na empresa. Com o uso de algoritmos preditivos, podemos não apenas nos contentar com descrição e análise de fatos passados, mas fazer previsões de comportamento de órgãos públicos baseadas nos padrões observados até o momento.

Na sua opinião, qual o maior obstáculo quando se quer introduzir inovações tecnológicas no ambiente de trabalho? As barreiras culturais ou as limitações financeiras? Por quê?

Existem, claro, as limitações financeiras, mas o custo da tecnologia vem caindo e continuará caindo de forma expressiva. Hoje, por cerca de cem reais, compram-se dispositivos de armazenamento de um terabyte (um trilhão de bytes) que custavam centenas de milhares de reais há pouco mais de dez anos. A grande barreira é a cultural. O nosso pensar de forma linear, quando a evolução é expo-

nencial, nos leva à terrível armadilha de subestimar o impacto das transformações. Um exemplo de como a mudança exponencial é subestimada foi o Human Genome Project. Foi lançado em 1990, com estimativa de ser concluído em 15 anos a um custo de US\$ 6 bilhões. Em 1997, metade do prazo, apenas 1% do genoma humano tinha sido sequenciado. Pelo planejamento linear que nós adotamos, supondo 1% em 7 anos, levaríamos 700 anos para concluir o sequenciamento. Parece lógico, não? A pressão para encerrar o projeto foi imensa, mas quando perguntaram ao futurista Ray Kurzweil, ele disse “1% significa metade do caminho. Vão em frente!”. Ele pensou exponencialmente. 1% dobrando a cada ano significa chegar aos 100% em 7 anos. O projeto foi concluído em 2001, quatro anos antes do planejado e custando muito menos dinheiro que o estimado. O pensamento linear, tradicional, errou o alvo por 696 anos! Muitas vezes o pensar fora da caixa é insuficiente, pois a caixa ainda nos prende a uma fonte de referência, limitando o pensamento inovador...

O envolvimento do cidadão nas análises e decisões tomadas pelo governo parece ser uma abordagem que veio para ficar. Hoje em dia essa participação é, inclusive, considerada elemento essencial em uma democracia. Quais as ferramentas tecnológicas que você considera mais eficazes na promoção do diálogo entre as instituições públicas e a sociedade?

As redes sociais já fazem parte do dia a dia de grande parte da sociedade e esse número tende a aumentar. Se observarmos de perto esse fenômeno constatamos que a Internet e as redes sociais podem ser vistas como tecnologias cognitivas descentralizadas, onde mais e mais pessoas podem emitir suas ideias e opiniões de

forma muito rápida, fora do controle dos mecanismos centralizadores. Comparemos um evento transmitido ao vivo por uma TV unidirecional com uma rede social. Na TV somos obrigados a ver as imagens e ouvir a voz que a emissora libera, enquanto nas redes, cada um que esteja presente ao evento filma de um determinado ângulo e emite suas próprias percepções do fato, compartilhando em tempo real com milhares de outros. Cria uma topologia de poder menos vertical e muito mais aberta. Não existe mais o controle centralizado dos canais de distribuição de ideias e as trocas horizontais substituem o modelo vertical e centralizado.

Nos últimos anos, a atividade de fiscalização migrou de falta de informação para abundância de dados. Qual o caminho mais viável para uma instituição - cuja atividade principal depende da análise de informação - identificar os melhores dados e decidir como usá-los/interpretá-los?

A imensa disponibilidade de dados embute mudanças na maneira de pensarmos dados. Quando saímos do pensamento baseado na escassez para abundância de dados, devemos pensar diferente. Pela dificuldade e limitação tecnológica que existia, nós acabamos construindo um modelo mental de escassez de dados. Com isso refinamos uma série de práticas como análises estatísticas por amostragem. A partir de uma pequena amostra de dados, extrapolamos para um cenário mais amplo. Com o tempo refinamos os modelos e hoje eles são bastante confiáveis. Entretanto, há algumas lacunas. A precisão depende muito da amostragem. Por exemplo, uma pesquisa de opinião baseada em uma amostra randômica de usuários de telefones fixos embute um viés: se a coleta for feita no horário de trabalho, quem vai atender não necessariamente re-

“ Com o uso de algoritmos preditivos, podemos não apenas nos contentar com descrição e análise de fatos passados, mas fazer predições de comportamento de órgãos públicos baseadas nos padrões observados até o momento.”

presenta a opinião das pessoas que trabalham fora. Podem ter um ponto de vista bem diferente das que podem atender o telefone doméstico durante o dia. Além disso, se quisermos detalhar um pouco mais a pesquisa, um universo de amostras pequeno, como fazemos hoje, não terá representatividade estatística.

Qual seria um bom exemplo disso?

As pesquisas de intenção de voto. Geralmente pegam-se umas duas mil pessoas e tem-se um quadro geral. Mas se quisermos detalhar ao ponto de querermos saber a intenção de jovens entre 18 e 25 anos do estado da Paraíba, a amostragem será insuficiente. Estamos presos às perguntas iniciais e não podemos sair delas. Mas com volumes grandes o pensamento é outro. Quando a variável passa a ser “N=todos” podemos fazer granações inimagináveis no modelo de escassez. Podemos identificar tendências e descobrir correlações não pensadas antes. Podemos fazer novas perguntas e descer a novos níveis de segmentação. Saímos para um “mind set” mais oportuno ou seja, aproveitamos oportunidades de fazer perguntas não pensadas antes de

analisar os dados. Outra característica interessante que afeta nossa maneira de olhar os dados é que volumes grandes não demandam precisão extra de cada dado. Aliás, já fazemos isso hoje. Pensemos em um número grande como o PIB de um país. Não detalhamos os centavos, mas ficamos restritos aos grandes números e às tendências que eles apontam.

O que é necessário para se obter dados concretos, quando se está diante de uma abundância de informação?

Temos que seguir alguns passos. Primeiro é essencial ter um patrocinador de alto nível na organização. Não pode ser uma ação limitada a determinada área ou restrita ao setor de TI. Depois, definir claramente os objetivos de negócio, que problemas as iniciativas de análise massiva de dados irão resolver? Um fator fundamental: a equipe. Dificilmente conseguiremos pessoas que atendam à equação hackers + profundos conhecimento estatísticos e matemáticos + bons conhecimento de negócio. Uma sugestão é montar uma equipe multidisciplinar e operacionalizar os processos que envolvam os projetos de análise de dados. Claro que a equipe deve ter um bom gestor, que consiga entender as diversas linguagens faladas por profissionais tão diferentes entre si e que seja apaixonado pelo conceito de analisar dados. Um gerente burocrata não vai conseguir desfiar os inevitáveis problemas de comunicação inter e intra equipe. E montar uma equipe apenas com hackers, por exemplo, pode gerar um algoritmo preditivo sensacional, mas de pouco valor para a organização. Afinal, o objetivo não é gerar modelos analíticos fantásticos, mas sim resolver problemas de negócio, como eficácia e eficiência da fiscalização. E, finalmente, governança. Pelas características de lidar

com volumes muito grandes e variados, com dados não estruturados (antítese do modelo estruturado e relacional que estamos acostumados), tende-se a não documentar e nem criar processos de governança. Com isso existe o risco de reinventar-se a roda constantemente.

As ferramentas de TI são muito eficientes na coleta e classificação de dados, mas o senhor acredita que elas podem ser confiáveis quando se trata de interpretação?

Depende é claro da eficiência do algoritmo. Entretanto, a evolução é muito rápida. Em 2009, quando o Google falava em carros autônomos, era curiosidade futurista. Hoje já é praticamente uma realidade. Hoje os sistemas de reconhecimento e imagens do Facebook são mais eficientes que o ser humano em reconhecer rostos em fotografias. Portanto, temos que pensar de forma exponencial. Isso significa que se uma tecnologia digital, hoje, ainda não está tão eficaz, talvez em dois a três anos esteja dez ou vinte vezes melhor.

Inteligência Artificial (IA) e Aprendizagem Automática Avançada (Advanced Machine Learning) estão entre as chamadas Tendências Estratégicas de Tecnologia, uma vez que permitem a criação de sistemas capazes de compreender, aprender e prever eventos que podem melhorar a tomada de decisão. Na sua opinião, quais são os desafios e benefícios da utilização dessas tecnologias por uma Corte de Contas?

A revolução conduzida pela IA está chegando tão rápido que temos dificuldade em imaginar como ela se tornará. O imaginário de ficção científica ainda predomina. Lembro de uma frase, de um dos filmes que me marcaram e que tenho em casa, “2001: A Space Odyssey”, onde o

“Muitas vezes o pensar fora da caixa é insuficiente, pois a caixa ainda nos prende a uma fonte de referência, limitando o pensamento inovador...”

computador se intrometia na vida dos astronautas da tripulação: “Just what do you think you’re doing, Dave?”. De lá para cá vimos filmes de robôs como “Terminator”, “Eu, robô”, “Ela” e Jarvis, o assistente pessoal de Tony Stark em Iron Man. A propósito, Jarvis significa “Just Another Rather Very Intelligent System”. Ficção científica poderia muito bem ser definida como antecipação cinetífica. Muito que vemos em Jarvis, de alguma forma, já está no nosso dia a dia.

Jeff Hawkins, fundador da Numeta (e inventor do Palm Pilot), diz que a IA está hoje em um ponto similar ao da computação no início da década de 1950, quando os pioneiros estabeleceram as ideias básicas dos computadores. Menos de 20 anos depois, os computadores tornaram possíveis sistemas de reservas de companhias aéreas e ATMs (caixas eletrônicos) bancários e ajudaram a NASA a colocar o homem na lua, resultados que ninguém poderia ter previsto nos anos 50. Adivinhar o impacto da IA e dos robôs em uma década ou duas está se tornando ainda mais difícil. “Daqui a vinte anos, essa tecnologia será um dos principais motores da inovação e da tecnologia, se não a principal”, diz Hawkins. Indiscutível que a IA vai afetar a sociedade e o emprego como conhecemos. A automação, em seu início, afetou as linhas de produção nas fábricas.

Agora o risco de desemprego afeta funções que antes eram reservadas aos humanos. Por exemplo, motorista de caminhão. É um dos trabalhos mais comuns no mundo todo. São 3,5 milhões deles nos Estados Unidos e, aqui no Brasil, temos mais de um milhão registrados para o transporte de carga. O governo holandês já realizou um teste bem-sucedido de caminhões sem motorista cruzando a Europa. O Uber recentemente pagou US \$ 680 milhões para comprar Otto, uma startup que desenvolve tecnologia para caminhões autônomos e que foi fundada por especialistas de IA do Google. A consultoria McKinsey previu que, dentro de oito anos, um terço de todos os caminhões na estrada será autônomo, rodando sem motoristas. Em talvez 15 anos, o motorista de caminhão, como o ascensorista, será um anacronismo. O Uber investiu em Otto não apenas para operar caminhões, mas porque quer operar frotas de carros autônomos. Em setembro, começou a testar essa frota em Pittsburgh. O serviço postal do Canadá quer enviar aviões não tripulados, em vez de vans, para entregar correio rural.

No TCU, me parece claro que o papel da IA será o de um auxiliar extremamente eficaz na análise de contas, analisando de forma ampla e rápida milhões de documentos, fazendo comparações e identificando correlações que nem sempre nós humanos conseguimos fazer. Não temos capacidade de lidar com grandes volumes de dados muito rapidamente. Assim, nossas decisões muitas vezes são influenciadas por experiência pessoais.

As instituições governamentais estão cada vez mais dependentes de sistemas, tecnologia e dados. Nesse cenário, segurança da informação e segurança cibernética desempenham

um papel extremamente importante. Na sua opinião, quais são os principais desafios e entraves para tornar sistemas e informações mais seguros?

Esse é um dos grandes desafios. Temos dois pontos a validar. Um é a garantia de segurança dos dados. Outro é garantir privacidade. E esse é um assunto que é bem fluído. Deixamos uma imensa pegada digital no nosso dia a dia. As redes sociais, como o Twitter e Facebook, são gigantescos repositórios de opiniões e comentários. Por exemplo, haverá mais palavras escritas no Twitter nos próximos dois anos do que as contidas em todos os livros já impressos. O Facebook, com seus mais de 1,4 bilhão de usuários (936 milhões entram todo dia nele), é hoje a maior rede da história da humanidade. Olhando esta imensa quantidade de usuários e considerando que aproximadamente um quarto do mundo tem menos de 14 anos de idade, isso significa que mais ou menos 25% dos adultos do planeta têm conta no Facebook. Outra fonte inesgotável de informações sobre quem somos é a famosa página inicial do Google, contendo apenas um campo para entrada de dados. É um repositório do id coletivo da humanidade. Ele ouve nossas confissões, preocupações e segredos. Nós digitamos naquele retângulo o que queremos, sem censura. Como são buscas que as pessoas fazem sem censura, expressam nelas sentimentos de ódio e preconceitos, que geralmente são camuflados pelo comportamento social em público. Este é o problema do cientista social: o que eles mais querem saber é justamente o que seus objetos de estudo mais tentam esconder. O simples ato de perguntar algo incômodo gera

“ No TCU, o papel da IA será o de um auxiliar extremamente eficaz na análise de contas, analisando de forma ampla e rápida milhões de documentos, fazendo comparações e identificando correlações que nem sempre nós humanos conseguimos fazer.”

autocensura. Nas buscas, consegue-se obter informações praticamente impossíveis de se obter por pesquisas tradicionais. O Google Trends pode gerar excelentes amostras da mente privada, do que as pessoas realmente querem saber e que nem sempre compartilham com outras. Buscam sozinhas. Apenas o Google sabe... Vários projetos já mostram o quanto é possível saber com o Google Trends, como previsão do mercado de ações, o que move a produtividade econômica, epidemias de gripe e dengue, incidência de racismo e preconceitos arraigados. Como estamos falando de algo muito recente, existe ainda uma longa caminhada de aprendizado. A Internet é muito jovem como registro humano predominante e o próprio Facebook só ganhou esta proporção toda nos últimos seis anos. As informações sobre o comportamento humano ainda estão sendo construídas e talvez daqui a dez ou vinte anos

conseguiremos responder mais precisamente a perguntas como nos relacionamos, como novas ideias se infiltram e se disseminam pela sociedade, como a timeline do Facebook expressará a vida de uma pessoa (hoje ele recebe em média 0,6 Mb de novos dados por usuário, por dia), como os tuites mostrarão a reação da sociedade a determinados eventos e como o retângulo do Google explicitará os cantos das nossas mentes.

A inovação é, hoje em dia, um fator crítico de sucesso para as organizações, incluindo o governo. A adoção de novas metodologias e tecnologias é frequentemente um requisito essencial para a inovação. Nesse sentido, quais métodos e técnicas, na sua avaliação, teriam potencial para promover práticas de controle inovadoras?

O impacto da transformação digital deverá ser muito maior que o do conceito de e-commerce de uns 15 anos atrás. Hoje a maior parte das transações que setores inteiros de indústria fazem com seus clientes já é via Internet, como bancos, comércio e empresas aéreas. Comprar online é rotina. Novos modelos de negócio surgem, como o AirBnB, e colocam em cheque setores consolidados, como o hotelheiro. Em muitos países, novos modelos de empréstimos, como os fornecidos pelo LendingTree (EUA) e Kiva, estão mudando a relação entre a sociedade e os fornecedores tradicionais de empréstimos, as instituições financeiras. Na Alemanha o Friendsurance é uma ruptura no modelo tradicional da indústria de seguros.

Está nítido que a sociedade já está acostumada a usar a Web e os apps dos smartphones para suas atividades do dia a dia, seja a de localizar um táxi, comprar um produto, fazer um

check-in do voo ou uma transferência eletrônica entre contas correntes. Cosméticos é um exemplo interessante. Há cinco anos, os brasileiros não compravam pela Web. Em 2014, gastaram 1,3 bilhão de reais. Ainda apenas 1,5% das vendas deste mercado, contra 6% nos EUA. Imaginem o potencial de crescimento. E, novamente, quem começou foi um negócio fora das tradicionais empresas do setor, a BelezanaWeb. O uso de aplicativos como WhatsApp mudou a maneira como as pessoas interagem. Famílias trocam mais mensagens entre si que falam por fone. E praticamente todos nós compartilhamos, até compulsivamente, nossas ideias, opiniões e quase tudo que fazemos em plataformas de mídias sociais como Facebook e Instagram. A sociedade está cada vez mais tecnológica e o processo de consumerização de TI é um movimento que pressiona as empresas a adotar as mesmas tecnologias que seus clientes já usam. Eles adotam primeiro!

As organizações têm que se ajustar à velocidade da transformação digital. A transformação de uma organização depende de mudança da mentalidade na alta administração. Os gestores precisam entender a urgência da mudança para poderem provocá-la. Com este apoio, uma nova mentalidade incentivadora de mudanças provoca a contratação de talentos que hoje inexistem. A transformação digital exige que a organização saia da sua zona de conforto. A velocidade das mudanças não permite mais planejamentos de longo prazo baseados apenas em evoluções incrementais, como ampliação de um mercado ou lançamento de novos produtos similares. O inesperado surge a cada instante. O cenário de negócios passa a ser cada vez mais volátil, ambíguo, incerto e complexo. Os grandes bancos só se moveram com rapidez em direção

“ A transformação de uma organização depende de mudança da mentalidade na alta administração.”

ao banco digital porque foram provocados pelas Fintechs. A velocidade das mudanças não foi opção deles, mas necessidade de reação. Para visualizar e construir o futuro, temos que ser, ao mesmo tempo, otimistas e realistas. Não é fácil, mas precisamos construí-lo, para não sermos desconstruídos por ele.

A gestão do conhecimento (Knowledge Management, KM) é ainda um grande desafio para organizações como o TCU, que têm na informação e no conhecimento a principal matéria prima para desempenhar suas atividades. Qual abordagem o senhor acredita que terá maior impacto nas estratégias de KM dos próximos anos e por quê?

A gestão do conhecimento tem muito a ver com o uso de algoritmos e IA. Vamos pegar, para comparação, uma tradicional profissão, a advocacia. Uma provocação que podemos fazer é: “ainda existirão advogados no futuro?”. A taxa de acertos em previsões futuristas é a mesma de chimpanzés jogando dardos e acertando o alvo, mas podemos debater algumas ideias e tirarmos conclusões por nós mesmos. Claro, desde que não nos apeguemos a crenças e paradigmas que nos limitam o olhar crítico. Vamos analisar o contexto. As práticas de trabalho dos advogados não mudaram muito nas últimas décadas. De maneira geral, os advogados oferecem assessoria personalizada de alto custo e os sócios dos prestigiados

escritórios presidem organizações em forma de pirâmide, recebendo altas comissões, enquanto batalhões de advogados principiantes fazem o trabalho árduo de buscar precedentes e elaborar contratos. Os altos custos destes escritórios e dos seus honorários propiciam um cenário aberto a disrupções. Já existem algumas iniciativas muito interessantes que, ainda desdenhadas pelos advogados tradicionais, podem provocar um efeito Uber nos próximos anos.

Alguns escritórios de advocacia nos EUA já usam IA como “associado digital”, delegando a algoritmos preditivos a tarefa de executar buscas inteligentes por documentos, pareceres e jurisprudências referentes aos casos em análise. Interessante que uma análise feita na Europa e EUA, sobre o uso de IA na advocacia, mostra que, salvo raras exceções, não são as tradicionais bancas de advogados, mas novos entrantes, que investem no conceito. Vemos que o momento do Uber, Airbnb, Skype e Whatsapp se repete. As empresas estabelecidas tendem a ser conservadoras e lutam para preservar seu modelo de negócios. Enfim, estamos diante de mudanças significativas na sociedade e praticamente nenhuma função ou setor de negócios estará a salvo das transformações. O cenário conturbado, como o contexto que vemos hoje envolvendo o Uber e os taxistas, certamente vai acontecer quando os escritórios de advocacia sentirem reais ameaças ao seu modelo atual. Mas, alguns entenderão que o processo é irreversível e os vencedores serão os que conseguirem fazer o mix certo entre advogados e tecnologia. O mesmo poderíamos dizer sobre o TCU. O conhecimento agregado torna-se realmente útil quando manipulado de forma ágil por motores de IA, apoiando o trabalho dos ministros.

Opinião

Os pilares da estratégia de análise de dados e consumo de informações no TCU

Eficiência e efetividade: palavras que resumem necessidades fundamentais impostas a todas as organizações, públicas ou privadas – ou seja, entregar resultados relevantes com a menor utilização de recursos. Fazer o melhor possível aquilo que deve ser feito torna eficiência e efetividade conceitos fortemente conectados. O desalinhamento entre os dois pode representar excelência ao inverso: estar fazendo muito bem algo que não deveria estar sendo feito. O TCU, com a missão de aprimorar a administração pública em benefício da sociedade por meio do controle externo, também deve buscar fazer mais e melhor o que a ele compete.

Especificamente sobre o TCU, enquanto auxiliar do poder de fiscalizar a legalidade e legitimidade dos gastos, bem como o bom funcionamento das políticas públicas, pode-se sintetizar eficiência e efetividade como a necessidade de “fazer melhor e escolher melhor o que fazer”. Dada a limitação de recursos, é cada vez mais importante ser certo na fase de planejamento, projetando ações de controle com base no risco, materialidade e relevância identificados em cada um dos objetos. O “fazer melhor” representa o bom funcionamento dos processos de trabalho relacionados às atividades de controle externo, em que se demanda o pleno funcionamento dos métodos e disponibilidade das ferramentas e de profissionais capacitados.

Para planejar e executar da maneira desejada, um dos insumos fundamentais é a informação. É o uso correto e tempestivo da informação que permite que as organizações tomem decisões mais apropriadas sobre o que fazer e como fazer.

Em que pese o desafio da eficiência e efetividade ser perene no mundo corporativo, o cenário atual nos exige pensar em novas alternativas. Considerando que a quantidade de sistemas de informação tem crescido a ponto de os principais atos e fatos administrativos (contratações, licitações, pagamentos, concessões de benefícios etc.) poderem ser representados digitalmente, informações sobre o funcionamento do Estado e de interesse do controle externo estão cada vez mais disponíveis, prontas para serem analisadas e consumidas.

A tendência de utilização maciça dos dados públicos está posta e é similar à realidade atual das organizações privadas. A informação está e estará cada vez mais abundante, distribuída e fluída, adaptável à interpretação de diversos atores. Por isso será cada vez mais difícil (se não impossível) reter as informações ou impedir que elas sejam combinadas para gerar outras. *A abundância das informações, somada às infinitas formas de combiná-las, nos impõe outro desafio: consumi-las.* Ter à sua disposição a rota mais rápida para se locomover de um ponto a outro em dias de chuva torrencial é alentador, mesmo sabendo que aquilo só ocorre porque existe uma comunidade de usuários que contribuem produzindo e consumindo informações, de maneira simples e barata. Efetivo e eficiente, não?

Considerando que a utilização correta da informação abundante sobre o funcionamento do Estado é condição fundamental para a boa gestão dos recursos públicos, como fazer que órgãos públicos, incluindo os de controle, consigam utilizar plenamente esse insumo para viabilizar o aumento da sua eficiência e efetividade a custos possíveis? Em outras palavras, o que é preciso fazer, de maneira pragmática, para que as instituições públicas se adaptem à realidade atual de exigência de melhores produtos e serviços e funcionem melhor, sob a ótica do uso de informação como recurso essencial? Como lidar com as dificuldades já conhecidas



Wesley Vaz Silva
Secretário

Secretaria de Gestão de Informações
para o Controle Externo



e avançar na direção da instrumentalização dos processos de trabalho com informações necessárias e suficientes para cumprirem a sua finalidade?

A seguir, apresenta-se três pilares inspirados no funcionamento de instituições públicas e privadas nacionais e internacionais que têm servido como base da estratégia do TCU, visando o pleno uso da informação em prol do controle. A execução de ações coordenadas tem permitido ao TCU avançar na utilização de técnicas de análise de dados, na produção e no consumo de informações úteis ao planejamento e à execução das ações de controle. Acredita-se que os fundamentos aqui apresentados possam ser considerados por outros órgãos de controle, bem como por qualquer instituição pública brasileira que deseje repensar seu funcionamento a partir do consumo intensivo de informação.

Duas premissas foram fundamentais para o estabelecimento dos pilares. Ambas colidem frontalmente alguns mitos da Administração Pública e, por essa razão, podem ser contraintuitivas em uma primeira leitura. Aos mitos:

1. A estabilidade. Há que se questionar se, como era (ou ainda é) amplamente entendido, as

instituições públicas gozam de uma estabilidade institucional e de recursos suficientes para executar as políticas públicas sob sua responsabilidade. O ambiente corporativo mundial, público e privado, é de incerteza, submetido a mudanças constantes e cada vez mais drásticas.

2. A infalibilidade. É preciso reconhecer que soluções complexas estão naturalmente sujeitas à falhas e insucessos. Em determinados ambientes, em especial para as instituições públicas que atuam em ambiente altamente regulado, as falhas podem gerar desconforto e, em última instância, responsabilização. A reação a isso se impõe na forma de inibição ao surgimento de novas iniciativas por medo do fracasso. Nesse contexto, uma nova relação com os riscos deve ser estabelecida sem que se afronte, por óbvio, o princípio da legalidade ao qual todas as iniciativas públicas se vinculam.

Portanto, é preciso encontrar soluções que forneçam novos produtos e serviços para o bom funcionamento das instituições,

caracterizadas pela inovação em *um ambiente de incerteza*. Apresentado por Eric Ries no seu livro *Lean Startup*, esse é precisamente o conceito de startup, sobre o qual os pilares de fomento ao uso pleno de informações pelo controle e para o controle estão sendo sedimentados na estratégia do TCU. Vamos aos pilares:

I. Governança

Apoio da alta administração e liderança. Esses costumam ser os viabilizadores mais importantes da governança corporativa, por representarem os pré-requisitos para as diretrizes superiores que suportam a missão. Sem o claro apoio da alta direção, iniciativas transformadoras não são mais do que boas ideias prestes a serem vencidas pelo cansaço.

Contudo, somente o direcionamento não basta. É preciso que as diretrizes se transformem em estratégias. No caso do TCU, que contou com direcionamento voltado ao estímulo da utilização de informações para a atividade de controle, a escolha foi a de incentivar a descentralização das atividades de análise de dados para as unidades técnicas. As secretarias de controle externo passaram a capacitar auditores do seu corpo para iniciar os trabalhos de análise

de dados, bem como a buscar acesso a informações úteis ao seu trabalho, utilizar ferramentas tecnológicas de suporte e compor um grupo de interesse e pesquisa no assunto.

A soma do apoio da alta direção com a formação de times de auditores tem gerado bons resultados e reforçado o fomento à construção de uma nova cultura, em que a descentralização e a coordenação entre os papéis estimulam o consumo de informações para as atividades de controle externo. O sucesso dessa estratégia depende, entre outros fatores, da clareza dos papéis e responsabilidades de cada um dos envolvidos, da confiança entre os atores, da experimentação e testes ágeis de procedimentos sobre os dados e da avaliação rápida dos resultados, positivos ou negativos.

Em dois anos, várias unidades técnicas do TCU lideraram trabalhos de fiscalização utilizando técnicas de análise de dados. Sem a pretensão de ser exaustivo, cita-se:

- Acórdão 539/2015-Plenário, derivado de Fiscalização de Orientação Centralizada liderada pela Secretária de Controle Externo do Mato Grosso (Secex-MT) e sob a relatoria do ministro Benjamin Zymler, que tinha como objetivo validar a seleção de objetos de controle com base em análise preditiva de dados relativos a transferências voluntárias.
- Acórdão 718/2016-Plenário, realizado pela Secretaria de Controle Externo da Previdência, do Trabalho e da Assistência Social (SecexPrevi) sob a relatoria do Ministro Vital do Rêgo, que objetivava estruturar fiscalização contínua de benefícios previdenciários quanto à sua concessão, manutenção e pagamento no âmbito do Instituto Nacional do Seguro Social.

Em 2016, algumas unidades técnicas do TCU (Secex-MT, SecexPrevi, Secex-CE e Secex Educação, por exemplo) formalizaram

a existência de um núcleo próprio de análise de dados, composto por auditores capacitados para utilizar técnicas de análise de dados nos seus trabalhos. O trabalho desses núcleos se baseia na colaboração e integração com as demais unidades de apoio do TCU, visando inserir nos processos de trabalho cotidianos novos métodos e ferramentas, com uma estrutura leve e voltada a resultados.

No âmbito externo, novos grupos institucionais têm surgido para intensificar a colaboração e os trabalhos conjuntos relacionados à análise de dados. Destaca-se a rede nacional de informações estratégicas para o controle externo – Rede Infocontas, criada em 2013 e que, recentemente, fomentou a criação de uma unidade de análise de informações em cada um dos tribunais de contas estaduais do Brasil. Em nível internacional, destaca-se a criação, no último congresso internacional das entidades de fiscalização superior realizado em dezembro de 2016, de um grupo de trabalho sobre Big Data e Analytics, composto por diversos países, entre eles o Brasil. O objetivo é estimular a colaboração e a troca de boas práticas, além de discutir e fomentar o desenho de novas estratégias e o uso de novos métodos e ferramentas baseadas em informação para a melhoria da efetividade e eficiência das entidades de fiscalização

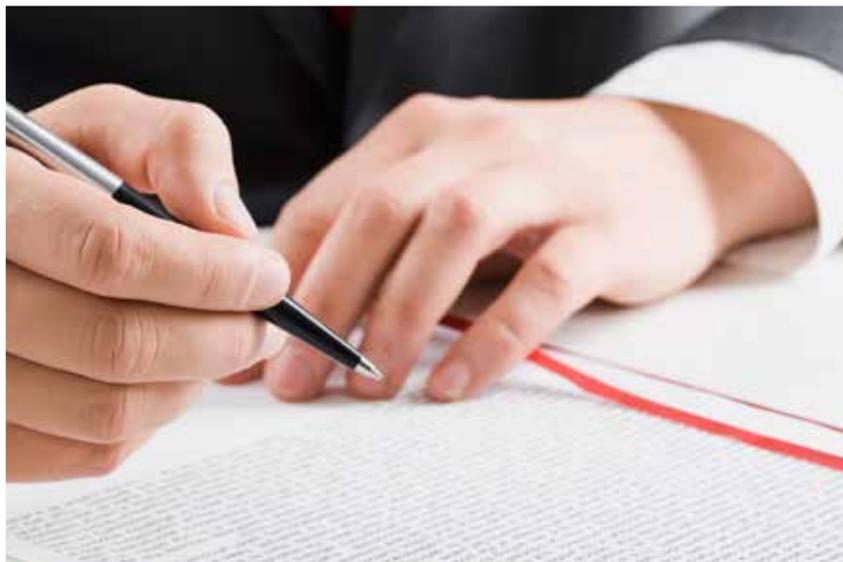
superior. Além disso, a exemplo do que já existe no TCU com o papel atribuído à Secretaria de Gestão de Informações para o Controle Externo (Seginf), as entidades de fiscalização superiores dos EUA, Índia e China criaram estruturas organizacionais responsáveis pelo estabelecimento e execução da estratégia corporativa para o uso de informações e de técnicas de análise de dados no âmbito das suas operações.

Estabelecidas as diretrizes, definida a estratégia, capacitados os profissionais e definidos os objetivos, é preciso conseguir as ferramentas e os insumos.

II. Plataforma

Nas palavras do pesquisador Doug Laney, a “informação só é útil quando pode ser plenamente utilizada”. Se informação é o insumo, a matéria-prima, o seu pleno uso não depende somente da sua existência ou disponibilização. A tarefa de tornar verdadeiramente simples o consumo da informação implica desenvolver novos métodos, ferramentas e competências profissionais, mas antes é preciso obter as informações e torná-las plenamente disponíveis para o uso.

Essa é a essência do Labcontas, ambiente virtual criado e gerenciado pelo TCU, que permite aos auditores terem acesso a



informações internalizadas a partir de dezenas de bases de dados oriundas de acordos de cooperação entre instituições da Administração Pública Federal e o próprio TCU.

No intuito de se consolidar como uma plataforma colaborativa de fato, o Labcontas também é utilizado por parceiros externos, principalmente tribunais de contas estaduais (TCEs). Atualmente, já existem 25 desses parceiros, sendo 20 TCEs, que utilizam os softwares e as informações públicas da plataforma e, em contrapartida, contribuem trazendo informações públicas sobre o funcionamento de instituições de sua jurisdição e de interesse mútuo dos parceiros de controle (contratações de determinado estado da federação, por exemplo).

A existência de um ponto central que congrega informações e softwares para todos os seus profissionais colaboradores é primordial para que os trabalhos de análise de dados e consumo de informação sejam mais eficientes. O recente estímulo à abertura de dados no poder público, consubstanciado com a publicação de normas que apoiam a troca de informações entre os órgãos públicos (vide Decreto 8.789/2016) e somado ao aumento na quantidade das informações disponíveis (nem sempre de qualidade), sugere o surgimento de iniciativas estruturantes visando a *construção de uma plataforma aberta de dados públicos* do Estado, voltada à análise avançada de informações por parte das organizações públicas e que serviria como ponto de contato para as iniciativas tanto no âmbito das ações de controle como na intenção de se transformar em instrumento efetivo de gestão.

Com o ambiente pronto, informações disponíveis e estrutura montada, é hora dos produtos.

III. Soluções baseadas em informação

As organizações globais que consolidaram seu funcionamento a partir de produção e consumo de informações possuem verdadeira neurose em tentar manter as suas informações prontas para uso da maneira mais simples possível e a

um custo marginal, próximo de zero. Ou seja, o principal objetivo é permitir que as informações sejam muito fáceis de usar e tenham baixo custo.

O ponto é que, por mais que tenhamos equipes qualificadas e dados a nossa disposição, se o intuito é fomentar o ecossistema interno e externo de consumo de informações, é preciso construir soluções que resolvam o problema (seja planejar uma fiscalização ou ter insumos para responsabilizar ou não determinado gestor) de maneira muito simples, na visão daquele que usará esse insumo para tomar decisões. Mesmo que as soluções utilizem algoritmos complexos baseados em aprendizado de máquina e processamento cognitivo ou técnicas mais recentes de indexação e cálculo de relevância do conteúdo, o que deve importar para o auditor são os resultados que se obtém desses produtos e o quão confiáveis eles são para cada propósito.

No âmbito do TCU, cabe citar o DGI Consultas, solução que provê uma interface de busca às informações custodiadas e produzidas pelo TCU de maneira simples e direta e que exemplifica o modo como um grande volume de informações pode ser acessado por auditores de maneira muito simples e em prol do seu trabalho cotidiano. Inspirado nesse exemplo e viabilizado a partir de uma cooperação com o Ministério da Transparência, Fiscalização e Controladoria-Geral da União, o ALICE (Analisador de Licitações, Contratos e Editais) produz e envia mensagens eletrônicas automáticas contendo um apontamento de riscos sobre as licitações publicadas no dia anterior, considerando aspectos como os valores envolvidos e indícios de irregularidades obtidos diretamente dos textos dos editais em comparação com a jurisprudência do TCU.

Esses são dois exemplos de soluções baseadas em informação que servem como mais uma ferramenta de apoio aos processos de planejamento baseados em risco, além de fornecer insumos relevantes durante a execução de fiscalizações e instruções de processos por meio da utilização abrangente das informações internas e externas à

disposição do TCU. Não tardará o tempo em que até mesmo análises preliminares automáticas das evidências possam – por que não? – ser entregues ao auditor, de forma a subsidiá-lo no seu trabalho de avaliação.

O reconhecimento da importância da utilização plena das informações nas instituições públicas de controle é um fato, mas que carrega consigo muitas dúvidas. A maturidade heterogênea das organizações e o passivo de problemas que afeta o seu pleno funcionamento tomam qualquer estratégia de mudança na forma de trabalhar muito desafiadora. Contudo, as exigências atuais não nos parecem fornecer alternativas a não ser adentrar na realidade que se impõe da quarta revolução industrial, do pensar digital, em que o consumo de bens e serviços é caracterizado pela exigência do baixo custo e da qualidade. E esses desejos da sociedade afloram e, cada vez mais, são direcionados para o funcionamento do Estado.

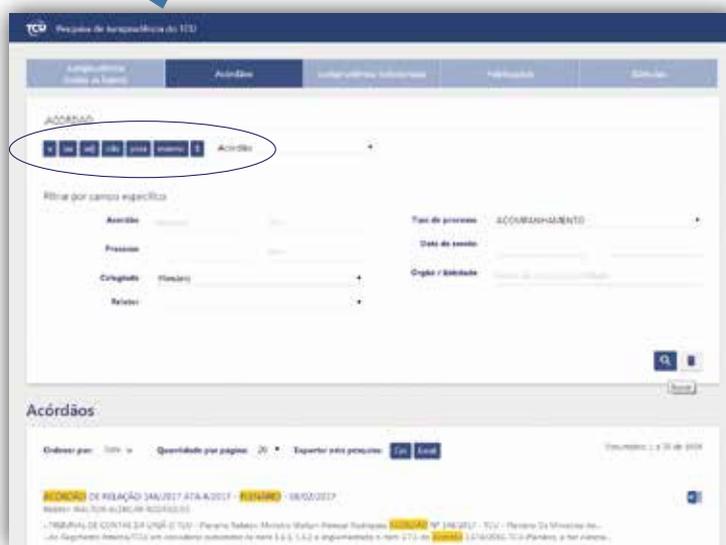
A necessidade de se moldar a essa nova realidade impõe aos profissionais públicos, além de uma análise crítica sobre o senso comum do funcionamento das organizações públicas, uma honesta reflexão sobre seu papel enquanto parte de uma organização que nunca estará suficientemente pronta, mas que precisa entregar mais e funcionar melhor.

De modo pragmático, já é possível constatar bons resultados oriundos da execução da estratégia de utilização de informações e análise de dados nos processos de trabalho do TCU. Afora os alarmismos de tempos em tempos, mas mantendo viva a capacidade de se inspirar e se comparar a uma realidade global e local, não há alternativa senão o constante repensar a maneira como a missão das instituições públicas deve ser plenamente cumprida. O aumento desejado da eficiência e da efetividade certamente envolve o *pleno uso da informação*, que demanda o envolvimento e o aumento da interação entre profissionais de perfis complementares, a utilização de novas tecnologias, a gestão das informações e, principalmente, uma mudança no modo de planejar e executar os processos de trabalho.

Nova Pesquisa de Jurisprudência

Mesmo padrão adotado pelos Tribunais Superiores

Pesquisa mais simples, rápida e precisa



Destaque entre as principais mudanças:

- Pesquisa simultânea em todas as bases de dados de jurisprudência;
- Operadores lógicos utilizados pelos tribunais superiores (“e”, “ou”, “adj”, “não”, “prox”, “mesmo” e “\$”);
- Novos recursos: histórico de pesquisa, índice e navegação pelo resultado;
- Na Jurisprudência Seleccionada: pesquisa por referência legal, navegação por árvore de classificação e utilização do Vocabulário de Controle Externo (VCE) para possibilitar a pesquisa por sinônimos.
- Saneamento do conteúdo das bases.

Fundamentos de uma auditoria em realidade misturada



Aroldo Cedraz de Oliveira

é presidente do Tribunal de Contas da União, doutor em Medicina Veterinária pela Universität Hannover – Alemanha, mestre em Medicina Veterinária pela Universidade Federal de Santa Maria/RS, além de Professor Honorário (h.c.) da Universidade de Auditoria de Nanjing (NAU) e ex-congressista por quatro legislaturas. É coordenador de edição da obra “O controle da administração na era digital”.



Francisco Osório Ramos

é servidor do Tribunal de Contas da União, formado em Engenharia Elétrica pela UnB, com mestrado em Física pela UnB. Professor de Física da UnB, de Engenharia Elétrica da UNIP e do IESB e de Ciência da Computação do UniCEUB, ministrando as disciplinas Inteligência Artificial, Computação Gráfica e Arquitetura de Computadores.

RESUMO

Com suporte em informações colhidas no site da *National Aeronautics and Space Administration–NASA* e do *Jet Propulsion Laboratory–JPL*, são analisadas as técnicas de computação gráfica empregadas no projeto *OnSight–NASA*, aplicando-as à auditoria de obras públicas. Com base nesta reconstrução teórica, estabelecem-se os alicerces de uma nova modalidade de auditoria, denominada *MRA–Mixed Reality Auditing*, ou auditoria em realidade misturada. Esta consiste em uma fiscalização remota, em tempo real, que utiliza imagens de câmeras digitais no local de fiscalização, transmitidas para *Head-mounted Displays–HMDs*, ou óculos especiais de realidade misturada, utilizados por um observador situado em uma localidade distinta.

Palavras-chave: Computação gráfica. Realidade misturada. Renderização. *Frustum*. *Quadtree*. Motor de jogo. Fiscalização de obras.

1. INTRODUÇÃO

O *Jet Propulsion Laboratory–JPL*, um centro tecnológico de pesquisa e desenvolvimento ligado à NASA, começou a oferecer ao público, a partir de março de 2016, um passeio dirigido em uma área do planeta Marte (JPL, 2016). Trata-se da missão *Destination: Mars*, uma exibição interativa utilizando óculos de realidade misturada, um dos desdobramentos da realidade virtual. Os visitantes



podem percorrer vários locais de Marte, reconstruídos a partir de imagens reais enviadas para a Terra pelo veículo *Curiosity*, um jipe robô motorizado que explora a superfície do planeta vermelho desde agosto de 2012 (Ibid., 2016).

Destination: Mars é uma adaptação do projeto *OnSight*, uma ferramenta de software para missões em Marte. Os dados e imagens transmitidos diariamente pelo veículo *Curiosity* são captados na Terra por três antenas gigantes de 80m, sendo utilizado um dispositivo de realidade misturada para transportar essas informações da superfície de Marte até um laboratório na Terra. Os cientistas são envolvidos pelas imagens captadas nas proximidades do veículo em Marte, podendo passear em torno de superfícies rochosas ou agachar-se para examinar formações geológicas de diferentes ângulos. O *OnSight* fornece aos cientistas um meio de planejar e, juntamente com o veículo *Curiosity*, conduzir operações em Marte, explorando o planeta diretamente de seus escritórios em Pasadena, na Califórnia (Ibid., 2016).

2. HMDS E REALIDADE DIGITAL

HMD ou *Head-mounted display* é um dispositivo de vídeo usado sobre a cabeça como um capacete, contendo amplo visor, fones de ouvido e uma interface através da qual o usuário pode experimentar um ambiente de realidade virtual (MELZER; MOFFITT, 2011). Consiste em um visor transparente montado na cabeça do observador, de onde a informação é apresentada sem exigir que o usuário olhe para fora de seu campo de visão normal, como ilustra a figura 1 (CANON, 2016). Desde os estudos pioneiros,

o campo de pesquisa da realidade virtual tem utilizado HMDs como dispositivos essenciais de exibição visual.

Figura 1:

Dispositivo de vídeo HMD



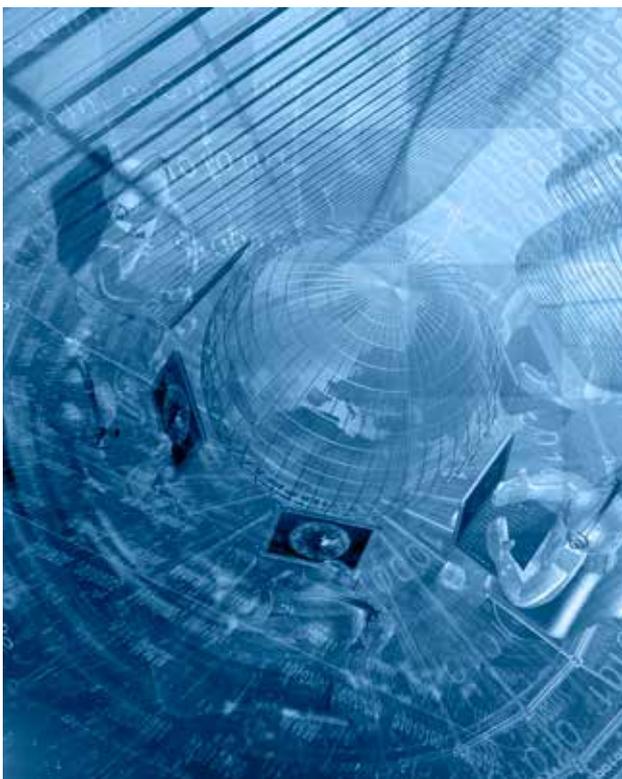
Fonte: CANON (2016)

A realidade virtual consiste em uma realidade multimídia imersiva ou simulada por computador que replica um ambiente, simulando a presença física em locais do mundo real ou de um mundo imaginário, permitindo ao usuário interagir com este mundo específico. Em outras

palavras, trata-se de um termo genérico, aplicável a todos os tipos de experiências imersivas, criadas usando somente conteúdo do mundo real, sintético, ou um híbrido de ambos (LACKEY; SHUMAKER, 2016). São exemplos de óculos de realidade virtual o Oculus Rift, do Facebook; o Gear VR, da Samsung; e o Projeto Morpheus, da Sony.

Ainda segundo Lackey e Shumaker (2016), realidade aumentada é uma visão direta ou indireta de um ambiente físico do mundo real, cujos elementos são aumentados ou suplementados com entradas geradas por sensores, tais como som, vídeo, gráficos ou dados de GPS. Trata-se de uma sobreposição de conteúdo sintético sobre o mundo real, em que este conteúdo não é ligado a ele ou mesmo parte dele. O conteúdo de mundo real e o de computação gráfica não são capazes de interagir um com o outro. Constituem-se em óculos de realidade aumentada o Google Glass, o Daqri Smart Helmet e o Moverio, da Epson.

Realidade misturada, também denominada realidade híbrida, consiste em uma fusão entre os mundos real e virtual para produzir novos ambientes e visualizações, onde objetos físicos e digitais coexistem e interagem em tempo real. Trata-se de uma sobreposição de conteúdo sintético sobre o mundo real do usuário, ligado a ele ou que com ele interage (Ibid., 2016). Contam-se entre os óculos de realidade misturada o HoloLens, o Canon MREAL Display MD-10 e o Magic Leap.



Modernamente, admite-se que as realidades virtual, aumentada e misturada podem ser reunidas sob uma denominação mais genérica: realidade digital.

3. NÍVEL DE DETALHE E RENDERIZAÇÃO

A visão estereoscópica por computador (*stereovision*) consiste na extração de informação 3D de imagens, como aquelas obtidas por uma câmera digital. Comparando a cena a partir de dois pontos selecionados, a informação 3D pode ser extraída a partir do exame das posições relativas de objetos nos dois cenários. As informações detectadas a partir destas imagens estereoscópicas são lançadas sobre uma superfície livre, criada com esta finalidade (AKENINE-MÖLLER; HAINES; HOFFMAN, 2008).

Duas câmeras, afastadas horizontalmente uma da outra, são usadas para obter duas visões diferentes de uma cena, de forma semelhante à visão humana binocular. Comparando estas duas imagens, a informação de profundidade relativa pode ser obtida na forma de um mapa de disparidade (*disparity map*), que codifica a diferença entre as coordenadas horizontais de pontos correspondentes nas imagens. Os valores neste mapa de disparidade são inversamente proporcionais à profundidade da cena na localização correspondente dos pixels (Ibid., 2008).

Utilizando a visão estereoscópica por computador, para cada imagem que chega ao HMD do observador, é gerado um modelo 3D completamente novo por meio de algoritmos automáticos que constroem um modelo matemático e estatisticamente acurado da superfície analisada, denominada correlação estereoscópica (COLANER, 2016). Observando duas imagens, encontrando as diferenças entre elas com auxílio de um mapa de disparidade e usando um modelo específico de câmera, conclui-se quão distante cada ponto se encontra e geram-se os mapas de alcance destas imagens.

Com a finalidade de acelerar o processamento das imagens tratadas, todos os objetos que se encontram fora do campo de visão do observador são removidos no processo de renderização (processo de gerar uma imagem a partir de um modelo 2D ou 3D por meio de processamento digital, consistindo em três etapas básicas: determinação da câmera virtual, da superfície visível e das fontes de luz). Somente os pixels que são visíveis no *frustum* necessitam ser renderizados. Todos os objetos que são visíveis a partir de um ponto em particular são pré-computados, sendo imediatamente removidos todos os objetos não visíveis, reduzindo-se dessa forma, o número de objetos que interceptam o *frustum* (Ibid., 2016).

Frustum é o campo de visão a partir do olho do observador, podendo ser representado por um volume espa-

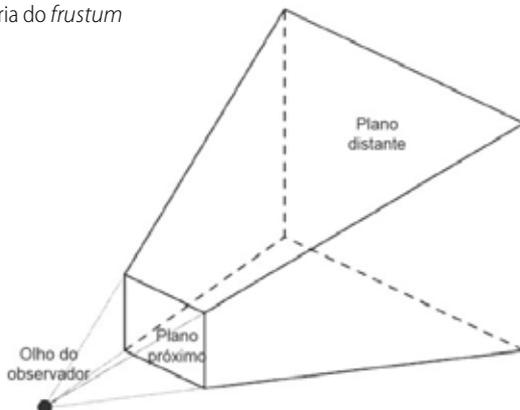
cial imaginário contendo tudo o que é visível em uma cena tridimensional. Pode ser representado por uma pirâmide truncada constituída por seis planos. Quatro desses planos correspondem aos lados da tela, sendo chamados de *frustum* direito, esquerdo, de base e de topo. Os dois planos restantes são chamados planos *frustum* próximo e distante, e definem as distâncias mínima e máxima em que os objetos de uma cena são visíveis para o observador (DUNN; PARBERRY, 2002). A figura 2 ilustra a geometria do *frustum*.

Para gerenciar o nível de detalhe, a superfície da cena é dividida em ladrilhos ou *tiles* de diferentes tamanhos, cada um com um determinado nível de detalhe, seguindo uma divisão da superfície (espaço 2D) conhecida como *quadtree* (COLANER, 2016).

O *quadtree* consiste em uma estrutura de dados em árvore, na qual cada nó interno possui exatamente quatro filhos, sendo utilizada para particionar um espaço bidimensional (2D), por exemplo, uma superfície mediante subdivisão recursiva em quatro quadrantes, regiões ou células adaptáveis. Cada célula ou reservatório possui uma capacidade máxima. Quando esta capacidade é atingida, o reservatório sofre uma divisão. A estrutura da árvore segue a decomposição espacial do *quadtree*, ou seja, cada nó sofre uma divisão em quatro sub-nós (LENGYEL, 2004).

A figura 3 exibe um *quadtree* construído para uma área contendo um único objeto. A ilustração à direita mostra como a estrutura de dados correspondente é organizada. Cada nó possui quatro sub-nós. Se nenhuma geometria de mundo interceptar um quadrante, então este quadrante não é subdividido. Qualquer quadrante que não contém objetos é removido da árvore. Assume-se também que quaisquer quadrantes ausentes estão vazios.

Figura 2:
Geometria do *frustum*



Fonte: Dunn e Parberry (2002)

Segundo Lengyel (2004), organizar a geometria em um *quadtree* traz o benefício de que sempre que um nó da árvore puder ser determinado como não visível, tem-se a informação imediata de que todos os sub-nós daquele são também não visíveis e podem ser eliminados, melhorando o gerenciamento do nível de detalhe.

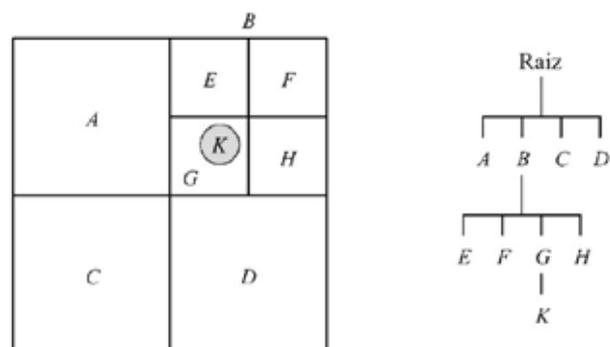
4. O MOTOR DE JOGOS

Um motor para criação de jogos digitais como o *Unity*¹ gera o software para integração com o hardware envolvido, mostrando-se adequado para a interação do usuário com a realidade misturada, visto que os óculos disponíveis possuem uma interface com o referido motor. Utiliza-se um *frame rate* de 60 fps (quadros por segundo). *Frame rate* é a frequência (*rate*) com que um dispositivo de processamento de imagens as exibe de forma consecutiva, sendo estas chamadas de *frames* (quadros).

O motor de jogos *Unity* emprega a técnica de *mipmaps*, que consistem em sequências de texturas do mundo 3D otimizadas e pré-calculadas, sendo cada uma delas uma representação em resolução progressivamente mais baixa da mesma textura, o que incrementa a velocidade de renderização. Texturas *mipmaps* de alta resolução são usadas para amostras de alta densidade, como objetos próximos à câmera. Texturas de baixa resolução são usadas quando o objeto aparece mais distante. A técnica de *mipmap* pode melhorar o desempenho de processamento em até 33% no *Unity*.

Dessa forma, para os *tiles* mais próximos do observador, a resolução torna-se máxima. Para *tiles* mais distantes de seu campo de visão, a resolução vai diminuindo exponencialmente. Quando o observador dirige seu olhar mais para frente, as texturas de resolução menor

Figura 3:
Particionamento de um espaço 2D com *quadtree*



Fonte: Lengyel (2004)

são sobrepostas às texturas de mais alta resolução. Com esta técnica, imagens locais captadas por câmeras do veículo *Curiosity* podem ser combinadas com imagens de satélites orbitando Marte no local de observação, já que estas exibem uma resolução que é constante na região (COLANER, 2016).

Adicionalmente, se o observador clicar em algum detalhe da imagem podem ser exibidas outras imagens com zoom, tomadas de outro ângulo ou mesmo de um satélite orbital. Por seu turno, os óculos de realidade misturada (HMDs) suportam olhar fixo (*gaze*), gestos e comandos de voz.

5. AUDITORIA EM REALIDADE MISTURADA

A auditoria em realidade misturada consiste na realização de uma fiscalização em que o auditor não está presente no local da obra, encontrando-se distante, em outra localidade. Os sinais eletromagnéticos, relativos às imagens da superfície que são captadas por câmeras digitais no local, são transmitidos para o HMD do auditor que recebe as referidas imagens e consegue percorrer a obra virtualmente e em tempo real, mesmo estando a distância, por exemplo, em uma sala na sede do TCU. Assim, auditores do TCU podem realizar fiscalizações de obras em locais distantes de Brasília sem abandonarem a sede na Capital Federal.

O auditor é, portanto, envolvido pelas múltiplas imagens captadas nas proximidades da obra, todas com elevado nível de detalhe, além de várias posições e ângulos diferentes, podendo passear em torno dessa superfície ou

até agachar-se para examinar detalhes de diferentes perspectivas e reunir achados de auditoria.

Para o caso de fiscalizações em obras rodoviárias, podem ser observados remotamente, utilizando a Auditoria em Realidade Misturada–MRA, os seguintes defeitos na rodovia: formação de painelas; afundamentos; escorregamento de aterros; desgaste e fissuras em revestimentos; dispositivos de drenagem danificados; exsudação em trecho de curva; remendo; degradação, ruptura e erosões de borda etc.

No que diz respeito às fiscalizações em obras ferroviárias, a observação visual em MRA pode detectar os seguintes defeitos, tanto nas obras de terraplenagem e de artes correntes e especiais, quanto na plataforma ferroviária e na via permanente: apodrecimento; fratura ou excesso de furos em dormentes de madeira; queima por patinação; e corrosão ou flambagem da via em trilhos de aço. Pode ocorrer ainda deslocamento longitudinal dos trilhos (ou arrastamento), desnivelamento nas juntas, desaparecimento de folga entre eles, ocorrência de vazio ou bolsa de contração, bem como o seu desalinhamento.

Quanto a fiscalizações de obras de edificações com MRA, podem ser verificadas irregularidades na locação da obra (construção da estrutura na posição incorreta); na terraplenagem (escavação, transporte e aterro); nas fundações (fissuras nos tubulões ou nas sapatas); nas estruturas de concreto armado ou protendido (incorrecções nos prumos de cantos externos, pilares e poços de elevadores); na alvenaria de vedação (trincas ou rachaduras, utilização de tijolos danificados, locação das paredes e dos vãos das esquadrias); na cobertura



(vazamentos, telhas quebradas, inclinação do telhado e calhas), bem como nos revestimentos (placas soltas).

Podem ser inseridas no software de realidade misturada todas as informações relativas aos projetos básico e executivo, como quantidade de materiais, qualidade dos mesmos e preços por unidade para serem conferidas com as constatações visuais das obras sob fiscalização, simplesmente encarando um alvo e utilizando controles baseados em gestos para selecionar os comandos na barra de menu. Da mesma forma, pode ser inserida uma régua virtual para medir a distância entre dois pontos da obra.

No âmbito internacional, a auditoria de recursos provenientes de auxílios, no caso de grandes desastres naturais, constitui-se em um tema recorrente no *International Journal of Government Auditing* da INTOSAI. Com as técnicas de auditoria em realidade misturada, especialistas do mundo inteiro, entre auditores de EFSs, geólogos, sismólogos, físicos e ambientalistas poderão acompanhar as ocorrências *in loco*, sem abandonar seus países, simultaneamente e em tempo real.

6. CONCLUSÃO

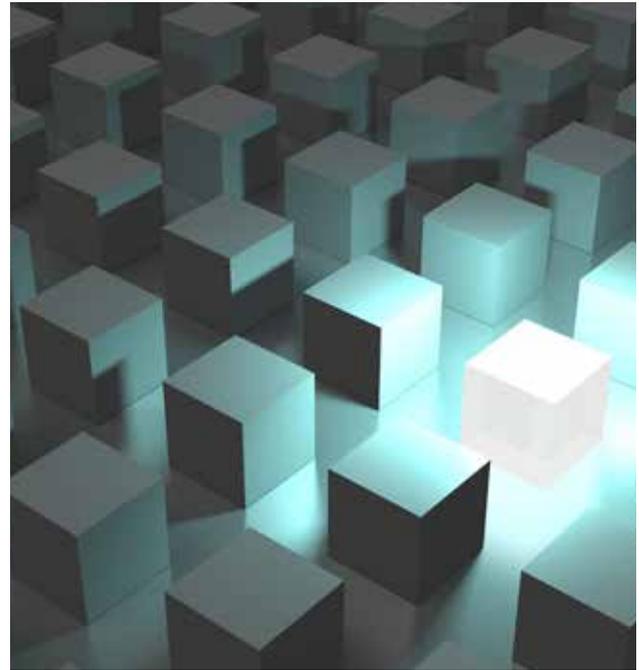
Este estudo teve como objetivo analisar as técnicas de computação gráfica empregadas no projeto *On-Sight- NASA*, aplicando-as à auditoria de obras públicas. Com esteio nesta reconstrução teórica, estabeleceram-se os alicerces de uma nova modalidade de auditoria denominada *MRA – Mixed Reality Auditing*, ou auditoria em realidade misturada.

A criação deste novo conceito de auditoria pelo Tribunal de Contas da União constitui-se em um grande avanço no campo da fiscalização de obras públicas e ambiental, com impacto imenso na forma como este trabalho é realizado até o momento, porquanto suprime as distâncias e a necessidade de viagens de especialistas aos locais de obras e desastres naturais, com evidente economia de tempo e recursos para viagens e hospedagens, bem como demais despesas concernentes.

As bases lançadas no presente estudo servem como diretrizes para que o TCU realize parceria com a NASA, efetivando projeto conjunto na área de interesse ou mesmo inicie projeto próprio, incluindo o desenvolvimento do software necessário, bem como a seleção do HMD mais adequado e a consequente integração com o hardware a ser utilizado.

NOTAS

1 Disponível em: <<http://unity3d.com>>.



REFERÊNCIAS

- AKENINE-MÖLLER, T.; HAINES, E.; HOFFMAN, N. *Real-Time Rendering*. 3rd ed. Massachusetts: A K Peters; Florida: CRC Press, 2008.
- CANON. MREAL Display MD-10, 2016. Disponível em: <<https://goo.gl/wBeCih>>. Acesso em: 14 set. 2016.
- COLANER, S. VR and AR go to Mars: Interview with NASA Scientists Jeff Norris and Alex Menzies. *Tom's Hardware*, New York, April 7, 2016. Disponível em: <<https://goo.gl/gvLN7H>>. Acesso em: 14 set. 2016.
- DUNN, F.; PARBERRY, I. *3D Math Primer for Graphics and Game Development*. Texas: Wordware Publishing, Inc., 2002.
- JPL. *Mixed Reality Technology Brings Mars to Earth*. Jet Propulsion Laboratory–NASA, California, March 30, 2016. Disponível em: <<https://goo.gl/RYrCrZ>>. Acesso em: 14 set. 2016.
- LACKEY, S.; SHUMAKER, R. *Virtual, Augmented and Mixed Reality: 8th International Conference on Human-Computer Interaction*. 1st ed. New York: Springer, 2016.
- LENGYEL, E. *Mathematics for 3D Game Programming & Computer Graphics*. 2nd ed. Massachusetts: Charles River Media Group, 2004.
- MELZER, J. E.; MOFFITT, K. *Head-mounted Displays: Designing for the User*. South Carolina: Create Space Independent Publishing, 2011.

Plataforma de serviços de dados abertos do TCU: *crowdsourcing*, nuvem cívica e aplicativos cívicos



Monique Louise de Barros Monteiro

é servidora do Tribunal de Contas da União, mestre e bacharel em Ciência da Computação pela UFPE. Possui 12 certificações profissionais: Big Data Certified Science Professional, Big Data Certified Professional, SOA Certified Governance Specialist, SOA Certified Architect, SOACP, PSM, SAFe Agilist, Oracle Master Java EE 5 Enterprise Architect, MTS, MSP, SCJP e IBM OOAD.



Marcelo Pacote

é servidor do Tribunal de Contas da União, mestre e bacharel em Ciência da Computação pela UnB. Possui especialização em Desenvolvimento de Aplicativos Móveis pela Uniara, além de 17 certificações profissionais: PMP, CBDP, SOACP, SOACC, CSM, CSD, SCJD, SCJP, SCJA, SCEA (I), SCBCD, SCWCDD, RUPF, IRIP, CTFI, ITILF e Oracle SQL Expert. Autor de livro na área de Tecnologia da Informação.



Remis Balaniuk

é servidor do Tribunal de Contas da União, pós-doutor em Realidade Virtual pela Stanford University, doutor em Informática pelo Institut National Polytechnique de Grenoble - França, mestre em Ciências da Computação pela Universidade Federal do Rio Grande do Sul e tecnólogo em Processamento de Dados pela Universidade de Brasília. Atualmente é professor e pesquisador da Universidade Católica de Brasília.



RESUMO

Este artigo descreve o trabalho de desenvolvimento, disponibilização e sustentação de uma plataforma de exposição de dados abertos e de coleta de informações por meio de *crowdsourcing*, tendo por base seu uso por aplicativos cívicos. Como resultados esperados da implementação dessa plataforma, podemos citar as oportunidades de 1) centralização e disponibilização de dados abertos e 2) obtenção de contribuições e percepções dos cidadãos. Tais informações poderão ter grande utilidade tanto para a definição de indicadores para serviços e políticas públicas como também para aumentar o conhecimento a respeito dos anseios do cidadão em diferentes localidades e esferas populacionais, incluindo a percepção da sociedade com relação aos serviços a ela prestados.

Palavras-chave: aplicativos cívicos, *web services*, *crowdsourcing*, nuvem cívica, dados abertos

1. INTRODUÇÃO

Um tema recorrente na Administração Pública no Brasil e em outras nações é o compartilhamento de dados abertos para maior transparência em relação às ações governamentais.

Atualmente, temos no Brasil, por exemplo, o Portal Brasileiro de Dados Abertos, ferramenta disponibi-

lizada pelo governo para que todos possam encontrar e utilizar os dados e as informações públicas. O portal inclui dados relativos à execução orçamentária, às informações cartográficas e sobre unidades de atendimento, além de possuir indicadores e estatísticas diversas.

Dentre os fatores responsáveis por catalisar iniciativas como o **Portal Brasileiro de Dados Abertos**¹ está a Lei 12.527 – Lei de Acesso à Informação Pública, sancionada em 18 de novembro de 2011. De acordo com as diretrizes fixadas pelo dispositivo legislativo em questão, a regra geral consiste em classificar as informações de interesse coletivo produzidas ou custodiadas pelo Estado como públicas, visando a assegurar o direito fundamental de acesso à informação. Assim, informações devem ser classificadas como sigilosas em casos excepcionais, quando for imprescindível à segurança da sociedade e do Estado.

O Tribunal de Contas da União (TCU) tem em seu planejamento ações para promover a divulgação de dados abertos por meio dos seus veículos de informação. No contexto desse conjunto de ações, desde o início de 2015, o TCU tem internalizado bases de dados provenientes do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), do Cadastro Nacional de Estabelecimentos de Saúde (CNES)², do Ministério do Desenvolvimento Social e Agrário (relativas à assistência social), do Sistema Nacional de Emprego (SINE), do Portal de Compras do Governo Federal (SIASG/Comprasnet)³, do Sistema Integrado de Admi-

nistração Financeira do Governo Federal (SIAFI)⁴, do Cadastro Nacional de Empresas (CNE)⁵, dentre outros.

Após a internalização das bases, os dados recebidos são disponibilizados para a sociedade, por meio de consulta por usuários finais ou por sistemas de software. A seguir, é detalhada a metodologia adotada para esta disponibilização dos chamados dados abertos.

2. METODOLOGIA

Tradicionalmente, ferramentas como o Portal Brasileiro de Dados Abertos, apesar de sua utilidade, têm apresentado como característica principal a disponibilização das informações de forma estática e/ou bruta (ex.: planilhas, arquivos em formato XML, CSV ou HTML), com grande quantidade de registros (ex.: planilhas com dezenas de milhares de linhas ou arquivos com centenas de *megabytes*). Porém, dados eletrônicos governamentais brutos são, em geral, de difícil compreensão para o cidadão comum. Faz-se necessária, portanto, a intervenção de pessoas, grupos e/ou empresas, com capacidade e disponibilidade para desenvolver aplicativos e outros elementos tecnológicos que possam traduzir o dado governamental aberto em produtos de interesse e utilidade pública ou privada.

Grande parte dos conjuntos de dados disponibilizados é de difícil apropriação também pelos desenvolvedores de tecnologias cívicas, já que essas bases de dados, oferecidas tipicamente em formato bruto, comumente têm um baixo valor informativo, além de serem disponibilizadas na forma de arquivos estáticos. Tal característica exige do desenvolvedor um trabalho periódico de transferência, entendimento, tratamento, adequação e posterior disponibilização em aplicativos. Também é requisito comum a necessidade de infraestrutura de TI remota e disponível continuamente, o que implica em custos de hospedagem (ex.: contratação de serviços de nuvens) com os quais os desenvolvedores precisam arcar. Tais dificuldades limitam o potencial de uso de dados abertos pelos aplicativos cívicos.

Ciente desse cenário, o TCU adotou um modelo de atuação no ecossistema de dados abertos que assumiu como premissa a necessidade de oferecer aos desenvolvedores de tecnologias cívicas uma plataforma de serviços na qual seus aplicativos possam acessar remotamente dados tratados e atualizados de uma forma mais adequada e otimizada para processamento por sistemas de software clientes. Outra premissa é que a plataforma oferecida permita aos aplicativos **armazenar** os dados por eles gerados.

Primeiramente, os dados recebidos a partir das bases citadas na introdução são persistidos em um banco relacional corporativo.

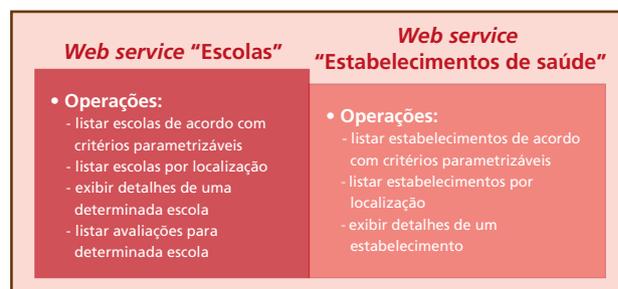
Concluída a etapa de persistência dos dados na infraestrutura de armazenamento do TCU, eles estão prontos para serem acessados tanto por usuários finais como por aplicativos através de *web services*. Segundo Erl et al. (2012), “a partir de uma perspectiva geral, um *serviço* é um programa de software que disponibiliza sua funcionalidade por meio de uma interface técnica publicada, denominada *contrato do serviço*”. Considerando que tais serviços de software podem ser implementados por meios diversos, utilizaremos neste artigo a denominação *web service* para nos referirmos a serviços disponíveis na *World Wide Web*.

Cada *web service* é responsável por agrupar e expor um conjunto de operações relacionadas. Tais operações podem ser acessadas por aplicações desenvolvidas externamente para a obtenção dos dados por meio de protocolos padronizados, de acordo com os padrões da *World Wide Web*. Tais protocolos incluem, por exemplo, *Hypertext Transfer Protocol* (HTTP) (FIELDING et al., 1999).

A figura 1 mostra o conceito de *web service* e suas operações em nível conceitual. Temos dois *web services* representados e cada um deles responde por determinado “tema”, agrupando operações de consulta que serão executadas sobre a base de dados persistida. Cada operação, por sua vez, admite o recebimento de parâmetros que permitem filtrar, dinamicamente, os crité-

Figura 1:

Representação em alto nível de dois *web services*. Um agrupa operações de consulta sobre escolas, e outro, operações de consulta sobre estabelecimentos de saúde





rios de busca, bem como os atributos e a quantidade de dados retornados.

Com relação ao formato dos dados retornados pelas operações, utilizamos JSON – *JavaScript Object Notation* (ECMA, 2013). Tal formato é compreensível por usuários finais, uma vez que é textual. Além disso, por ser semiestruturado, é amplamente utilizado por aplicativos para dispositivos móveis e outros tipos de sistemas computacionais, garantindo ampla interoperabilidade e facilidade de manipulação por desenvolvedores, principalmente em comparação com formatos proprietários, binários ou de maior complexidade, comumente encontrados em portais tradicionais de dados abertos.

3. DESENVOLVIMENTO

Os *web services* foram projetados inicialmente como ferramenta de apoio para o desenvolvimento de aplicativos móveis cívicos, no contexto do Projeto BE-PiD (Programa Educacional Brasileiro de Desenvolvimento para IOS), por meio de uma parceria entre o TCU e a Universidade Católica de Brasília (UCB)⁶. A parceria em questão envolveu o desenvolvimento de aplicativos cívicos por parte de alunos da referida universidade, que utilizassem os canais de informação fornecidos pelos *web services* desenvolvidos pelo TCU.

Foram especificados e desenvolvidos conjuntos de *web services* para disponibilização dos seguintes grupos de informações:

- **Saúde:** informações de especialidades médicas, estabelecimentos de saúde, profissionais, medicamentos e serviços especializados, todos no âmbito de serviços públicos.
- **Assistência social:** informações de centros de referência e centros de referência especializada de assistência social (respectivamente, CRAS e CREAS).
- **Empregos:** informações de postos do SINE.
- **Educação:** informações de educação, incluindo basicamente dados cadastrais dos estabelecimentos educacionais e de seus recursos.

Denominaremos o conjunto de *web services* acima citados de “nuvem cívica”. Daqui em diante, para nos referirmos especificamente ao conjunto de operações definidas no contrato ou interface disponibilizado pelos *web services*, utilizaremos o acrônimo API, sigla para *Application Programming Interface*.

A API inicialmente proposta e disponibilizada seguiu os requisitos esperados pelos aplicativos “Mapa da Saúde” e “Nossa Escola”, desenvolvidos por alunos da UCB.

Em virtude principalmente das necessidades de coletar *feedback* de cidadãos – prática conhecida mundialmente como *crowdsourcing* – os aplicativos foram desenvolvidos seguindo um modelo de rede social, de-

mandando, portanto, a existência de um cadastro de usuários e de avaliações de serviços públicos por eles fornecidas. A partir daí, surgiu o **metamodelo para aplicativos cívicos**, um conjunto adicional de *web services* que permite consultar, incluir, alterar e excluir informações relacionadas a usuários, aplicativos, grupos de usuários, *postagens*, *hashtags*, notificações etc.

Adicionalmente, para que o mesmo modelo pudesse ser adotado tanto para saúde pública como para educação, bem como para outras áreas de atuação do Estado, é possível que usuários enviem avaliações, sob a forma de **postagens**, de entidades genéricas denominadas aqui de **objetos**. Assim, um usuário pode avaliar um objeto, que pode ser, por exemplo, uma escola ou um hospital, a depender do contexto de utilização determinado pelo aplicativo em uso.

As massas de dados geradas pelo *crowdsourcing* cívico serão fontes inestimáveis de informação, em particular para a concepção de indicadores que permitirão a descoberta de conhecimento a respeito do funciona-

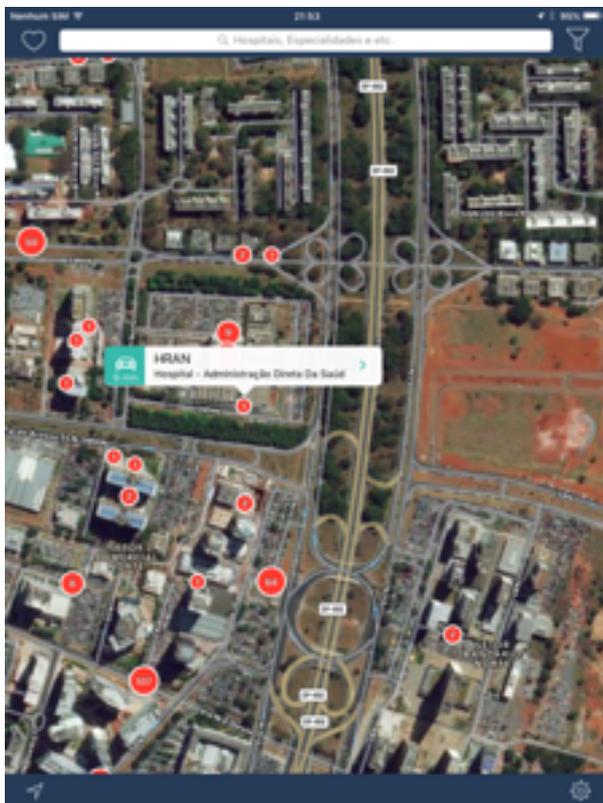
mento da máquina administrativa pública, das necessidades e da percepção do cidadão. O *crowdsourcing*, quando utilizado como uma prática de transparência, permite que grandes quantidades de dados ganhem sentido, podendo gerar ideias novas para o desenvolvimento de projetos para sociedade.

Esse conhecimento servirá futuramente como insumo no planejamento de ações de controle do TCU. Por fim, a coleta e o armazenamento dos dados em um formato definido pelo próprio TCU facilita iniciativas futuras de análise estatística e mineração de dados, tanto preditiva como prescritiva, podendo incluir desde análises mais simples até aplicações mais complexas baseadas em aprendizado de máquina.

Como exemplo de aplicativo cívico, as figuras 2 a 3 mostram capturas de tela do aplicativo Mapa da Saúde, que possui recursos de busca por estabelecimentos de saúde em um raio de localização, exibição de dados detalhados de cada estabelecimento (incluindo, embora não seja exibido nas capturas, listagem de especialida-

Figura 2:

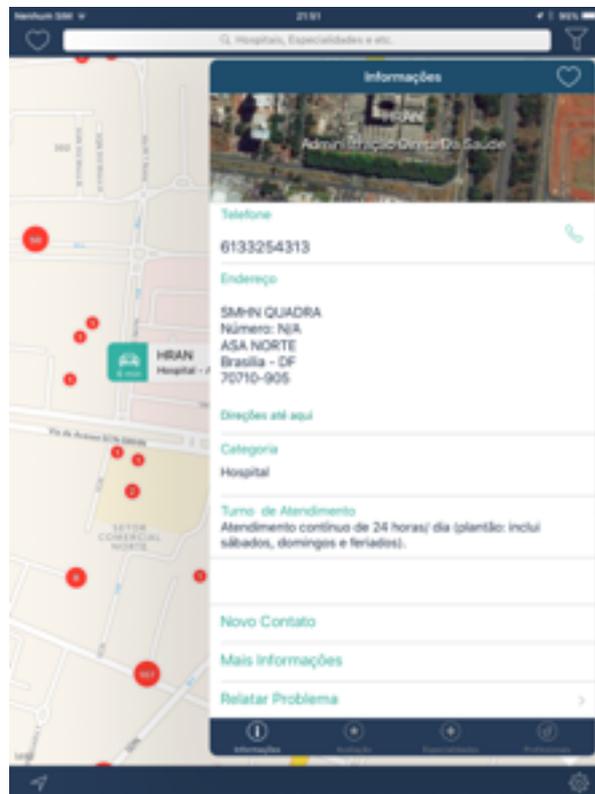
Mapa da Saúde – busca de estabelecimentos de saúde em raio, exibidos em um mapa (visualização em modo satélite)



Fonte: Elaboração própria

Figura 3:

Mapa da Saúde – detalhe de um estabelecimento



Fonte: Elaboração própria

des médicas e quantitativo de profissionais por função) e ainda uma funcionalidade para que o usuário avalie a qualidade do atendimento prestado.

4. ASPECTOS TECNOLÓGICOS

A nuvem cívica foi desenvolvida segundo o estilo REST – sigla para *Representational State Transfer* (FIELDING, 2000). REST engloba um estilo arquitetônico para a construção de aplicações e APIs e, grosso modo, provê um mecanismo no qual, no contexto dos *web services* aqui referenciados, cada operação de um *web service* pode ser representada da seguinte forma:

Método HTTP (GET | POST | PUT | DELETE) + URI (*Uniform Resource Identifier*)

Cada um dos **métodos** acima representa um dos possíveis métodos ou operações suportados pelo protocolo HTTP, a serem executados sobre um **recurso** identificado por uma **URI**. Os métodos possuem a seguinte semântica, resumidamente:

- GET: Utilizado para recuperar informações (ex.: operações de consulta);
- POST: Utilizado para criar novos recursos (ex.: cadastrar novo usuário);
- PUT: Utilizado para atualizar recursos (ex.: atualizar dados cadastrais de um usuário já existente);
- DELETE: Utilizado para excluir recursos (ex.: excluir um usuário).

Exemplos de chamadas a operações da API que podem ser executadas por aplicativos:

- GET <http://contas.tcu.gov.br/nossaEscolaRS/escolas/123>: recupera as informações da escola identificada externamente pela URL <http://contas.tcu.gov.br/nossaEscolaRS/escolas/123>. Considera-se neste exemplo que “123” é o identificador da escola interno à aplicação (ex.: chave em um banco de dados);
- DELETE <http://contas.tcu.gov.br/appCivicoRS/pessoas/456>: excluir o usuário identificado externamente pela URI <http://contas.tcu.gov.br/appCivicoRS/pessoas/456>. Considera-se neste

exemplo que “456” é o identificador da escola interno à aplicação (ex.: chave em um banco de dados);

- POST <http://contas.tcu.gov.br/appCivicoRS/pessoas> {"nome": "José da Silva", "email": "josedasilva@tcu.gov.br", "dataNascimento": "12/12/1950"}: cadastra um novo usuário, enviando suas informações no corpo da mensagem;
- PUT <http://contas.tcu.gov.br/appCivicoRS/pessoas/789> {"nome": "José da Silva", "email": "novoemail@tcu.gov.br", "dataNascimento": "12/12/1950"}: atualiza os dados de um usuário existente, enviando suas informações no corpo da mensagem.

A API é documentada no formato Swagger⁷. Tal formato permite não apenas a visualização do contrato da nuvem cívica (operações suportadas, formatos das URIs, formato dos dados trafegados, parâmetros para cada operação) como também o teste das funcionalidades. Além disso, a documentação é gerada dinamicamente, sendo atualizada a cada mudança no contrato da API.

As figuras 4 a 6 ilustram os contratos da API do metamodelo para aplicativos cívicos e sua visualização por meio da interface Swagger. As outras APIs – saúde, educação, assistência social e empregos – possuem interfaces de documentação semelhantes.

5. CONCLUSÃO E PERSPECTIVAS FUTURAS

Os aplicativos Mapa da Saúde, Nossa Escola, Meu Remédio, Mami e Vacin App constituem exemplos reais de aplicações desenvolvidas tendo por base a API disponibilizada. Estão disponíveis nas lojas *App Store* e *Google Play*.

Em 2016, o TCU promoveu mais uma oportunidade de validar o modelo para aplicativos cívicos aqui descrito: o **Desafio de Aplicativos Cívicos**⁸. Trata-se de uma *hackathon* para que desenvolvedores externos inscrevam aplicativos que façam uso dos *web services* disponibilizados.

Na época em que este artigo foi escrito, o concurso estava em andamento, com cerca de noventa aplicativos inscritos. Ao longo do Desafio, mantivemos um canal aberto diretamente com os desenvolvedores, através de caixa postal institucional e de um site no GitHub⁹. Por

meio desse canal, recebemos mais de quarenta sugestões de ajustes e extensões às operações disponibilizadas pela API. A maioria das sugestões foram implementadas, e algumas encontram-se na lista de melhorias a serem implementadas na próxima versão da API.

Embora tenha sido proposta e implementada pelo TCU, a plataforma de *web services* disponibilizada está aberta a contribuições e parcerias. Seu sucesso dependerá da convergência de esforços e ideias na construção de soluções inovadoras e úteis ao cidadão. Se bem-sucedida,

Figura 4:
Exemplo de listagem de *web services* (metamodelo para aplicativos cívicos)

AppCívicoRS		
Web Services referentes ao modelo comum para aplicativos cívicos.		
Contact the developer		
TCU		
aplicativos : Aplicativos	Show/Hide	List Operations Expand Operations
grupos : Grupos	Show/Hide	List Operations Expand Operations
hashtags : Hashtags	Show/Hide	List Operations Expand Operations
instalacoes : Instalações	Show/Hide	List Operations Expand Operations
notificacoes : Notificações	Show/Hide	List Operations Expand Operations
peessoas : Pessoas	Show/Hide	List Operations Expand Operations
postagens : Postagens	Show/Hide	List Operations Expand Operations
tipos de objeto : Tipos de Objeto	Show/Hide	List Operations Expand Operations
tipos de postagem : Tipos de Postagem	Show/Hide	List Operations Expand Operations

Fonte: Elaboração própria

Figura 5:
Exemplo de listagem de operações de um *web service*

aplicativos : Aplicativos		Show/Hide	List Operations	Expand Operations
GET	/rest/aplicativos			Retorna o conjunto de aplicativos cadastrados.
POST	/rest/aplicativos			Cadastra um novo aplicativo.
GET	/rest/aplicativos/pessoa/{codPessoa}			Retorna o conjunto de aplicativos de um responsável.
GET	/rest/aplicativos/{codAplicativo}			Recupera os dados de um determinado aplicativo pelo código.
PUT	/rest/aplicativos/{codAplicativo}			Atualiza os dados de um aplicativo já cadastrado.
GET	/rest/aplicativos/{codAplicativo}/hashtags			Recupera as hashtags cadastradas de um determinado aplicativo.
GET	/rest/aplicativos/{codAplicativo}/tipos-perfil			Recupera os tipos de perfil criados para um determinado aplicativo.
POST	/rest/aplicativos/{codAplicativo}/tipos-perfil			Cadastra um novo tipo de perfil para um dado aplicativo.
GET	/rest/aplicativos/{codAplicativo}/tipos-perfil/{codTipoPerfil}			Recupera os dados de um determinado tipo de perfil.
PUT	/rest/aplicativos/{codAplicativo}/tipos-perfil/{codTipoPerfil}			Atualiza os dados de um tipo de perfil já cadastrado.

Fonte: Elaboração própria

Figura 6:

Exemplo de detalhamento de contrato de uma operação

aplicativos : Aplicativos Show/Hide | List Operations | Expand Operations

GET /rest/aplicativos Retorna o conjunto de aplicativos cadastrados.

Response Class (Status 200)

Model | Model Schema

```

Aplicativo {
  cod (long, optional),
  descricao (string, optional),
  links (Array(Link), optional),
  nome (string, optional)
}
Link {
  href (string, optional),
  rel (string, optional),
  templated (boolean, optional)
}

```

Response Content Type

Fonte: Elaboração própria

trará resultados que poderão contribuir para a melhoria dos serviços e políticas públicas, provendo novos serviços e recursos para a sociedade como um todo.

NOTAS

- 1 Disponível em: <<http://dados.gov.br>>. Acesso em: 10 out. 2016.
- 2 Disponível em: <<http://dados.gov.br/dataset/cnes>>. Acesso em: 10 out. 2016.
- 3 Disponível em: <<http://www.comprasgovernamentais.gov.br/acesso-aos-sistemas/comprasnet-siasg>>. Acesso em: 10 out. 2016.
- 4 Disponível em: <<http://www.tesouro.fazenda.gov.br/siafi>>. Acesso em: 10 out. 2016.
- 5 Disponível em: <<http://cne.smpe.gov.br/>>. Acesso em: 10 out. 2016.
- 6 Disponível em: <<http://www.bepiducb.com.br/index.html>>. Acesso em: 10 out. 2016.
- 7 Disponível em: <<http://swagger.io/>>. Acesso em: 10 out. 2016.
- 8 Disponível em: <<http://portal.tcu.gov.br/desafio-aplicativos-civicos/>>. Acesso em: 10 out. 2016.
- 9 Disponível em: <<https://github.com/>>. Acesso em: 10 out. 2016.

REFERÊNCIAS

- BOOTH, D. et al. (Eds.). W3C. Web Services Architecture. February 2004. Seção 1.4. Disponível em: <<https://www.w3.org/TR/ws-arch/#whatis>>. Acesso em: 10 out. 2016.
- BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Dispõe sobre os procedimentos a serem observados pela União, Estados, Distrito Federal e Municípios, com o fim de garantir o acesso a informações, e dá outras providências. Diário Oficial da União, Brasília, DF, 18 nov. 2011. Edição extra.
- ECMA. The JSON Data Interchange Format. Geneva, out. 2013. Disponível em: <<http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf>>. Acesso em: 10 out. 2016.
- ERL, T. et al. SOA with REST: Principles, Patterns & Constraints for Building Enterprise Solutions with REST. New Jersey: Prentice Hall, 2012. p. 24.
- FIELDING, R. T. Architectural styles and the design of network-based software architectures. 2000. Dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Information and Computer Science. University of California, Irvine, 2000.
- FIELDING, R. et al. Hypertext Transfer Protocol – HTTP/1.1. Junho 1999. Disponível em: <<https://www.w3.org/Protocols/rfc2616/rfc2616.txt>>. Acesso em: 10 out. 2016.

A recuperação semântica da informação no contexto do controle externo



Márcia Martins de Araújo Altounian

é servidora do Tribunal de Contas da União, bacharel em Biblioteconomia e Documentação pela UnB, com especialização em Gestão do Conhecimento pelo Instituto Blaise Pascal e Arquitetura e Organização da Informação pela UFMG.



Beatriz Pinheiro de Melo Gomes

é servidora do Tribunal de Contas da União, bacharel em Biblioteconomia e Documentação pela UnB, com especialização em Gestão Estratégica do Conhecimento e Inteligência Empresarial pela PUC/Paraná.

RESUMO

Nos sistemas de recuperação da informação (SRI), as técnicas de base sintática têm sido suplantadas pela crescente exploração das técnicas de recuperação semântica, que possibilitam a compreensão dos conceitos em seu contexto e finalidade. Algumas tecnologias têm contribuído para essa realidade, como a marcação semântica dos dados, utilizada na web semântica, o processamento de linguagem natural e as redes neurais. O tesauro também apresenta-se como um componente semântico que impacta no desempenho dos SRI. Tesouros são ferramentas da linguagem artificial em um domínio específico, formados por um sistema de conceitos relacionados entre si. Este artigo apresenta a aplicação do tesauro do TCU, denominado Vocabulário de Controle Externo, em alguns sistemas de informação corporativos.

Palavras-chave: Semântica. Recuperação da informação. Tesauro.

1. INTRODUÇÃO

A recuperação da informação (RI) é um campo de interesse comum da ciência da computação e da ciência da informação que se preocupa em desenvolver e estudar os aspectos relativos a eficiência e eficácia das buscas em um sistema de informação, de modo que os resultados sejam



coerentes com sua expressão de busca e, sobretudo, relevantes para o usuário do sistema.

Efetuar buscas em grandes repositórios de informações não estruturadas e não padronizadas torna-se uma tarefa árdua, levando muitas vezes a ambiguidades, a conteúdos fora de contexto e até mesmo a não recuperação da informação desejada. As ferramentas de busca, ao fazerem uso da linguagem natural, necessitam de conhecimento sobre o significado das expressões utilizadas e das relações entre elas, o que possibilita contextualização e tratamento de fenômenos linguísticos que afetam a qualidade da recuperação.

Para o processo de classificação e recuperação de informação, vários métodos automáticos vêm sendo desenvolvidos com o objetivo de obter respostas relevantes para qualquer pesquisa realizada. Até o momento, a maioria deles baseia-se em aspectos sintáticos e estatísticos, levando em consideração frequência e distribuição de palavras presentes em documentos. Por se basear nos aspectos sintáticos da informação, a classificação realizada por esses métodos ainda está muito distante, em termos de qualidade, da realizada com base na indexação dos assuntos dos documentos feita por especialistas.

Recentemente, vêm sendo explorados novos métodos de classificação automática de documentos baseados em aspectos semânticos da linguagem que se referem ao significado dos termos em seu contexto e finalidade. Uma dessas iniciativas consiste no uso de instrumentos de representação de relacionamentos semânticos e conceituais como os tesouros, que se caracterizam por representar domínio

do conhecimento e objetivam solucionar problemas de ambiguidade inerentes às palavras usadas cotidianamente.

O Vocabulário de Controle Externo, tesouro do TCU, foi desenvolvido com o objetivo de padronizar o tratamento de conteúdos especializados e contribuir com a recuperação semântica no contexto do controle externo, além de conferir maior agilidade e precisão às pesquisas nos sistemas de informação do Tribunal.

2. SEMÂNTICA

O termo “semântica”, tradicionalmente estudado na linguística e na filosofia, vem sendo amplamente usado nas mais variadas áreas de atuação, especialmente nas relacionadas à tecnologia da informação. A chamada web semântica (WS), que tem origem na expansão da web e nas limitações dos instrumentos de busca baseados em sintaxe, é certamente um dos motivos para essa associação.

Mas o que é semântica? A palavra tem origem no termo grego *semantiké*, e muitas acepções podem ser encontradas de acordo com a perspectiva utilizada, pois há, entre outras vertentes, semântica textual, cognitiva, lexical, formal e argumentativa. Em geral, todas convergem num mesmo ponto: estudam significado, ou significação. A ampla variedade de possibilidades atesta que o estudo do significado pode ser feito de vários ângulos.

Três propriedades básicas destacam-se na semântica – são elas a sinonímia, a antonímia e a polissemia. A sinonímia é a divisão da semântica que estuda a relação entre expressões linguísticas que têm o mesmo sentido. Por

exemplo, “garota” e “menina” são palavras substantivas que possuem um mesmo significado, remetendo à figura de uma jovem; assim também acontece com verbos como “renunciar”, “recusar” e “rejeitar”, que transmitem a ideia de repulsa. Ressaltando-se a rara ocorrência de sinonímia perfeita, pode-se afirmar que sinonímia é a relação entre palavras e expressões que possuem sentido e significado comuns.

Se por um lado a sinonímia estuda as palavras com significados semelhantes na língua, a antonímia trata do estudo das palavras que indicam sentidos opostos. Na mesma linha do exemplo anterior, podemos citar os substantivos “garota” e “senhora”, bem como os verbos “renunciar” e “aceitar” – tratam-se todos de termos com significação oposta.

Também é possível que uma mesma palavra assumam significados diferentes – nesse caso, o contexto em que estiver inserida ditará o seu sentido. Um bom exemplo do cotidiano é a palavra “manga”, da qual rapidamente podemos pensar em duas acepções bastante diferentes: uma fruta e uma parte de peça do vestuário. Esse é um exemplo comum de polissemia, termo que, formado pelo prefixo “poli” (“muitos”) e pelo sufixo “semos” (“significados”), é a parte da semântica que estuda as significações que uma palavra assume em determinado contexto linguístico.

Há ainda outras duas propriedades semânticas, na mesma linha de estudo da significação, que merecem ser citadas: a homonímia e a paronímia. A homonímia estuda a relação de duas ou mais palavras que, apesar de terem significados diferentes, têm a mesma forma e o mesmo som – é o caso de termos como concerto/conserto e rio/rio (substantivo/verbo), entre tantos outros. A paronímia estuda

as particularidades de palavras que, ao contrário, têm grafia e pronúncia semelhantes, mas significados diferentes – são exemplos: eminente/iminente e absolver/absorver.

Finalmente, a semântica estuda ainda as propriedades de denotação e conotação das palavras. Denotação é a propriedade que uma palavra tem de limitar-se ao seu próprio conceito – por exemplo, o termo “estrelas” em “as estrelas do céu”. Conotação é a propriedade que uma palavra tem de ampliar-se no seu campo semântico, dentro de um contexto, gerando várias possibilidades interpretativas – por exemplo, o mesmo termo “estrelas” em “as estrelas do cinema”.

Quando transportamos os conceitos aqui explorados para o que hoje conhecemos como WS, fica fácil perceber que estamos falando de promover melhorias nos processos de representação e recuperação da informação na web. Desde 1990, a web é caracterizada por usar linguagens de marcação que objetivam a apresentação e a leitura por pessoas e por mecanismos de busca baseados em algoritmos com orientação à sintaxe. O uso da semântica pode ampliar a possibilidade de associações dos documentos a seus significados por meio de metadados descritivos. Portanto, a questão do significado na semântica é fundamental à WS.

Das diversas linhas de estudo relacionadas à semântica, a da semântica formal parece ser a mais pertinente à tecnologia da informação. Três aspectos principais permeiam os estudos sobre semântica formal:

1. O princípio da composicionalidade, que estabelece que o significado das sentenças depende do



significado das palavras que as compõem – ou seja, o significado do todo é função do significado das partes e da combinação sintática entre elas. Para saber o significado de uma sentença, é necessário conhecer o significado das suas partes, bem como as regras que definem sua combinação.

2. A condição de verdade, que determina as condições em que tal sentença é verdadeira. Nesse contexto, saber o significado de uma sentença equivale a conhecer suas condições verdade, o que não é o mesmo que saber o seu valor verdade, ou seja, se o fato é verdadeiro ou falso.
3. Os modelos em semântica, em que são construídos sistemas simples, em relação aos sistemas complexos que se desejam estudar. Constrói-se uma teoria lógica para o modelo e, se os resultados forem razoavelmente positivos em relação ao sistema complexo, diz-se que o sistema simples é um bom modelo; caso contrário, ele é abandonado.

A semântica formal considera ainda o fato de que as línguas naturais são utilizadas para falar sobre objetos, indivíduos, fatos, eventos e propriedades, descritos como externos à própria língua – assim, a referencialidade é um de seus aspectos fundamentais. Por essa razão, na semântica formal o significado é entendido, por um lado, como uma relação com a linguagem, e, por outro, com aquilo sobre o que a linguagem fala.

A semântica formal procura responder às seguintes perguntas: o que representam ou denotam as expressões linguísticas?, como calculamos o significado de expressões complexas a partir dos significados de suas partes?

3. RECUPERAÇÃO SEMÂNTICA DA INFORMAÇÃO

Os sistemas de recuperação de informação (SRI) procuram representar o conteúdo dos documentos e apresentá-los ao usuário de maneira a atender rapidamente sua necessidade de informação.

Para tanto, a RI pesquisa técnicas para tratamento, organização e busca de conteúdos a partir do uso de padrões. As ferramentas de RI, geralmente, trabalham com técnicas de indexação capazes de indicar e acessar rapidamente documentos de um banco de dados textual.

Existem três tipos principais de indexação:



- indexação tradicional, em que se determinam os termos descritivos ou caracterizadores dos documentos;
- indexação *full-text* (ou indexação do texto todo), em que todos os termos que compõem o documento fazem parte do índice;
- indexação por *tags* (por partes do texto), em que apenas algumas partes do texto são escolhidas para gerar as entradas no índice (somente as consideradas mais importantes ou mais caracterizadoras).

As buscas são geralmente feitas através de termos fornecidos pelo usuário ou escolhidos por ele entre alguns apresentados. Esses termos podem significar o assunto ou classe a que pertencem os documentos desejados (na indexação tradicional) ou os termos que devem estar presentes nos documentos desejados (nas indexações *full-text* e por *tags*).

Nos sistemas convencionais de busca, as técnicas utilizadas são de base sintática. Entretanto, quando a busca do usuário envolve informação cuja relevância não pode ser dada por palavras-chave, esse modelo não satisfaz. Observa-se, portanto, a crescente exploração das informações semânticas, o que possibilita compreensão dos conceitos em seu contexto e finalidade.

A marcação semântica dos dados na origem é um exemplo das novas tecnologias utilizadas para a RI. A WS tem-se utilizado dessa estratégia para implementar padrões de metadados que adicionem aos dados informações significativas sobre seus contextos, marcando-os semanticamente.

A exploração da semântica intrínseca dos dados busca fundamentos da linguística e da ciência da informação para ampliar o universo de informações recuperadas e a aferição de contextos, por meio do uso de estruturas da linguagem natural, como os sintagmas verbais e nominais, e de ferramentas de representação de relacionamentos semânticos e conceituais.

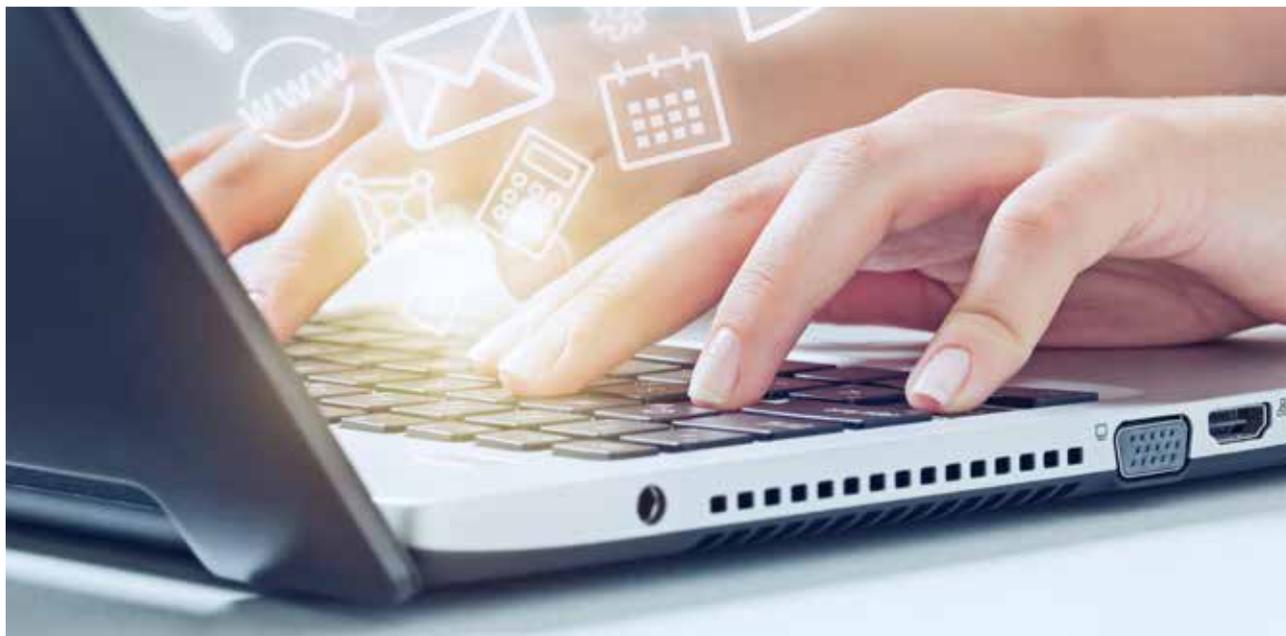
Técnicas de processamento de linguagem natural (PLN) que permitam a construção de algoritmos para a busca de informação relevante em grande quantidade de documentos estão em crescente demanda. Métodos eficientes de PLN devem fundamentar-se em conhecimentos básicos sobre propriedades da linguagem e principalmente sobre a semântica dos conceitos. Desse contexto emerge a ideia de memória semântica, tema que tem sido objeto de teorias psicolinguísticas e é uma fonte rica para desenvolvimento de modelos computacionais que pretendem se aproximar dos processos mentais usados pelo cérebro humano na compreensão da linguagem. Além dos algoritmos, os modelos computacionais necessitam de dados que representam o conhecimento sobre a linguagem e as associações de senso comum entre os conceitos lexicais e suas propriedades. Atualmente, isso é muito difícil, pois essa tarefa somente pode ser produzida e verificada por humanos. A memória semântica trabalha com um léxico mental – ou seja, conceitos e unidades de conhecimento – e contém informações sobre as relações entre conceitos, formando uma rede conceitual de elementos conectados uns com outros por diferentes tipos de associações.

As redes neurais, por sua vez, são uma representação que têm muitas características comuns à memória humana: podem lidar com informação incompleta ou distorcida, permitem generalizações automáticas e exibem conteúdo baseado no contexto. Essas funções possibilitam várias aplicações na RI guardada na memória humana, essenciais em contextos precisos, e são essas aplicações que se pretendem reproduzir no ambiente computacional.

Um componente semântico que impacta o desempenho de um SRI é o tesouro, ferramenta da linguagem artificial de um domínio conhecido, construído por especialistas para representar através de conceitos o conteúdo informacional, especificando as relações entre eles. Trata-se de um sistema de conceitos que se relacionam entre si e são representados por termos.

Cada termo tem obrigatoriamente vinculação com outro, e é essa vinculação que forma a estrutura do tesouro. Os termos são utilizados pelos indexadores no momento da indexação e devem ser disponibilizados ao usuário no momento da RI.

Silveira e Ribeiro-Neto (2004) estudam o uso automático de tesouro para melhorar resultados de buscas na web, por meio de ranking baseado em conceitos que tem sido estudado para a RI em domínios específicos. Em experimento realizado pelos autores, os termos de busca utilizados em um SRI na web foram comparados aos conceitos de um tesouro, utilizado para encontrar conceitos relacionados. Cada conceito relacionado foi interpretado de forma independente e processado separadamente, e então combinado em uma rede bayesiana¹, para permitir um





ranking final, baseado em conceitos. O objetivo era verificar se a informação acrescida de conceito aumentaria a média da precisão dos resultados da busca.

No estudo foram utilizadas seis fontes de informação: palavra chave, conceito, termo específico, termo geral, termo relacionado e sinônimo. Os autores verificaram, entre outras coisas, que a utilização de um tesouro específico para um domínio específico é fundamental para a melhoria do desempenho das buscas.

O experimento demonstrou o aumento de 30% na média da precisão dos resultados das buscas e, portanto, propõe a visão de que os tesouros melhoram tanto a revocação como a precisão em um SRI.

4. TESAURO

De acordo com Moreira, Alvarenga e Oliveira (2004), o termo “tesouros” se origina do grego thesaurós e significa tesouro ou repositório. Esse termo surgiu com a publicação do dicionário analógico de Peter Mark Roget, em Londres, em 1852, intitulado *Thesaurus of English words and phrases*. O termo também designa vocabulário, dicionário ou léxico, porém o dicionário de Roget se diferenciava dos outros por ser um vocabulário organizado de acordo com seu significado, e não por ordem alfabética. A obra teve o mérito de estabelecer a denominação para vocabulários que relacionam seus termos por meio de algum tipo de relação de significado.

Em sua introdução, Roget define seu dicionário como uma “classificação de ideias” e explica que, diferen-

temente dos outros, o seu permite que se chegue à palavra mais adequada ou à que melhor se ajuste às necessidades do escritor, sem que, de início, ele saiba qual é ela (GOMES, 1996).

Nos anos 1960, o cientista da informação Brian Campbell Vickery apresentou quatro significados para o termo “tesouro” na literatura de sua área, sendo que o significado mais comum é o de uma lista alfabética de palavras, em que cada palavra é seguida de outras relacionadas a ela (VICKERY, 1980 apud MOREIRA; ALVARENGA; OLIVEIRA, 2004).

Currás (1995, p. 88) define tesouro como “uma linguagem especializada, normalizada, pós-coordenada, usada com fins documentários, onde os elementos linguísticos que a compõem – termos, simples ou compostos – encontram-se relacionados entre si sintática e semanticamente”.

Tristão (2004, p. 167) o define como “vocabulário de termos, que nada mais é do que uma seleção de termos, baseados em análise de conceitos, na qual se define o termo geral, de maior abrangência, e sua relação com termos mais específicos, que representam os conceitos menores”. A Organização Nacional de Padrões de Informação especifica:

Um tesouro é um vocabulário controlado organizado em uma ordem preestabelecida e estruturado de modo que os relacionamentos de equivalência, de homografia, de hierarquia, e de associação entre termos sejam indicados claramente e identificados por indicadores de relacionamento padronizados empregados reciprocamente. As finalidades primordiais de um te-

sauro são (a) facilitar a recuperação dos documentos e (b) alcançar a consistência na indexação dos documentos escritos ou registrados de outra forma e outros tipos, principalmente para sistemas de armazenamento e de recuperação de informação pós-coordenados. (ANSI/NISO Z39.19, 2003 apud SALES; CAFÉ, 2008).

Conforme Campos e Gomes (2006), a evolução na construção de tesouros baseia-se em duas vertentes, a americana e a europeia. Os tesouros elaborados nos Estados Unidos a partir da década de 1950 foram fruto do desenvolvimento ocorrido a partir do cabeçalho de assuntos para o unitermo, passando de um sistema pré coordenado para sistemas pós-coordenados.

Silva (2008) afirma que na mesma época, na Inglaterra, o Classification Research Group (CRG) – a partir da Teoria da Classificação Facetada, desenvolvida pelo matemático e bibliotecário indiano Shiyali Ramamrita Ranganathan – ampliou as categorias de personalidade, matéria, energia, espaço e tempo (PMEST) e desenvolveu diversas tabelas de classificação, dando origem a uma técnica denominada *Thesaurofacet*, que permitiu melhor posicionamento do conceito no sistema de conceitos em uma dada área de assunto, por meio do uso de suas categorias. Conjuntamente, também embasaram os tesouros terminológicos a Teoria da Classificação Facetada, a Teoria do Conceito e alguns princípios terminológicos. Esses instrumentos têm nas características do conceito um elemento essencial para evidenciar as relações entre os conceitos e seu posicionamento no sistema, além de defini-lo.

A citada Teoria do Conceito, voltada para o referente e originalmente chamada Teoria Analítica do Conceito, foi

lançada no final da década de 1970 pela cientista da informação Ingetraut Dahlberg, agregando princípios terminológicos relacionados ao conteúdo conceitual e à sua definição. Para Campos e Gomes (2006), essa é uma teoria consolidada para a determinação do que se entenderia por menor unidade em um tesouro: o conceito representado por um termo. Além disso, Moreira (2003) aponta como inovação o uso de definições para o posicionamento do conceito no sistema.

Bräscher (2010) aponta como função dos tesouros a tradução da linguagem dos documentos, dos indexadores e dos pesquisadores a uma linguagem controlada, usada na indexação e na RI em sistemas de informação. Conforme Sales e Café (2008), o ANSI/NISO Z39.19, de 2003, ressalta que os tesouros não são utilizados somente pelos especialistas da informação no momento da indexação, mas também por usuários da informação no momento da busca de documentos.

Segundo Café e Bräscher (2011), nos tesouros as “relações semânticas são estabelecidas por meio da análise das características ou propriedades dos conceitos, as quais permitem identificar diferenças e semelhanças que evidenciam determinados tipos de relacionamentos”. Um termo presente em um tesouro pode ser caracterizado de maneiras diferentes dependendo do assunto em questão e também do tipo de sistema que se deseja construir. A estrutura dos tesouros compreende três tipos principais de relações semânticas para relacionar os termos: a hierarquia, a equivalência e a associação.

Nas relações hierárquicas os termos são organizados em gênero/espécie. As relações de equivalência são de sinonímia, ou seja, há termos sinônimos presentes no tesouro e deve-se indicar qual o termo adequado para representar





determinado conceito. As relações de associação, por sua vez, apresentam associações entre os termos, sem especificar qual tipo de relação propriamente existe – são apenas termos que se relacionam de alguma forma.

Nos tesouros são tratados ainda casos de ambiguidade (possibilidade de que uma comunicação linguística se preste a mais de uma interpretação) e de polissemia (possibilidade de que uma palavra comporte mais de um significado).

5. VOCABULÁRIO DE CONTROLE EXTERNO

O tesouro do TCU, denominado Vocabulário de Controle Externo (VCE), foi lançado em 2015 e objetiva ser instrumento de controle terminológico que possibilite padronização no tratamento técnico e maior precisão na recuperação dos conteúdos presentes nos sistemas de informação do TCU.

O inter-relacionamento dos conceitos no VCE foi expresso por meio de relações de três tipos: de equivalência, hierárquicas e associativas. O objetivo das relações é apresentar os descritores em seu contexto semântico.

- **Relação de equivalência:** se termos sinônimos ou quase sinônimos são considerados, para efeito do vocabulário, como representando um mesmo conceito, um deles é escolhido como descritor e os demais são proibidos.

- **Relação hierárquica:** relacionamento que exprime os graus ou os níveis de superordenação e de subordinação dos termos; o termo superordenado representa o gênero de que o termo subordinado é tipo ou espécie.
- **Relação associativa:** reunião de conceitos afins que merecem estar relacionados, mas que não estão ligados por relacionamentos de equivalência nem de hierarquia.

No VCE, cada termo corresponde a um conceito, e todos os termos possuem relacionamentos, sendo o relacionamento determinado pelo significado do termo. As relações entre termos ajudam a compreender os conceitos específicos de controle externo e de áreas correlatas que compõem o tesouro.

Bem além de uma lista hierarquizada de palavras, o VCE é composto de três partes distintas, mas inter-relacionadas. A primeira é formada por palavras chave, relativas às áreas de atuação do Tribunal, acompanhadas por definições e sinônimos; a segunda, correspondente à clientela do TCU e às Entidades de Fiscalização Superiores (EFS) associadas à Organização Internacional de Entidades Fiscalizadoras Superiores (Intosai), traz informações como histórico, nomes alternativos, CNPJ e instituições afins. A terceira parte contempla a toponímia nacional formada pelas regiões, mesorregiões, unidades federativas e municípios brasileiros.

6. APLICAÇÕES DO VCE

6.1 E-JURIS

O e-Juris é uma ferramenta corporativa que faz parte do e-TCU, o qual agrega todos os sistemas de instrução e controle de processos do Tribunal, contendo a mesma lógica, estrutura e apresentação dos demais sistemas de processos. Além disso, é integrado com o portal do TCU, com o sistema corporativo de busca e com outros sistemas do TCU, como o Sagas e o VCE. As premissas adotadas para o novo sistema foram seletividade, qualidade, relevância, tempestividade e simplicidade.

A principal finalidade do e-Juris é a divulgação das teses relevantes, do ponto de vista jurisprudencial, que fundamentaram acórdãos do TCU, por meio de publicações periódicas (boletins e informativos) e de formação e disponibilização, para pesquisas e consultas, da base de dados da jurisprudência do Tribunal. Com isso, o novo sistema unifica processos de trabalho que antes eram feitos separadamente e sem integração.

As teses jurisprudenciais relevantes são representadas por enunciados, sob formato de ementa. Os enunciados representam precedentes jurisprudenciais, não o “entendimento” ou a jurisprudência prevalecente do Tribunal sobre determinada questão.

A adoção da indexação pelo e-Juris permite maior precisão na recuperação dos enunciados. Além da previsão de busca por termos sinônimos, o VCE agrega ao sistema a funcionalidade de sugestão de assuntos correlatos a ser

pesquisados, uma vez que a ferramenta é estruturada em um sistema de conceitos inter-relacionados por hierarquia, equivalência e associação.

6.2 BIBLIOTECA DIGITAL DO PORTAL DO TCU

Repositórios corporativos de conhecimento prestam-se a disseminar a informação produzida internamente, a permitir o acesso à cultura organizacional e a subsidiar a informação transformada em conhecimento. Sendo mais que um depósito de documentos, os repositórios podem atuar de forma decisiva, dando suporte ao desenvolvimento de novos produtos e serviços e à formação e à capacitação da força de trabalho da organização. Além disso, costumam servir de fonte oficial de informação aos parceiros e colaboradores, amparar as atividades cotidianas e auxiliar o processo decisório.

No TCU, a biblioteca digital é um dos repositórios corporativos de conhecimento de mais fácil acesso e utilização. Desenvolvida para organizar, tratar e disseminar informações que possam gerar novos conhecimentos, a biblioteca digital permite a inserção de material com diversas tipologias documentais. Assim, no mesmo ambiente é possível encontrar livros, trabalhos acadêmicos, apresentações, cartilhas, periódicos, acordos, contratos, expedientes e normativos, além de imagens diversas e outros tipos de recursos. A ferramenta permite que sejam estabelecidos dois níveis de acesso aos conteúdos depositados: é possível permitir o livre acesso a documentos de caráter público, da mesma forma que é possível limitar o



acesso a documentos de cunho interno ou que possuam algum tipo de restrição.

A inserção de documentos no ambiente é realizada descentralizadamente, e existem gestores de conteúdo responsáveis pela aprovação do que ficará disponível no repositório. A biblioteca digital possui também um formulário de entrada de dados que pode ser usado em todo o Tribunal. Esse formulário é composto de metadados controlados e possibilita a descrição do conteúdo, com elementos tais como título, autoria e data. Mais do que isso, requer que os documentos sejam classificados em uma árvore temática e que os assuntos tratados sejam traduzidos por palavras-chave oriundas de um vocabulário controlado.

Em outras palavras, a biblioteca digital do TCU constituiu-se num repositório corporativo de conhecimento que exige que as informações sejam classificadas e indexadas. Para a indexação dos conteúdos, o ambiente está integrado ao VCE. Além disso, como a biblioteca também está integrada à ferramenta de pesquisa textual do portal corporativo, é possível fazer buscas diretamente no portal e recuperar o conteúdo depositado no ambiente dessa biblioteca.

6.3 SISTEMA ORIENTAR

O Sistema Orientar foi concebido para ser ferramenta de orientação, gestão e disseminação do conhecimento sobre controle externo. Ele permite que qualquer servidor do Tribunal possa encaminhar perguntas sobre temas pré-selecionados de controle externo, tais como auditoria, planejamento, contas anuais, tomada de contas especial, representação, denúncia, solicitação, avaliação de qualidade, cobrança executiva, normas de controle externo e demais procedimentos processuais, além de tirar dúvidas sobre o sistema Fiscalis.

Após a seleção do tema, o sistema direciona a pergunta para a unidade responsável pela área. A partir da coletânea das perguntas e respostas, cada unidade respondente cria automaticamente um banco de dados de perguntas mais frequentes (FAQ), que fica armazenado no sistema e disponível para consultas e pesquisas por todos os servidores.

Diversas unidades do TCU já foram cadastradas como unidades respondentes, e o sistema é integrado ao VCE. A adoção do vocabulário controlado permite maior precisão na recuperação da informação desejada, uma vez que são atribuídos, tanto às perguntas quanto às respostas, termos de indexação que representam os assuntos tratados.

6.4 WIKI DE CONTROLE EXTERNO

Entre as diversas possibilidades de armazenar o conhecimento de uma instituição, uma em especial merece

destaque numa era em que produzimos conhecimento por meio de variados colaboradores e parceiros: as wikis – ferramenta de tecnologia conhecida como “software social” e desenhada com um conjunto de características que permitem a criação e a organização de conhecimento no mundo colaborativo. A utilização das wikis tem se revelado uma solução de baixo custo, com alto grau de eficiência, para o fomento à criação cooperativa de conhecimento dentro das organizações.

O ambiente wiki é a evolução do conceito de *computer supported cooperative work* (CSCW), que surgiu da necessidade de que organizações tenham pessoas trabalhando em lugares físicos diferentes, ao mesmo tempo que precisando alcançar resultados rápidos conjuntamente; ou seja, surgiu para facilitar a comunicação e a produtividade de grupos remotos.

Desde 2009, o TCU utiliza o software livre Media-wiki para gerir uma wiki de controle externo e acesso restrito aos seus servidores. Trata-se de um importante espaço colaborativo de construção de conhecimento que reúne tutoriais informais e verbetes especializados oriundos do VCE.

A Wiki pode ser acessada e editada por todos os servidores; eles agregam informações aos verbetes e tutoriais em tópicos que abrangem norma, legislação e doutrina. A Wiki é importante ferramenta de gestão do conhecimento, na medida em que fornece informações e documentos úteis ao trabalho cotidiano dos auditores, de acordo com sua área de atuação, e se configura como um genuíno ambiente de inteligência coletiva da organização.

6.5 FUTURAS APLICAÇÕES

O VCE apresenta ainda potenciais aplicações em diversos sistemas de informação do Tribunal. A integração da terminologia na ferramenta de busca do portal do TCU, por meio da adoção do vocabulário controlado tanto para o tratamento quanto para a busca, é fator primordial para o aumento de precisão e rapidez na RI.

Outra possibilidade de uso do VCE é como dicionário em softwares de indexação automática de grandes volumes de documentos, servindo de parâmetro terminológico na área de controle externo. Essa funcionalidade já foi testada durante o processo de indexação automática de enunciados, constante em uma prova de conceito para aquisição de software de *data mining* e análise semântica de dados.

Em um cenário ideal, vislumbra-se que o Tribunal adote padrões, a exemplo de metadados de assunto e da ferramenta de controle terminológico, para garantir a melhoria do desempenho de seus SRI.

NOTAS

- 1 Redes bayesianas constituem um modelo gráfico que representa simplesmente as relações de causalidade das variáveis de um sistema. Em resumo, redes bayesianas, também conhecidas como redes de opinião, redes causais e gráficos de dependência probabilística, são modelos gráficos para raciocínio (conclusões) baseado na incerteza, em que os nós representam as variáveis (discretas ou contínuas), e os arcos representam a conexão direta entre eles (SILVEIRA; RIBEIRO NETO, 2004).

REFERÊNCIAS

- ALMEIDA, M. B.; SOUZA, R. R. Avaliação do espectro semântico de instrumentos para organização da informação. *Encontros Bibli – Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 16, n. 31, p. 25-50, 2011. Disponível em: <<http://mba.eci.ufmg.br/downloads/11963-60907-1-PB.pdf>>. Acesso em: 23 nov. 2016.
- ALTOUNIAN, M.; ZAULI, A. A semântica na recuperação da informação na web: novas tendências. 2013. Trabalho apresentado como requisito parcial para aprovação na disciplina Recuperação da Informação, Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2013.
- BRÄSCHER, M. *Elaboração de tesouros*. Brasília, 2010.
- CAFÉ, L.; BRÄSCHER, M. Organização do conhecimento: teorias semânticas como base para estudo e representação de conceitos. *Informação & Informação*, Londrina, v. 16, n. 3, p. 25-51, jan./jun. 2011.
- CAMPOS, M. L. A.; GOMES, H. E. Metodologia de elaboração de tesouro conceitual: a categorização como princípio norteador. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 11, n. 3, p. 348-359, set./dez. 2006.
- CAPRI, D.; GARRIDO, I.; DUARTE, R. Recuperação semântica da informação. 2009. Trabalho apresentado como requisito parcial para aprovação na disciplina Recuperação da Informação, Centro de Ciências da Educação, Universidade Federal de Santa Catarina, Florianópolis, 2009. Disponível em: <<http://pt.slideshare.net/doritchka/angel-recuperao-semntica-da-informao>>. Acesso em: 23 nov. 2016.
- CURRÁS, E. *Tesouros: linguagens terminológicas*. Brasília, DF: CNPq; Ibict, 1995.
- GOMES, H.E. *Classificação, tesouro e terminologia; fundamentos comuns*. Rio de Janeiro: UNIRIO, 1996.
- MOREIRA, A. Tesouros e ontologias: estudo de definições presentes na literatura das áreas das ciências da computação e da informação, utilizando-se o método analítico-sintético. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 8, n. 2, p. 216-226, jul./dez. 2003.
- _____; ALVARENGA, L.; OLIVEIRA, A. P. Thesaurus and ontology: a study of the definitions found in the Computer and Information Science Literature, by means of an analytical-synthetic method. *Knowledge Organization*, v. 31, n. 4, p. 231-244, 2004.
- SALES, R., CAFÉ, Lúcia. Semelhanças e Diferenças entre Tesouros e Ontologias. *DataGramaZero - Revista de Ciência da Informação*, v.9 n.4 Ago 2008.
- SILVA, A. *Análise das relações semânticas em tesouros jurídicos brasileiros: orientações das normas e aplicação prática*. 2013. Trabalho de Conclusão de Curso (Bacharel em Biblioteconomia) – Centro de Ciências da Educação, Universidade Federal de Santa Catarina, Florianópolis, 2013. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/103801/TCC_Aline_da_Silva_20131PDFA.pdf?sequence=1>. Acesso em: 23 nov. 2016.
- SILVA, D., SOUZA, R., ALMEIDA, M.B. Ontologias e vocabulários controlados: comparação de metodologias para construção. *Ciência da Informação*, v. 37, n. 3, p. 60-75, set./dez. 2008.
- SILVEIRA, M. L.; RIBEIRO-NETO, B. Concept-based ranking: a case study in the juridical domain. *Information Processing & Management*, Doha, v. 4, n. 5, p. 791-805, Sept. 2004.
- SOUZA, A. et al. Recuperação semântica de objetos de aprendizagem: uma abordagem baseada em tesouros de propósito genérico. In: *SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO*, 19., 2008, Uberlândia. Anais... Uberlândia: SBIE, 2008. p. 603-612.
- SOUZA, A.; ROCHA, R. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 11, n. 2, p. 161-173, maio/ago. 2006. Disponível em: <<http://www.scielo.br/pdf/pci/v11n2/v11n2a02.pdf>>. Acesso em: 23 nov. 2016.
- SZYMANSKI, J.; DUCH, W. Information retrieval with semantic memory model. *Cognitive Systems Research*, Kalamazoo, v. 14, n. 1, p. 84-100, Apr. 2012. Disponível em: <<http://ethologie.unige.ch/etho5.10/themes/semantic.memory/szymanski.duch.2011.information.retrieval.in.semantic.memory.neural.network.models.pdf>>. Acesso em: 23 nov. 2016.

Geotecnologias e o monitoramento dos Objetivos de Desenvolvimento Sustentável pelas Entidades de Fiscalização Superior



Rherman Radicchi Teixeira Vieira

é servidor do Tribunal de Contas da União, lotado na Secretaria de Fiscalização de Infraestrutura Portuária, Hídrica e Ferroviária, formado em Agronomia pela Universidade de Brasília.



Leonardo Pereira Garcia

é servidor do Tribunal de Contas da União, lotado no Centro de Pesquisa e Inovação/ISC. Graduado em Administração pela Universidade de Brasília, com especialização em Educação Corporativa.



André Delgado de Souza

é servidor do Tribunal de Contas da União lotado na Secex/PB, bacharel, mestre e doutor em Engenharia Civil pela Universidade Federal de Pernambuco, com doutorado sanduíche pela Universidade Técnica de Berlim.



Erick Fonseca dos Santos

é servidor do Tribunal de Contas da União, mestrando em computação aplicada, com foco em visão computacional e reconhecimento de objetos em imagens de satélite, associado ao projeto GeoControle, do TCU, para acompanhamento de obras de infraestrutura por meio de sensoriamento remoto. É formado em computação.



RESUMO

Dentre as iniciativas propostas pela Organização das Nações Unidas (ONU) para que os Objetivos de Desenvolvimento Sustentável (ODS) superem os desafios encontrados pelos Objetivos de Desenvolvimento do Milênio (ODM), está a utilização de dados geoespaciais e investimentos em capacitação para utilizar as novas tecnologias. Além disso, Normas Internacionais das Entidades Fiscalizadoras Superiores (ISSAI) reconhecem a importância da capacitação de seus servidores para atender às novas demandas. A proposta de ação concreta deste trabalho visou apoiar a estratégia de investimento em geotecnologias para o monitoramento dos ODS pelas Entidades de Fiscalização Superior (EFS). Essa atividade foi desempenhada por meio da revisão da literatura sobre as referências técnicas das Nações Unidas e da Intosai que possam contribuir com a utilização de geotecnologias para o monitoramento dos ODS pelas EFS.

Palavras-chave: Objetivos de Desenvolvimento Sustentável. Entidades de Fiscalização Superior. Geotecnologias. Sensoriamento remoto. Sistema de Informação Geográfica (SIG). *Global Positioning System* (GPS). Intosai. ISSAI. Padrões internacionais em auditoria. Competências técnicas. Capacitação. Diagnóstico.

1. INTRODUÇÃO

A Agenda para o Desenvolvimento Sustentável de 2030 da Organização das Nações Unidas (ONU) lança um desafio sobre as atividades desempenhadas pelas Entidades de Fiscalização Superior (EFS). Recentes resoluções da Assembleia Geral da ONU enfatizam o papel-chave das EFS e da International Organization of Supreme Audit Institutions (Intosai) no alcance dos Objetivos de Desenvolvimento Sustentável (ODS) (NAÇÕES UNIDAS, 2016).

A Resolução da ONU A66/209 de 2011 (Id., 2011) salienta que as EFS desempenham um papel importante na promoção da eficiência, eficácia, responsabilização e transparência da administração pública, que propiciam o desenvolvimento nacional em direção às ODS. Além disso, a Resolução A69/228 de 2014 (Id., 2014a) reforça explicitamente o papel-chave das EFS na Agenda 2030.

No relatório do 23º Simpósio das Nações Unidas/Intosai (Id., 2015d) sobre o papel das EFS na Agenda 2030, destaca-se a importância do uso intensivo de análise de dados (*data analytics*). Dentre as iniciativas propostas pela ONU para que os ODS superem os desafios encontrados pelos Objetivos de Desenvolvimento do Milênio (ODM), está a utilização de dados geoespaciais (Id., 2015c). Além disso, as Normas Internacionais das Entidades Fiscalizadoras Superiores (ISSAI), ressaltam a possibilidade de aplicação das geotecnologias às diver-

sas fases da auditoria (INTERNATIONAL ORGANISATION OF SUPREME AUDIT INSTITUTIONS, 2013b).

No entanto, inexistem nas fontes bibliográficas pesquisadas obras com o objetivo de sistematizar conjuntamente as referências técnicas da ONU em relação ao uso de geotecnologias para o monitoramento dos ODS e referências técnicas da Intosai que abordem o tema geotecnologias.

Pelos motivos expostos e visando auxiliar o uso de novas tecnologias aplicadas ao controle, este artigo se propõe a fazer a revisão da literatura de referências técnicas das Nações Unidas e da Intosai que possam contribuir com o uso de geotecnologias para o monitoramento dos ODS pelas EFS.

2. GEOTECNOLOGIAS COMO FERRAMENTAIS DE APOIO AO MONITORAMENTO DOS ODS

As geotecnologias favorecem a abordagem universal e padronizada para o monitoramento de outras informações relevantes, tais como indicadores econômicos, educacionais, ambientais e de saúde. A informação geográfica permite a análise e a modelagem de dados, a criação de mapas e a detecção e o monitoramento de suas alterações ao longo do tempo de forma consistente e padronizada (NAÇÕES UNIDAS, 2015a).

Souza (2016, p. 40) descreve as geotecnologias como “o conjunto de tecnologias para coleta, processa-

mento, análise e disponibilização de informações georreferenciadas”. Dentre elas têm-se o Global Navigation Satellite System (GNSS), o Sensoriamento Remoto e os Sistemas de Informações Geográficas (SIG). O GNSS permitiu às pessoas saberem com precisão a localização delas na superfície terrestre, destacando-se entre os sistemas dessa natureza o Global Positioning System (GPS) (SOUZA, 2016).

Souza (2016, p. 41) caracteriza o sensoriamento remoto “pela utilização de sensores para captação e registro a distância, sem contato direto, da energia refletida ou absorvida pela superfície do alvo”. A partir da utilização de programas computacionais é possível realizar o armazenamento, a manipulação e a análise de dados gerados por essa tecnologia. Os veículos aéreos não tripulados (VANT), pequenas aeronaves não tripuladas, podem realizar monitoramento por meio de imagens aéreas e filmagens (SOUZA, 2016).

Nesse contexto, os SIG são “sistemas que realizam o tratamento computacional de dados geográficos” (SOUZA, 2016, p. 41). Câmara (2015, p. 2) explica que “a principal diferença de um SIG para um sistema de informação convencional é sua capacidade de armazenar tanto os atributos descritivos como as geometrias dos diferentes tipos de dados geográficos”. Com os SIG é possível integrar dados de diversas fontes, permitindo a realização de análises mais complexas (SOUZA, 2016). Podem-se inserir, em uma única base, dados provenientes de diferentes fontes e com-



biná-los, com apoio de algoritmos de manipulação e análise (CÂMARA, 2005).

O documento “The millennium development goals: report 2015” (NAÇÕES UNIDAS, 2015e) sintetiza os avanços, os desafios e as lições aprendidas durante os quinze anos de monitoramento dos ODM. Ao tratar sobre os desafios para o monitoramento, afirma-se que grandes lacunas de dados, dados de baixa qualidade, dados desatualizados e não desagregados em importantes dimensões figuram entre os principais desafios. Dentre as iniciativas propostas para superar esses desafios, está a **utilização de dados geoespaciais** (NAÇÕES UNIDAS, 2015e).

No relatório “A world that counts: mobizing the data revolution for sustainable development” (INDEPENDENT EXPERT ADVISORY GROUP, 2015), a ONU convoca os países signatários da Agenda 2030, as empresas e a sociedade civil a coordenarem esforços com a finalidade de melhorar a disponibilidade, a qualidade, a atualização e a desagregação das informações, objetivando apoiar a implementação em todos os níveis.

De acordo com o documento “Assessing gaps in indicator availability and coverage” (CASSIDY, 2014), um grande desafio a ser superado refere-se à baixa cobertura geral dos indicadores propostos. Quase um terço dos indicadores carece de dados em mais da metade dos países (SHUANG et al., 2013) e, em média, apenas 46% dos dados foram coletados, o que apresenta um desafio internacional no que se refere à geração de dados estatísticos (NAÇÕES UNIDAS, 2014e).

O relatório “Indicators and a monitoring framework for the sustainable development goals” (NAÇÕES UNIDAS, 2015c) destacou as oportunidades trazidas pela revolução dos dados, por meio de big data, dados sociais e geofísicos e novas formas de compartilhamento de dados. Além disso, esse relatório trouxe uma série de exemplos de formas de suporte das geotecnologias aos ODS, conforme demonstra o Anexo 1 ao final deste artigo.

Autores como Jeffrey Sachs (2012) apontam uma série de diferenças entre os ODM e os ODS, conforme detalha no Quadro 1.

No documento “Data for development: an action plan to finance the data revolution for sustainable development” (OPEN DATA WATCH, 2015, p. 8, tradução nossa), elaborado pela ONU e pelo Open Data Watch, afirma que

os ODS vão depender mais de observações geoespaciais e terrestres que os ODM. Imagens de satélite estão cada vez mais disponíveis gratuitamente em uma resolução moderada e com custos progressivamente menores para fontes de alta resolução. Dados de satélite têm o potencial para serem utilizados no monitoramento de mais de 23 indicadores dos ODS em potencial, desde a medição da qualidade do ar global até cobertura de florestas e cultivos, impactos de desastres e recursos hídricos. Novas imagens via satélite são um exemplo de tecnologias emergentes que oferecem oportunidades significativas para uma plataforma de monitoramento de recursos hídricos.

Quadro 1:

Diferenças entre os Objetivos de Desenvolvimento do Milênio (ODM) e os Objetivos de Desenvolvimento Sustentável (ODS)

	ODM	ODS
Duração	2000-2015	2016-2030
Pontos de controles intermediários (SACHS, 2012)	Ausentes	Presentes
Complexidade do monitoramento (INTERNATIONAL ORGANISATION OF SUPREME AUDIT INSTITUTIONS, 2015)	8 objetivos, 20 metas	17 objetivos, 196 metas
Abrangência geográfica (SACHS, 2012)	Países em desenvolvimento	Todos os países signatários
Formas de monitoramento	Não definidas previamente	Negociadas entre os governos; desenvolvimento de sistema de monitoramento
Papel da tecnologia (SACHS, 2012)	Secundário	Fundamental
Foco (NAÇÕES UNIDAS, 2014e)	Alívio da pobreza	Desenvolvimento sustentável em sentido amplo

Fonte: Elaboração dos autores

Não há dúvidas de que a coleta de dados compreende um grande desafio. Por isso, deve-se fazer uso de novas tecnologias e métodos disponíveis, como aqueles fornecidos por meio das tecnologias de informação geográfica e *big data* (NAÇÕES UNIDAS, 2015a). Como há cada vez mais dados disponíveis, o estudo de métodos com utilização de bases de dados vem ganhando maior destaque (FREITAS; DACORSO, 2014).

Dada a amplitude e complexidade da Agenda 2030, são necessários diversos tipos de dados com diferentes níveis de cobertura (SUSTAINABLE DEVELOPMENT SOLUTIONS NETWORK, 2015). Nesse contexto, cada tipo de dado dá sustentação e também sustenta os outros tipos.

Para o monitoramento dos ODS são necessários a composição e o fomento de um ecossistema de dados. Por esse motivo, a ONU especifica no relatório “Data for development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development” (SDSN, 2015) a tipologia dos dados, esclarece quais são as fontes de dados primários e estabelece os princípios-chave para a seleção de indicadores de monitoramento robustos.

A tipologia dos dados refere-se a dados de censo, registros civis e estatísticas sobre nascimento, pesquisas domiciliares, pesquisas agrícolas, dados administrativos, estatísticas econômicas e dados sobre meio ambiente, incluindo dados geoespaciais. Estes últimos são cruciais para determinar indicadores ambientais dos ODS, assim como para análise desagregada dos indicadores socioeconômicos destes (SDSN, 2015).

Quadro 2:

Fontes de dados primários para os indicadores dos ODS

Fonte de dados primários	
Dados administrativos	33%
Pesquisas domiciliares	26%
Dados internacionais	13%
Dados sobre meio ambiente (pesquisas agrícolas ou dados geoespaciais)	11%
Registro civil e estatísticas de nascimento	8%
TBD	6%
Pesquisas sobre força de trabalho	2%
Outros dados econômicos	2%
Censos	Transversal
TOTAL	100%

Fonte: Sustainable Development Solutions Network (2015, tradução e grifo nosso)

Das fontes primárias dos indicadores propostos por esse relatório (SDSN, 2015), 11% estão relacionadas aos dados sobre meio ambiente, estes que incluem dados geoespaciais. O Quadro 2 mostra as fontes de dados primários para esses indicadores.

Outro componente da ecologia dos dados remete aos princípios-chave para selecionar indicadores de monitoramento global, propostos pelo relatório “Follow-up and Review of the SDGs: fulfilling our commitments” (NAÇÕES UNIDAS, 2015b). De acordo com esses princípios, os indicadores precisam ser: limitados em número e em concordância global (para indicadores de monitoramento global); simples (por exemplo, indicadores de uma variável com implicações políticas diretas); apresentados com alta frequência, o que permite monitoramento regular, preferencialmente anual; baseados em consenso, alinhado com padrões internacionais e informações baseadas em sistemas; construídos a partir de fontes de dados bem estabelecidas; mais desagregados quanto possível; universais; focados no resultado; baseados em dados científicos; e ter uma boa aproximação (*proxy*) para questões e condições mais amplas.

Parte integrante da ecologia de dados são as ferramentas e metodologias para coleta, análise e comunicação dos dados. Essas ferramentas e metodologias estão intrinsecamente relacionadas aos campos geoespaciais. Trata-se da fotogrametria, da cartografia, dos sistemas de informação geográfica (SIG) e análises geoespaciais, de sensoriamento remoto e da inteligência geoespacial.

Desses campos, há alguns que são especialmente relevantes para o monitoramento dos ODS pelas EFS: a cartografia²; os sistemas de informação geográfica³ e o sensoriamento remoto⁴. Esses campos fazem parte do que caracterizamos neste trabalho como geotecnologias, ou seja: “o conjunto de tecnologias para coleta, processamento, análise e disponibilização de informações georreferenciadas” (SOUZA, 2016, p. 40).

No evento “Unleashing the power of ‘Where’ to make the world a better place: How geographic information contributes to achieving the SDGs” (NAÇÕES UNIDAS, 2015f), Lawrence Friedl, representando a National Aeronautics and Space Administration (NASA), afirmou que os ODS chegaram em um momento de convergência privilegiada para aproveitar as potencialidades dos dados espaciais. A informação geográfica é um elemento essencial no complexo contexto da implementação e monitoramento da Agenda 2030 (Id., 2015a), sendo mencionada na meta número 18 do ODS número 17, que busca o fortalecimento dos meios de implementação e a revitalização da parceria global para



o desenvolvimento sustentável como uma das fontes relevantes de dados. A “meta 17.18” objetiva incentivar o desenvolvimento de capacidades em países em desenvolvimento até 2020, visando ampliar significativamente a disponibilidade de dados confiáveis, atualizados e de alta qualidade, desagregados, por exemplo, por localização geográfica (Id., 2015c).

Ao analisar os indicadores nacionais propostos pelo Sistema das Nações Unidas no Brasil (PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO, 2015), percebe-se que uma quantidade significativa desses necessita de desagregação geográfica. A maioria dos indicadores propostos utiliza alguma forma de desagregação geográfica e, de acordo com o relatório “Acompanhando a Agenda 2030 para o desenvolvimento sustentável”, a desagregação de dados que a nova agenda apresenta é um desafio a ser superado.

3. REFERÊNCIAS TÉCNICAS DA INTOSAI QUE ABORDAM O TEMA GEOTECNOLOGIAS

O documento “ISSAI 5130 – Sustainable development: the role of Supreme Audit Institutions” (INTERNATIONAL ORGANISATION OF SUPREME AUDIT INSTITUTIONS, 2004), ao tratar sobre as novas atribuições derivadas dos ODS, reconhece que as instituições precisarão questionar se as metodologias existentes para realizar as auditorias são adequadas para esse contexto.

Também afirma que a depender das especificidades de cada EFS, pode-se incluir especialistas no processo, seja por contratação ou por consultoria.

Já “Auditing forests: guidance for Supreme Audit Institutions” ((INTERNATIONAL ORGANISATION OF SUPREME AUDIT INSTITUTIONS, 2010), elaborado pelo Grupo de Trabalho em Auditoria Ambiental (WGEA) da Intosai, descreve como as EFS podem utilizar a tecnologia do Sistema de Informação Geográfica (SIG). Ao tratar de aspectos metodológicos, o documento afirma que “tecnologias computacionais podem ser excepcionalmente úteis em auditorias. Dois exemplos dessas tecnologias são o GPS e o SIG” ((INTERNATIONAL ORGANISATION OF SUPREME AUDIT INSTITUTIONS, 2010, p. 9). Dados geográficos podem ser utilizados para diversas finalidades e em diferentes fases da auditoria. O SIG pode ser utilizado na fase de planejamento e o GPS na fase de execução como uma ferramenta de suporte. Dados geográficos podem ser empregados em diversas finalidades, como em diferentes fases do planejamento, aplicativos baseados em redes viárias, aplicativos baseados em recursos naturais, análises panorâmicas, gestão de instalações, entre outros (INTERNATIONAL ORGANISATION OF SUPREME AUDIT INSTITUTIONS, 2010).

De acordo com o documento “Environmental data: resources and options for Supreme Audit Institutions” (INTERNATIONAL ORGANISATION OF SUPREME AUDIT INSTITUTIONS, 2013a) tanto os

gestores públicos auditados quanto as EFS se beneficiarão do uso de SIG. O referido documento descreve dados geoespaciais separadamente, em uma seção específica, por considerar que estes trazem considerações únicas para as EFS. Devido a restrições orçamentárias dos governos, os gestores públicos sentem-se pressionados a demonstrar resultados, o que os leva a utilizar em maior escala dados ambientais para demonstrar que seus programas atingiram os objetivos estabelecidos. Esse tipo de mudança pode afetar como o desempenho é medido e como gestores públicos e EFS avaliam esses programas (INTERNATIONAL ORGANISATION OF SUPREME AUDIT INSTITUTIONS, 2013a).

O referido documento afirma que fontes de dados espaciais podem ser especialmente úteis para EFS quando estas estão realizando a verificação de questões ambientais que tenham aspecto geográfico explícito, como áreas de proteção ambiental ou identificação de áreas poluídas. Dados geoespaciais são também úteis para selecionar amostras de diferentes locais, para identificar áreas de alto risco e padrões nos dados, o que não seria possível se estes não tivessem seu componente espacial. As EFS também podem utilizar os dados espaciais como forma de apresentar seus resultados, tornando-os mais tangíveis.

Por fim, no “Environmental Data: Resources and Options for Supreme Audit Institutions” (INTOSAI, 2013a) afirma-se que a utilização de dados atrelados a informações geográficas torna as bases de dados mais complexas pela necessidade de se registrar o quê e onde está acontecendo, utilizando-se, para isso, de coordenadas geográficas. Dessa forma, gera-se uma grande demanda de controle de qualidade, além do correspon-

dente aumento em complexidade para se avaliar a base de dados. Por isso, as EFS que consideram a possibilidade de utilizar esse tipo de dado necessitam investir em acesso às ferramentas e também no desenvolvimento das competências técnicas necessárias.

No relatório “The 7th survey on environmental auditing” (Id., 2012), a WGEA apresenta o resultado da pesquisa respondida por mais de 112 EFS. A pesquisa solicitou que as EFS descrevessem metodologias inovadoras de trabalho que vinham aplicando em auditorias ambientais. A tecnologia geoespacial foi o item mais votado⁵.

A *ISSAI 5540* (INTERNATIONAL ORGANISATION OF SUPREME AUDIT INSTITUTIONS, 2013b), que aborda questões relacionadas à utilização de dados geoespaciais em auditorias de gestão de desastres ambientais, afirma que os SIG e o sensoriamento remoto podem fornecer valor agregado a todos os estágios da auditoria, como demonstra o Quadro 3⁶. Segundo o documento, os SIG podem auxiliar na análise extensiva e complexa de dados de diversas formas: apresentar os dados espacialmente, consultar dados por localização espacial, analisar localizações espaciais e armazenar e visualizar dados em camadas, entre outras. Os SIG permitem que os usuários produzam mapas de alta qualidade em qualquer escala, armazenem grande quantidade de informações geográficas, visualizem dados complexos e produzam novos dados com base nos dados existentes.

A *ISSAI 5540* propõe um *checklist* sobre a utilização de informações geoespaciais em auditorias, reproduzido integralmente no Quadro 4.

A referida *ISSAI* também afirma que a utilização de informações geoespaciais e do SIG no setor público

Quadro 3:

Aplicação das geotecnologias às diversas fases da auditoria de acordo com a *ISSAI 5540*

ETAPA DA AUDITORIA	APLICAÇÃO DAS GEOTECNOLOGIAS
Avaliação de riscos relevantes	Os SIG tornam possível a análise de vários dados ou camadas de dados em um contexto geográfico. O sensoriamento remoto pode ser utilizado para verificar informações em bases de dados com informações de campo
Planejando a auditoria	Os SIG e o sensoriamento remoto podem auxiliar nesse ponto ao decidir sobre o foco da auditoria
Conduzindo a auditoria	A equipe pode utilizar dispositivos de GPS e mapas baseados em satélite para conectar dados de auditoria em campo a dados geográficos. Dessa forma, é possível analisar os dados de campo imediatamente após carregá-los e combiná-los com mapas
Análise dos resultados da auditoria	Com os SIG é possível analisar diferentes camadas de informação geográfica, análise a partir da qual o desempenho pode ser medido. Ainda, a visualização de resultados com os SIG permite a visão sobre as diferenças geográficas no desempenho das organizações públicas
Comunicação dos resultados	Utilizando os SIG e o sensoriamento remoto, resultados de auditorias podem ser mapeados e apresentados para apoiar as principais conclusões e recomendações das auditorias e facilitar a comunicação dos resultados. A visualização dos resultados das auditorias em mapas permite a transmissão de uma mensagem clara e forte em comparação àquela somente escrita

Fonte: Adaptado de International Organisation of Supreme Audit Institutions (2013b, tradução nossa)

Quadro 4:

Checklist sobre utilização de dados geoespaciais em auditorias

Checklist: utilização de dados geoespaciais em auditorias
Que dados geoespaciais são necessários para responder às perguntas da auditoria?
Qual a acurácia dos dados geoespaciais requerida?
Qual é o prazo exigido para os dados geoespaciais?
Que dados geoespaciais estão disponíveis?
De que fontes os dados geoespaciais podem derivar e quão confiáveis elas são?
Qual a qualidade dos dados geoespaciais disponíveis?
Quais os custos dos dados geoespaciais disponíveis?
Se os dados geoespaciais requeridos não estiverem disponíveis, eles podem ser coletados como parte do processo e do orçamento da auditoria?
Os auditores envolvidos possuem o conhecimento necessário para coletar e analisar os dados geoespaciais requeridos ou é necessário incluir um especialista externo?

Fonte: International Organisation of Supreme Audit Institutions (2013b, tradução nossa)

tem crescido por diversas razões. Uma das principais é a extensão e a complexidade da informação, que deve ser considerada e analisada enquanto tomam-se decisões. Muitas decisões requerem informações geoespaciais e os SIG apoiam a análise desse tipo de informação. Dentro do contexto de políticas públicas, as informações geoespaciais cumprem diversas funções, como a definição de objetivos, a formulação de medidas, o monitoramento e a avaliação. Além disso, são muitas as áreas da política pública em que se podem utilizar informações geoespaciais. Exemplos⁷ delas são gestão de recursos naturais, proteção ao meio ambiente, economia, educação, segurança e saúde (INTERNATIONAL ORGANISATION OF SUPREME AUDIT INSTITUTIONS, 2013b).

4. CONSIDERAÇÕES FINAIS

Dentre as iniciativas propostas pela ONU para que os ODS superem os desafios encontrados pelos Objetivos de Desenvolvimento do Milênio, está a utilização de dados geoespaciais (NAÇÕES UNIDAS, 2015e). Além disso, nas Normas Internacionais das IS-SAI (INTERNATIONAL ORGANISATION OF SUPREME AUDIT INSTITUTIONS, 2004) são reconhecidas a importância da capacitação de seus servidores para atender às novas demandas.

Este artigo se propôs a fazer a revisão da literatura sobre as referências técnicas das Nações Unidas e da Intosai que pudessem contribuir com a utilização de geotecnologias para o monitoramento dos ODS pelas Entidades de Fiscalização Superior. Espera-se que essa

sistematização subsidie futuros estudos a serem desenvolvidos no âmbito das EFS sobre o tema.

NOTAS

- 1 Por exemplo, a aplicação de imagens de satélite ou o uso do sensoriamento remoto para ações de controle.
- 2 Disciplina que se encarrega da concepção, produção, disseminação e estudo de mapas tanto como objetos tangíveis quanto digitais.
- 3 Qualquer sistema que capture, armazene, administre e visualize dados associados à localização.
- 4 Ciência de mensurar a propriedade de algum objeto ou fenômeno por um sensor que não tem contato físico com o objeto/fenômeno estudado.
- 5 Com a mesma quantidade de votos, compartilhando o primeiro lugar como metodologia inovadora por essa pesquisa, ficou o uso de expertise externa.
- 6 Cabe ressaltar as similaridades entre as etapas da auditoria e algumas das funções descritas no DACUM Research Chart for Geospatial Analyst (capture data, manage data, analyse data, produce deliverables). Disponível em: <<http://bit.ly/2gqLhVT>>. Acesso em: 12 jul. 2016.
- 7 O artigo denominado "O uso de geotecnologias como uma nova ferramenta para o controle externo" (2015), publicado na edição nº 133 da Revista do TCU, exemplifica oito aplicações de geotecnologias das quais o controle público pode se beneficiar.

Anexo 1:

Exemplos de suporte das geotecnologias aos Objetivos de Desenvolvimento Sustentável (Nações Unidas, 2015)

ODS	INDICADOR	DESCRIÇÃO DO INDICADOR	DESAGREGAÇÃO	FONTE ¹		
1	Acabar com a pobreza em todas as suas formas, em todos os lugares	6	Perdas por desastres naturais, relacionados ou não a mudanças climáticas (em dólares ou vidas perdidas)	Mede perdas humanas e financeiras em áreas rurais e urbanas devidas a desastres naturais, desagregadas em eventos relacionados ou não a mudanças climáticas	Este indicador pode ser desagregado espacialmente (incluindo a segregação entre urbano e rural)	Fontes não geoespaciais
2	Acabar com a fome, alcançar a segurança alimentar e melhoria da nutrição e promover a agricultura sustentável	13	Lacuna no rendimento de culturas	Acompanha as lacunas de desenvolvimento das principais culturas, ou seja, rendimento atual comparado àquele que seria possível em perfeitas condições	Adequado para desagregação espacial, de escalas locais a globais	Dados geoespaciais, incluindo sensoriamento remoto e satélite
9	Construir infraestrutura resiliente, promover a industrialização inclusiva e sustentável e fomentar a inovação	58	Acesso durante todo ano a estradas transitáveis (% acesso a estradas dentro de um raio de [x] km de distância)	Acesso a estradas transitáveis durante todo o ano é questão crítica para processos de desenvolvimento rural, incluindo acesso a insumos, mercados, educação e serviços de saúde. Este indicador acompanha a parte da população que vive dentro de um raio de [x] km de estradas que são transitáveis durante todo o ano	Este indicador pode ser desagregado espacialmente	Dados geoespaciais, incluindo sensoriamento remoto e satélite
11	Tornar as cidades e os assentamentos humanos inclusivos, seguros, resilientes e sustentáveis	69	A média de poluição do ar urbano de material particulado (MP10 e MP2.5)	Acompanhamento da média de poluição do ar urbano	Por cidade e estado	Dados geoespaciais, incluindo sensoriamento remoto e satélite
12	Assegurar padrões de produção e de consumo sustentáveis	75	Profundidade ótica de aerossol	Medida dos aerossóis totais (por exemplo: partículas de fumaça, poeira, sal do mar) distribuídos em uma coluna de ar da superfície da Terra até a parte superior da atmosfera	Este indicador pode ser reportado com um alto grau de desagregação espacial (incluindo cidades e no nível de bairro)	Dados geoespaciais, incluindo sensoriamento remoto e satélite
15	Proteger, recuperar e promover o uso sustentável dos ecossistemas terrestres, gerir de forma sustentável as florestas, combater a desertificação, deter e reverter a degradação da terra e deter a perda de biodiversidade	83	Varição anual em área florestal e área cultivada (indicador alterado dos ODM)	Acompanha a variação líquida de área florestal e a expansão da agricultura em ecossistemas naturais, assim como a perda de terrenos produtivos para o crescimento de áreas urbanas, indústrias, estradas e outros usos	Este indicador pode ser desagregado espacialmente	Dados geoespaciais, incluindo sensoriamento remoto e satélite
		85	Varição anual em área degradada ou desertificada (% ou ha)	Componentes de degradação do solo incluem a salinização, a erosão, a perda de nutrientes do solo e o assoreamento	Desagregação geográfica por sub-região	Dados geoespaciais, incluindo sensoriamento remoto e satélite

Fonte: Adaptado de Nações Unidas (2015c, tradução nossa)

REFERÊNCIAS

CÂMARA, G. Representação computacional de dados geográficos. In: CASANOVA, M. et al. Bancos de Dados geográficos. Curitiba: MundoGEO, 2005. p. 1-44.

CASSIDY, M. Assessing gaps in indicator availability and coverage. 2014. Disponível em: <<http://bit.ly/2glULel>>. Acesso em: 8 jul. 2016.

FREITAS, R. K. V. de; DACORSO, A. L. R. Inovação aberta na gestão pública: análise do plano de ação brasileiro para a Open Government Partnership. Revista de Administração Pública, Rio de Janeiro, v. 48, n. 4, p. 869-888, 2014.

INDEPENDENT EXPERT ADVISORY GROUP. A world that counts: mobilising the data revolution for sustainable development UN report. 2015. 30 p. Disponível em: <<http://bit.ly/1DcbDol>>. Acesso em: 8 jun. 2016.

INTERNATIONAL ORGANISATION OF SUPREME AUDIT INSTITUTIONS. Auditing forests: guidance for supreme audit institutions. 2010. 71 p. Disponível em: <<http://bit.ly/2fJOrk5>>. Acesso em: 12 jul. 2016.

_____. Environmental data: resources and options for supreme audit institutions. 2013a. 64 p. Disponível em: <<http://bit.ly/2gEOStn>>. Acesso em: 12 jul. 2016.

_____. ISSAI 5130. Exposure draft. Sustainable on development: the role of supreme audit institutions. 2004. 53 p. Disponível em: <<http://bit.ly/2fcy2Br>>. Acesso em: 12 jul. 2016.

_____. ISSAI 5540. Use of geospatial information in auditing disaster management and disaster-related aid. 2013b. 36 p. Disponível em: <<http://bit.ly/2glYicn>>. Acesso em: 14 jul. 2016.

_____. The 7th survey on environmental auditing. 2012. 77 p. Disponível em: <<http://bit.ly/2fvUVOT>>. Acesso em: 12 jul. 2016.

_____. The role of sais and means of implementation for sustainable development (lessons learned from MDGs). 2015. Disponível em: <<http://bit.ly/2gkLcc0>>. Acesso em: 9 jul. 2016.

NAÇÕES UNIDAS. Ambassador statement: unleash ing the power of “Where” to make the world a better place: How geographic information contributes to achieving the SDGs. 2015a. 3 p. Disponível em: <<http://bit.ly/2fJOZ2z>>. Acesso em: 5 jul. 2016.

_____. Assessing gaps in indicator availability and coverage. 2014. 9 p. Disponível em: <<http://bit.ly/2glULel>> Acesso em: 5 jun. 2016.

_____. Follow-up and Review of the SDGs: fulfilling our commitments. 2015b. 39 p. Disponível em: <<http://bit.ly/2fvQkMQ>>. Acesso em: 9 jul. 2016.

_____. Indicators and a monitoring framework for the sustainable development goals: launching a data revolution for the SDGs. 2015c. 225 p. Disponível em: <<http://bit.ly/1DMsAfp>>. Acesso em: 9 jul. 2016.

_____. Resolution A/66/209. Promoting the efficiency, accountability, effectiveness and transparency of public administration by strengthening supreme audit institutions. 2011. Disponível em: <<http://bit.ly/2fctLy9>> Acesso em: 5 jul. 2016.

_____. Resolution A/69/228. Promoting and fostering the efficiency, accountability, effectiveness and transparency of public administration by strengthening supreme audit institutions, 2014a. 3 p. Disponível em: <<http://bit.ly/2gETHmm>>. Acesso em: 5 jun. 2016.

_____. 23e Symposium ONU/INTOSAI. 2015d. 8 p. Disponível em: <<http://bit.ly/2fJRstl>>. Acesso em: 5 jun. 2016.

_____. The millennium development goals report 2015. 2015e. 73 p. Disponível em: <<http://bit.ly/1Rg5uUm>>. Acesso em: 5 jul. 2016.

_____. Unleashing the power of “Where”: how geographic information contributes to achieving the SDGs. 2015f. 24 p. Disponível em: <<http://bit.ly/2fcsLth>>. Acesso em: 9 jul. 2016.

_____. NoTheme I: Sustainable Development Goals. How INTOSAI can contribute to the UN post 2015 agenda including good governance in order to strengthen the fight against corruption? Abu Dhabi: Incosai, 2016.

OPEN DATA WATCH. Data for development: an action plan to finance the data revolution for sustainable development. 2015. 81 p. Disponível em: <<http://bit.ly/2fvQaVD>>. Acesso em: 9 jul. 2016.

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO. Acompanhando a agenda 2030 para o desenvolvimento sustentável: subsídios iniciais do Sistema das Nações Unidas no Brasil sobre a identificação de indicadores nacionais referentes aos Objetivos de Desenvolvimento Sustentável. Brasília: PNUD, 2015. 250 p.

SACHS, J. D. From millennium development goals to sustainable development goals. *The Lancet*, London, v. 379, p. 2206-2211, 2012.

SUSTAINABLE DEVELOPMENT SOLUTIONS NETWORK. Data for development: a needs assessment for SDG monitoring and statistical capacity development. 2015. 81 p. Disponível em: <<http://bit.ly/1zeUVcf>>. Acesso em: 9 jul. 2016.

SHUANG, C. et al. Towards a Post-2015 Framework that Counts: Development National Statistical Capacity. 2013. Disponível em: <<http://bit.ly/2g8V35c>>. Acesso em: 9 jul. 2016.

SOUZA, A. D. Modelo de controle para obras de esgotamento sanitário utilizando sistema de informações geográficas. 2016. Tese (Doutorado em Engenharia Civil) – Universidade Federal de Pernambuco, Recife, 2016.

InfoSAS: um sistema de mineração de dados para controle da produção do SUS



Osvaldo Carvalho

é doutor de Estado em Ciência da Computação pela Universidade Pierre et Marie Curie e professor associado da UFMG. Trabalhou com algoritmos distribuídos, e hoje leciona Programação de Computadores e desenvolve sistemas de informação.



Wagner Meira Jr.

é PhD pela Universidade de Rochester e professor titular de Ciência da Computação na Universidade Federal de Minas Gerais. Suas áreas de interesse são mineração de dados, sistemas paralelos e distribuídos e suas aplicações.



Marcos Prates

é bacharel pela UFMG (2006) em Matemática Computacional, mestre pela UFMG (2008) em Estatística e PhD em Estatística pela University of Connecticut, EUA (2011). Desenvolve métodos estatísticos e algoritmos para análise de estatística espacial e aprendizado de máquina.



Renato M. Assunção

é PhD em Estatística pela University of Washington/EUA e professor titular do Departamento de Ciência da Computação da UFMG, além de especialista em desenvolvimento de algoritmos e métodos para a análise de dados estatísticos, especialmente aqueles com georreferenciamento.



Raquel Minardi

é bacharel em Ciência da Computação, doutora pela UFMG e pós-doutora pelo Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA). Desenvolve modelos e algoritmos em visualização de dados e em biologia computacional. Publicou mais de 20 artigos em periódicos internacionais e em conferências nacionais e internacionais.



José Nagib Cotrim Árabe

é PhD em Ciência da Computação pela University of California, Los Angeles (UCLA), além de professor da Universidade Federal de Minas Gerais e chefe do Departamento de Ciência da Computação.



RESUMO

Este trabalho apresenta o InfoSAS, um sistema de detecção de anomalias estatísticas nos registros da produção do SUS. O InfoSAS encontra taxas de atendimentos por habitante muito superiores à média nacional, ou valores de internação bem acima dos praticados pela maioria dos estabelecimentos para um mesmo procedimento. Os resultados encontrados indicam que centenas de milhões de reais gastos pelo SUS são destinados a atendimentos considerados anômalos por critérios conservadores. Anomalias estatísticas podem ser provocadas por fraudes, mas também por mutirões de saúde, epidemias, ou má distribuição do atendimento. Em qualquer caso, anomalias graves devem ser investigadas ou explicadas.

Palavras-chave: SUS, Sistema Único de Saúde, Detecção de anomalias, Mineração de dados, Taxa de atendimentos por habitante, Valor médio de internações, Distribuição log-normal, Priorização de auditorias.

1. INTRODUÇÃO

A grande dimensão do setor de saúde e o enorme volume de recursos envolvidos fazem dele um alvo atraente para fraudes em todo o mundo. Nos Estados Unidos, por exemplo, estima-se que mais de 270 bilhões de dólares sejam perdidos anualmente com fraudes (THE ECONOMIST, 2014). Não temos motivos para acreditar

que no Brasil a situação seja diferente. Neste documento apresentamos o InfoSAS, um sistema de mineração e visualização de dados do SUS que, como diversos outros (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996), identifica padrões e modelos variados nos dados, podendo ser de grande valia para a gestão do SUS e, em particular, para a identificação de anomalias estatísticas que podem ser indícios de irregularidades. O foco deste artigo é a sua utilização para fins de auditoria.

O InfoSAS encontra em profusão fatos que merecem a atenção de gestores, como taxas de atendimentos por habitante muito superiores à média nacional, como se pode ver na Figura 2; ou valores de internação bem acima dos praticados pela maioria dos estabelecimentos, como mostrado na Figura 4. Temos elementos, expostos na Seção 5, para acreditar que sua utilização na priorização de auditorias pode gerar retornos para o SUS da ordem de R\$400 milhões por ano. Não é exagero estimar em bilhões de reais o valor que pode ser recuperado com auditorias a serem feitas sobre a produção de um período de 8 anos (de 2009 a 2016), assim como com a inibição de comportamentos anômalos na produção futura.

2. DISCREPÂNCIAS ESTATÍSTICAS

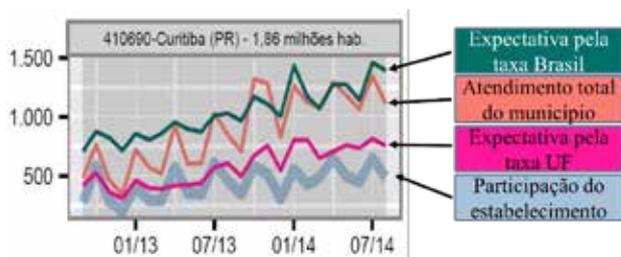
Os atendimentos prestados pelo SUS (denominados conjuntamente de produção do SUS) são registrados por diversos instrumentos e esses registros são armazenados nas bases SIA¹ e SIH². O SUS também mantém

a base CNES³, um cadastro de estabelecimentos. O sistema InfoSAS examina essas bases do SUS e também dados populacionais fornecidos pelo IBGE, e produz o que chamamos de folhas de fatos.

As bases de dados referidas são volumosas e essas folhas de fatos dão foco ao seu exame. Elas relacionam um alvo de mineração, um período de análise e um estabelecimento prestador, mostrando gráficos e tabelas obtidas pelo exame das bases de dados. Um alvo de mineração é um subconjunto da produção do SUS, definido por um ou mais procedimentos da Tabela do SUS⁴, e algumas vezes também pela faixa etária dos pacientes. É importante observar que fatos presentes nas folhas podem também ser verificados de forma independente consultando outras fontes de informação oferecidas pelo SUS, como o TabWin⁵.

Figura 1:

Análise da participação de um estabelecimento no atendimento de Curitiba para o alvo “Tratamento de doenças do aparelho da visão”



A Figura 1 mostra um exemplo de gráfico presente em uma folha de fatos. Nele vemos que, para o alvo “Tratamento de doenças do aparelho da visão”, o atendimento total de Curitiba, no período de setembro de 2013 a agosto de 2014, seguiu de perto a expectativa segundo a média brasileira, que este atendimento é superior à média do Paraná, e que o prestador contribuiu com menos da metade desse atendimento. Essas são informações compatíveis com uma situação de normalidade.

Mas nem sempre é assim. Alguns fatos chamam atenção do profissional de controle. Na Figura 2, diversos municípios estão com uma taxa de atendimentos por habitante muito acima da brasileira e da estadual, e o estabelecimento em questão é praticamente o único a atender esses municípios. Este é um exemplo claro do que consideramos anomalias ou discrepâncias estatísticas.

Discrepâncias estatísticas devem ser consideradas com cuidado, como ilustra o exemplo mostrado na Figura 3. Cada um dos seis gráficos mostrados se refere a um município e todos a um único prestador. Pode-se notar que somente o gráfico no canto superior esquerdo (que corresponde ao município sede do estabelecimento) não apresenta picos estranhos, com valores muito acima dos esperados pelas taxas brasileiras e da UF onde se localizam os municípios. Um indício de fraude? Não. Quando se nota que o alvo é “Mamografia” e que o prestador é uma unidade móvel, a anomalia estatística se explica: o prestador é um veículo equipado com um

Figura 2:

Análise da taxa de atendimento por habitante de 6 municípios atendidos por um estabelecimento para o alvo “Tratamento de doenças do aparelho da visão”

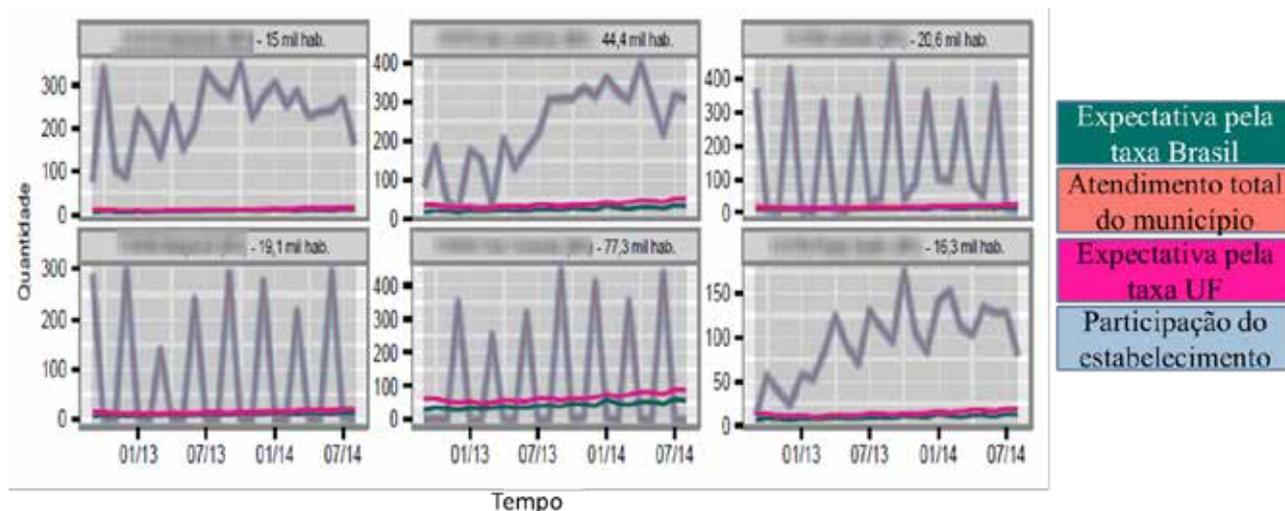
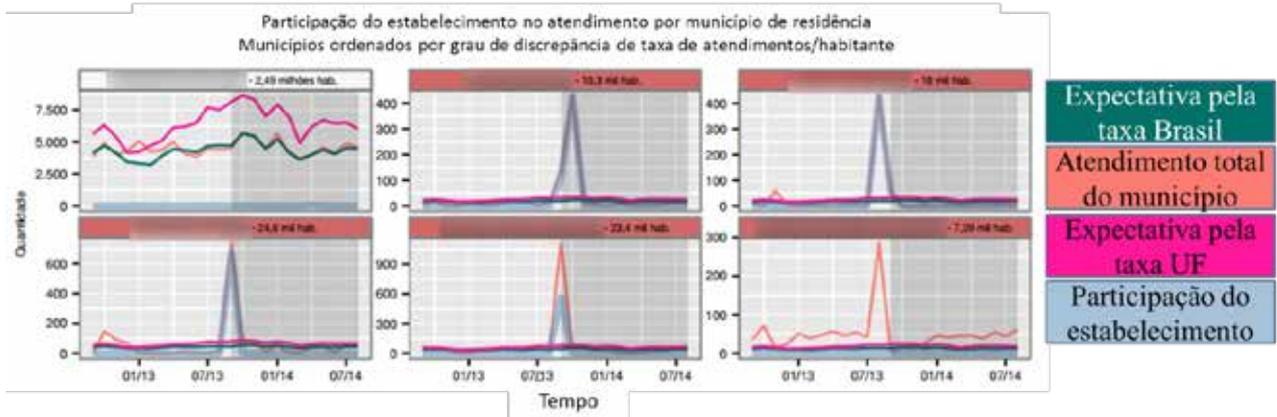


Figura 3:

Discrepâncias estatísticas explicáveis pelas características do alvo, “Mamografia bilateral para rastreamento”, e do prestador, uma Unidade Móvel de Saúde da Mulher



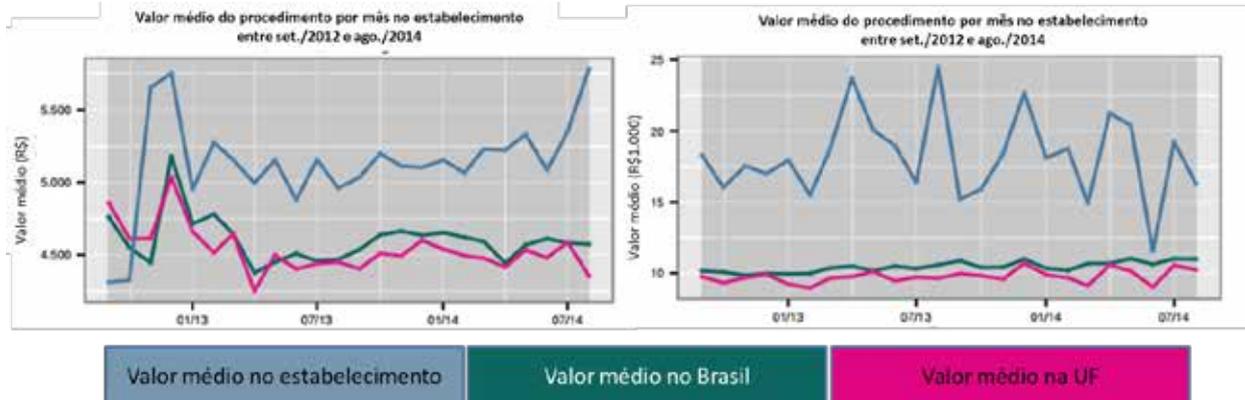
mamógrafo que realiza mutirões em pequenas cidades, provocando em suas visitas os picos de atendimentos.

De forma geral, é importante observar que uma anomalia estatística pode sim ser um indício de fraude, mas também pode resultar de processos corretos, como no exemplo acima, ou de informação incorreta, ou de epidemias. E é também possível que municípios com taxas de atendimento altas sejam, na verdade, os poucos bem atendidos no país. Portanto, a análise de uma folha de fatos deve ser feita com o máximo de discernimento e por um profissional de saúde pública que conheça o contexto local.

Na Figura 4 temos outros exemplos de discrepâncias estatísticas, agora com estabelecimentos praticando preços médios de internação muito superiores aos praticados no restante do Brasil para dois alvos de mineração. Novamente, é preciso que o gestor examine com cuidado cada caso de discrepância estatística, pois é possível que o estabelecimento tenha um perfil de pacientes com complicações acima da média, ou que utilize práticas que podem elevar o custo da internação, mas que também, digamos, diminuam sua taxa de óbitos.

Figura 4:

Análise do valor cobrado por um estabelecimento para internações dos alvos “Artroplastia total primária do quadril não cimentada/híbrida”, à esquerda, e por outro estabelecimento para “Cirurgia cardiovascular – Cirurgia cardiovascular marcapasso”, à direita



3. DETECÇÃO AUTOMÁTICA DE DISCREPÂNCIAS ESTATÍSTICAS

Encontrar folhas de fatos de interesse para o controle em uma Secretaria de Saúde é um fator de conveniência para a decisão de apuração. Entretanto, buscar ao acaso discrepâncias relevantes é como procurar uma agulha em um palheiro. Temos mais de 5000 alvos, aproximadamente 6000 prestadores e, considerando apenas períodos de 12 meses, em 3 anos de produção, seriam 36 janelas de tempo possíveis. Um cálculo simples mostra que temos, literalmente, bilhões de folhas de fatos que podem ser extraídas das bases de dados examinadas.

O InfoSAS mostra seu valor nesse momento, pois utiliza diversos algoritmos que procuram capturar discrepâncias, produzindo escores que permitem ordenação e priorização das folhas de fatos. Para a seleção, o InfoSAS também permite ao usuário concentrar-se em áreas definidas por filtros geográficos, por período de análise e por alvos, pois um profissional de controle e avaliação tem, muitas vezes, sua atenção dirigida para setores específicos da saúde, como cardiologia ou ortopedia e, claro, maior interesse em sua região de atuação.

O InfoSAS analisa as séries temporais de valor médio mensal por procedimento e de produção mensal em cada alvo desejado. Tais séries são calculadas por estabelecimento e por município de residência dos pacientes. Vários algoritmos de detecção de anomalias são utilizados e cada um deles calcula um escore. Estes escores são posteriormente combinados. Por falta de espaço para descrever em detalhes todos os escores, vamos apresentar a seguir a definição de dois deles. Uma descrição mais detalhada de outro algoritmo pode ser encontrada em (CARVALHO et al., 2015).

Um algoritmo tem como objetivo detectar oscilações abruptas na produção de um prestador. Para isso, a produção do estabelecimento é comparada com sua própria série histórica de produção (ou de valor médio). Se houver uma mudança brusca na série, o algoritmo atribui um escore de discrepância ao estabelecimento. Mais especificamente, em cada mês i e para cada estabelecimento l , nós calculamos $escore_{li} = (t_{li} + 1) / (mediana_{li} + 1)$, onde t_{li} é a dimensão de interesse (que pode ser o valor ou o número de atendimentos) para o l -ésimo estabelecimento no i -ésimo mês, $mediana_{li}$ é a mediana dos últimos m meses para a quantidade de interesse do l -ésimo estabelecimento no i -ésimo mês. Tipicamente, tomamos $m=12$ ou $m=6$ meses.

Outro algoritmo procura detectar discrepâncias em taxas de atendimento por habitante. Tomando como base a produção brasileira por 100 mil habitantes, a série



temporal de produção realizada por todos os estabelecimentos nos residentes de um município é analisada. Caso este município tenha uma produção acima do limiar, o algoritmo atribui um escore de discrepância a ele. Em seguida, o escore é atribuído aos estabelecimentos de forma proporcional à participação de cada um no atendimento a esse município.

O escore do estabelecimento é obtido somando-se os escores que ele obteve em cada município de atuação. Mais formalmente, o escore de produção do município é calculado como uma soma acumulada nos últimos 12 meses, $escore_{li} = \sum_{j=i-11}^i dif_{lj}$, onde $dif_{li} = \max\{0, tBayes_{li} - limiar_i\}$ e $limiar_i = k * taxaBrasil_i$. O valor de $taxaBrasil_i$ é a taxa de produção mensal brasileira por 100 mil habitantes no mês i , $tBayes_{li}$ é a taxa bayesiana empírica por 100 mil habitantes no mês i para o município l , e k é uma constante previamente definida. Em nossos estudos, temos usado $k=3$. A taxa bayesiana empírica (ASSUNÇÃO et al., 1998; MARSHALL, 1991) é uma técnica estatística para calcular taxas e razões que não é afetada pela flutuação de pequenas populações.

4. FUNCIONAMENTO DO INFOSAS

A Figura 5 mostra o fluxo de dados do InfoSAS. Mensalmente, dados das bases SIA, SIH e CNES, e tam-

bém dados populacionais do IBGE alimentam o servidor de mineração de dados, que é guiado por uma tabela de alvos de mineração e que executa algoritmos para detecção de discrepâncias estatísticas. Esta fase de mineração produz o que nós chamamos de cubo de fatos minerados. Este cubo é explorado pelo usuário do InfoSAS, utilizando uma ferramenta de BI. O usuário seleciona um modelo de relatório e especifica parâmetros para filtros por conjunto de alvos, período de análise e recortes geográficos. Um relatório é produzido de acordo com o modelo escolhido e com os filtros determinados, contendo escores de discrepância calculados pelos algoritmos de mineração. Do relatório, o usuário extrai para análises mais aprofundadas folhas de fatos que lhe chamam a atenção.

5. CONCLUSÃO E TRABALHOS FUTUROS

No momento atual (outubro de 2016), o InfoSAS está instalado e em funcionamento no DATASUS. Um curso a distância para capacitação de gestores quanto ao uso do InfoSAS está em preparação, devendo ser finalizado até novembro de 2016, com a primeira oferta prevista ainda para este ano. O sistema InfoSAS já foi apresentado em diversos eventos nacionais (DRAC SAS, 2015) e, em todos eles, seus resultados foram julgados muito interessantes por especialistas em saúde pública.

Os resultados já obtidos pelo projeto InfoSAS são importantes e representam um passo à frente na moder-

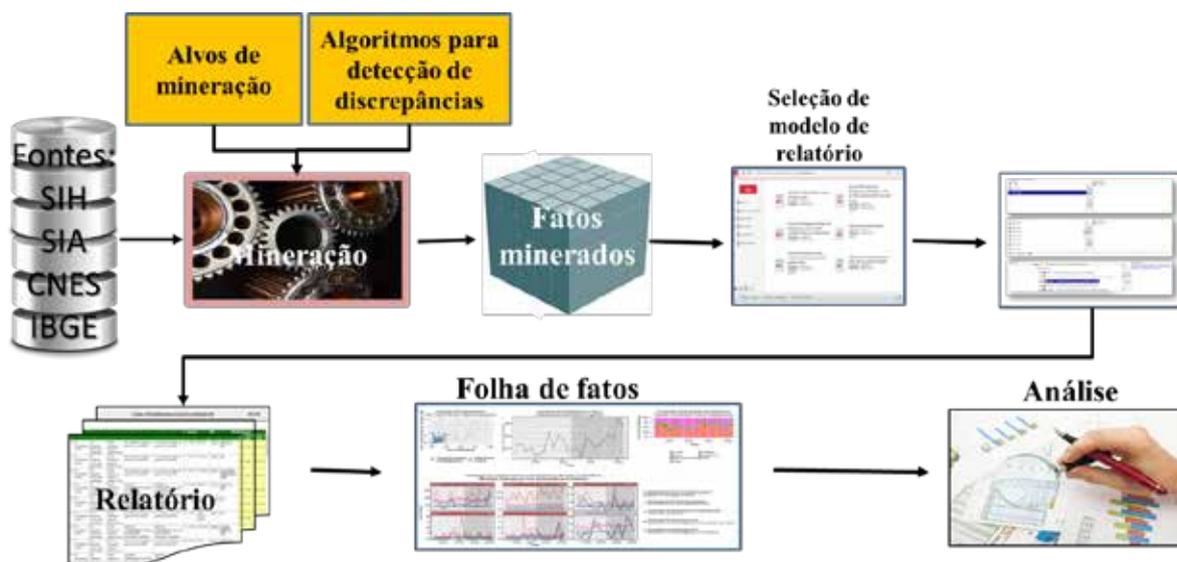
nização dos processos de seleção de itens para auditoria e controle. Além da necessidade de manutenção do sistema, constante atualização, correção e calibragem fina dos algoritmos já utilizados, existem muitas linhas para desenvolvimentos futuros. Vislumbramos uma nova interface com visualização mais flexível e interativa dos resultados; uma caracterização estatística de alvos com identificação de vazios assistenciais; a utilização da população SUS nos cálculos estatísticos; o cálculo de limites inferiores e superiores para a atribuição de produção anômala de um município a seus prestadores; a escolha adaptativa de recortes geográficos com o dimensionamento adequado à frequência de ocorrência de cada alvo; e a utilização de paralelismo no processamento da mineração e da visualização.

Além desses, queremos destacar dois desenvolvimentos que, a nosso ver, são os mais importantes a serem feitos. O primeiro destes seria viabilizado pela realização de auditorias, talvez guiadas por alertas do InfoSAS, classificando atendimentos um a um como fraudulentos ou regulares. Essa rotulação abre a possibilidade de utilização de algoritmos de aprendizado supervisionado, permitindo a emissão de alertas de forma mais precisa e elaborada.

O outro consiste em uma análise consolidada de anomalias em taxas de atendimentos por habitante. O InfoSAS, em sua versão atual utiliza diversos algoritmos que procuram capturar anomalias estatísticas. Temos

Figura 5:

Etapas da análise de dados pelo sistema InfoSAS



também uma estrutura de alvos de mineração onde um alvo pode incluir outros alvos. Com isso, conseguimos muitos achados, mas não temos como classificar estabelecimentos pelo conjunto de alvos nem obter estimativas globais de valor para produção anômala.

Para produzir um novo relatório superando estes problemas, nós decidimos ter como alvos somente conjuntos de procedimentos em uma mesma *forma de organização* da Tabela do SUS, o que faz com que a interseção entre dois alvos quaisquer seja vazia, e utilizar um único algoritmo com uma metodologia estatística para estimativa de excessos em taxas de atendimentos por habitante. Isso torna possível a obtenção de estimativas globais (em todos os alvos) de produção anômala para cada prestador, o que permite a priorização da apuração de fatos pelos gestores do SUS.

A metodologia que pretendemos utilizar, e que já aplicamos a dados de 2013, tem como ponto de partida a constatação de que a distribuição das taxas de atendimento por habitante observadas nos municípios segue uma distribuição log-normal, como ilustrado na Figura 6. Isso abre a possibilidade de definição para qualquer alvo e para qualquer município de um ponto de corte para o que deve ser considerado como normal ou anormal em uma taxa de atendimentos por habitante.

Nós arbitramos que taxas de atendimento por habitante devem ser consideradas anormais quando se

situarem à direita do ponto que divide a área sob a curva em uma parte “normal”, com 99% de probabilidade; e em uma parte “anormal”, com 1% de probabilidade, o que nos parece um critério bastante prudente. Para o exemplo da Figura 7, taxas acima de 3,2 atendimentos por 1.000 habitantes/mês são consideradas anormais. Suponhamos que para o alvo em questão, um município de 100.000 habitantes tenha recebido 400 atendimentos em um dado mês. Pela taxa limite de 3,2 atendimentos/1.000 habitantes, a população desse município deveria ter recebido, no máximo, 320 atendimentos. Nós consideramos, então, que o município teve 80 atendimentos anormais, cujo valor é distribuído entre os prestadores do município, mantendo as proporções do atendimento de cada prestador.

Os resultados da aplicação desta análise à produção registrada no SUS em 2013 são fortes. De um total de R\$19.912.491.904,00, foram considerados anormais atendimentos com valor somado de R\$413.920.365,56, correspondendo a 2,08% do total. O excesso em alvos registrados no SIHSUS foi de R\$350.354.969,25, e no SIA de R\$63.565.396,30.

Outro resultado deste estudo que pode ser utilizado para priorização de auditorias é uma análise da distribuição entre prestadores dos atendimentos considerados anormais. Em todo o Brasil, apenas 5 prestadores concentram quase 9% do valor total das anomalias;

Figura 6:

Densidade da probabilidade de taxas de atendimento por habitante: log-normal teórica vs. dados observados

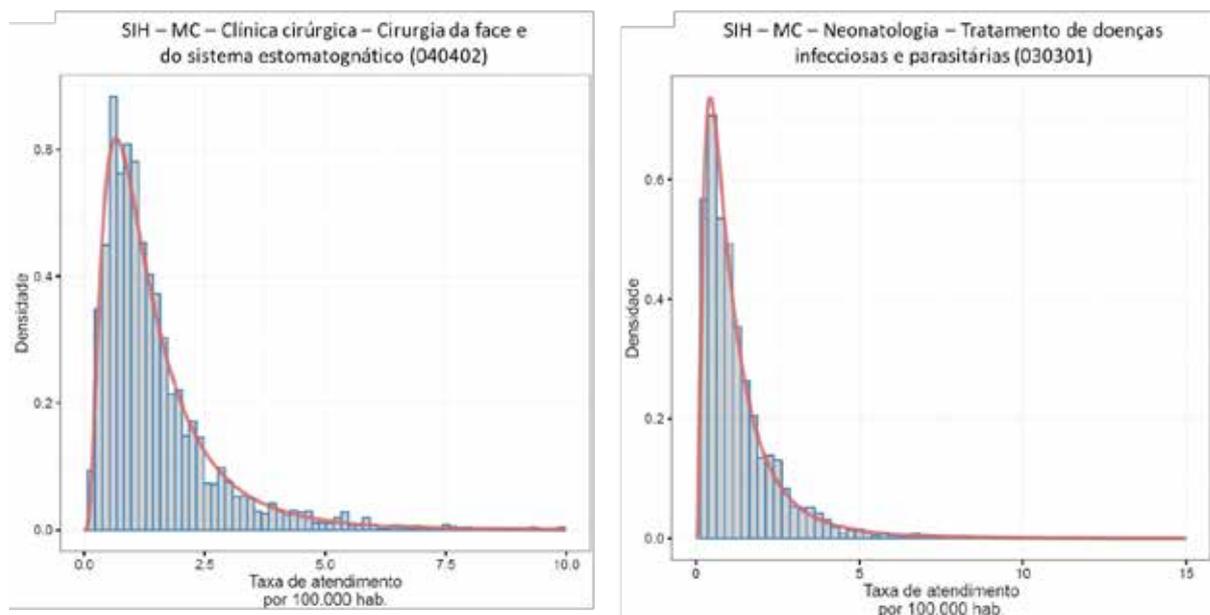
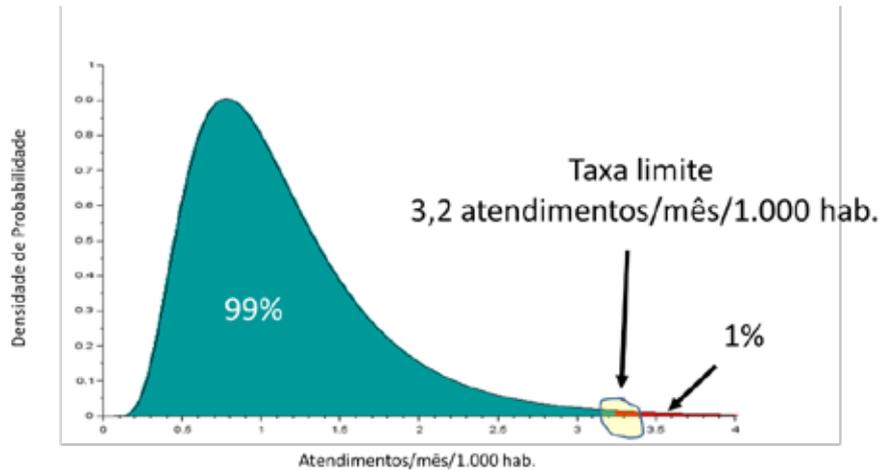


Figura 7:

Definição da taxa-limite de atendimentos por habitante no ponto de corte de 1% da distribuição log-normal



mais de 50% desse total está concentrado em 100 prestadores. Um único prestador teve quase 10 milhões de reais em atendimentos considerados como anômalos em 2013.

6. AGRADECIMENTOS

Em primeiro lugar temos que agradecer ao Departamento de Regulação, Avaliação e Controle da Secretaria de Atenção à Saúde do Ministério da Saúde, que solicitou e financiou todo o projeto.

O InfoSAS foi construído por uma equipe grande e multidisciplinar. Maria Helena Brandão foi a condutora do projeto por parte do DRAC. Ester Dias, Marcelo Campos, Sônia Gesteira e Suzana Rattes, da SMSA-BH, Luciana Moraes, da Funed, e Mônica Castro, da Unimed-BH, foram nossos consultores sanitários. Fabiana Peixoto e Letícia Neto foram as gerentes, e Tomas Schweizer foi o líder técnico. Edré Moreira, Carlos Teixeira, Douglas Azevedo, Larissa Santos, Luiz Fernando Carvalho, Luiz Gustavo Silva, Maurício Nascimento Jr., Milton Ferreira, José Carlos Serufo Jr., Pablo Fonseca, Renan Xavier e Raquel Ferreira, bolsistas de pós-graduação, participaram decisivamente na pesquisa, construção e implementação de diversos algoritmos. Felipe Caetano, Ícaro Braga, Geraldo Franciscani, João Paulo Pesce, João Victor Bárbara, Raphael de Faria e Wicriton Silva cuidaram de desenvolvimento, testes, visualização e bancos de dados. A toda essa equipe, verdadeira responsável pela construção do sistema, vão os nossos mais sinceros agradecimentos.

REFERÊNCIAS

- ASSUNÇÃO, R. M. et al. Maps of epidemiological rates: a Bayesian approach. *Cadernos de Saúde Pública*, Rio de Janeiro, v. 14, n. 4, p. 713–723, out./dez. 1998.
- CARVALHO, L. F. M. et al. A simple and effective method for anomaly detection in healthcare. In: 4TH WORKSHOP ON DATA MINING FOR MEDICINE AND HEALTHCARE, IN CONJUNCTION WITH THE 15TH SIAM INTERNATIONAL CONFERENCE ON DATA MINING, 2015, Vancouver, maio 2015. Disponível em <http://homepages.dcc.ufmg.br/~carlos/papers/sdm/dmmh2015.pdf>. Acesso em 29 nov. 2016.
- DRAC SAS. Ministério da Saúde. Ciclo de Oficinas do DRAC – Controle de Avaliação. Brasília, DF, 1 set. 2015. Disponível em: https://www.youtube.com/watch?v=vlaR_Q7T-U. Acesso em: 4 jul. 2016.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI magazine*, California, v. 17, n. 3, p. 37-54, 1996.
- MARSHALL, R. J. Mapping disease and mortality rates using empirical Bayes estimators. *Applied Statistics*, Auckland, v. 40, n. 2, p. 283–294, 1991.
- THE ECONOMIST. The \$272 billion swindle. Londres, 31 maio 2014. Disponível em <http://www.economist.com/news/usa/21603078-why-thieves-love-americas-health-care-system-272-billion-swindle>. Acesso em: 29 nov. 2016.

Aplicativos de auditoria: uma ferramenta eficaz para asseguuração de compras governamentais¹



Qiao Li

é candidata a doutorado na Rutgers, a Universidade Estadual de Nova Jersey, Estados Unidos. O trabalho de pesquisa desenvolvido por Li inclui auditoria analítica, planejamento e gestão de risco em auditoria, sistemas para suporte em auditoria e aprendizado de máquina (machine learning).



Jun Dai

é professora assistente na Southwestern University of Finance and Economics, na China, e candidata a doutorado na Rutgers, a Universidade Estadual de Nova Jersey, Estados Unidos. O trabalho de pesquisa desenvolvido por Dai inclui auditoria analítica, automação em auditoria, aplicativos para auditoria e blockchain.

RESUMO

Recentemente, governos de muitos países adotaram iniciativas de dados abertos com o objetivo de tornar suas operações mais transparentes para os cidadãos. Com os dados abertos, qualquer um que tenha interesse em monitoramento de gastos do governo pode aplicar tecnologias para realizar análises sobre dados de compras públicas governamentais. Este trabalho propõe 29 aplicativos de auditoria que podem ajudar diversas partes interessadas na análise de dados governamentais abertos. Esses aplicativos podem ajudar a investigar dados de aquisição sob diferentes perspectivas, tais como autenticação das qualificações do contratante, detecção de precificação duvidosa etc. Esse artigo usa dados de contratos de compras do Governo Federal brasileiro para ilustrar a funcionalidade desses aplicativos. No entanto, os aplicativos também podem ser utilizados por outras nações para analisar seus dados governamentais abertos.

Palavras-chave: Aplicativos de Auditoria, Análise de Dados, Compras Públicas Governamentais.

1. INTRODUÇÃO

De acordo com a Organização Mundial do Comércio (OMC), contratos públicos governamentais respondem, em média, por 10% a 15% do Produto Interno Bruto (PIB) de uma economia (OMC 2015a). Os governos são grandes compradores de bens e serviços e isso os tornam vítimas frequentes de fraudes em contratos públicos (Robbins Geller



Rudman & Dowd LLP 2016). Devido ao complicado processo de licitação e contratação, aos vários tipos de contratos, à confidencialidade de informações e descentralização do armazenamento de dados é muito difícil para as partes interessadas organizarem, analisarem e monitorarem os contratos dos governos federais e locais. Nos últimos anos, governos de muitos países, tais como Estados Unidos, Canadá e Brasil, começaram iniciativas de dados abertos, com o objetivo de tornar as informações sobre as operações do governo imediatamente disponíveis, úteis e transparentes para os seus cidadãos (Data.gov 2016).

Embora fontes de dados abertos disponibilizem informações governamentais para o público, poucos estudos ou metodologias avaliam como partes interessadas podem coletar e analisar dados. Dessa forma, este trabalho visa desenvolver e propor aplicativos de auditoria que podem servir como instrumentos eficazes para o monitoramento das despesas públicas e identificar possíveis anomalias nos contratos. Aplicativos de auditoria são procedimentos formalizados de auditoria realizados por meio de *scripts* de computador, que tem se popularizado recentemente. Esse trabalho propõe uma estrutura que fornece orientação para a concepção de aplicativos de auditoria eficazes que examinam contratos governamentais. Para ilustrar essa afirmativa, projetamos 29 aplicativos de auditoria que identificam potenciais contratos de alto risco. Oito desses aplicativos foram desenvolvidos para demonstrar a sua utilidade e benefícios ao utilizar dados contratuais do Governo Federal brasileiro.

2. ANTECEDENTES

2.1 DADOS ABERTOS DO GOVERNO E ASSUNTOS RELACIONADOS

No termo dados abertos, a palavra "aberto" indica que os dados estão disponíveis gratuitamente para todos utilizarem e redistribuírem (Auer et al., 2007). Tais dados são reunidos e mantidos pelos governos e podem ser acessados pelos cidadãos através de sítios online. Muitos países têm criado bancos de dados abertos, para tornar os dados disponíveis ao público. Por exemplo, o Brasil publicou um sistema de informação de aquisição federal chamado "SIASG". O Brasil também fornece Interfaces de Programação de Aplicativos (APIs) para que os cidadãos possam baixar dados sobre contratos federais associados a fornecedores, mercadorias etc. Apesar de ter sido divulgada uma grande variedade de dados do governo, esse estudo concentra-se apenas em dados abertos que se relacionam com contratos públicos. Alguns exemplos de países que criaram sites e bancos de dados que possuem dados de compras governamentais bem-sucedidos incluem os EUA, China, Austrália, Canadá, Brasil e Reino Unido. Por exemplo, o governo britânico publicou seus dados de contratos em 2010. Utilizando essa informação, um oficial britânico encontrou registros duplicados de compras em vários departamentos do governo que custaram mais de 4 milhões de libras (6 milhões de dólares). Em outro exemplo, oficiais do governo de São Francisco tomaram públicos os dados de transporte em 2012. Eles estimam que o número

de atendimentos de consultas telefônicas abaixou e com isso economizaram mais 1 milhão de dólares (*The Economist* 2015). No entanto, existem alguns problemas com dados abertos que ainda impedem seu sucesso. Em primeiro lugar, a qualidade dos dados não é boa o suficiente. Embora os governos de muitos países sejam obrigados a divulgar dados relacionados, o nível de divulgação varia. Todos os campos de dados úteis são fornecidos? Os detalhes dos dados são adequados? Existe alguma informação faltando? Estas questões podem alterar a eficácia da análise de dados abertos. Em segundo lugar, alguns dados abertos não são preparados em um formato legível por máquina, como o formato PDF digitalizado (que são na verdade fotos). Isto torna a coleta e análise de dados mais difícil. Em terceiro lugar, a quantidade de dados é enorme. Isto significa que procurar portais de dados abertos e extrair informações relevantes e úteis são muitas vezes uma tarefa árdua. Outro problema é que poucos indivíduos possuem habilidades para analisar e interpretar dados e, em seguida, aproveitar essas interpretações ou conclusões. Por fim, dados não publicados podem ser muito valiosos, mas são dificilmente acessíveis devido à questão da privacidade de dados (*The Economist* 2015).

2.2 USUÁRIOS INTERESSADOS EM DADOS ABERTOS

As partes interessadas em explorar dados abertos, identificar anomalias, descobrir irregularidades e detectar fraudes incluem, mas estão não limitados a: cidadãos, imprensa, empresas concorrentes e concorrentes políticos. Há pelo menos dois desafios que devem ser abordados antes

que esses grupos comecem a trabalhar com dados abertos. Primeiro, além de dados abertos do governo, eles podem também precisar coletar dados úteis provenientes de outras fontes. Essas fontes podem incluir mídias sociais, notícias, relatórios governamentais e relatórios de analistas. É difícil para as partes interessadas coletar essas informações e integrá-las com dados abertos governamentais. Um desafio ainda maior existe em analisar a grande quantidade de dados disponíveis. Poucas ferramentas analíticas foram especificamente concebidas para a variedade de usuários que querem investigar dados governamentais (O'Leary 2015). Embora existam muitos programas de software de análise de dados gerais sobre o mercado, a complexidade inerente dessas ferramentas analíticas pode impedir a compreensão e a utilização pelos usuários do software. Este problema é evidente, especialmente, para usuários com conhecimento limitado de auditoria e análise de dados. Portanto, desenvolver ferramentas de análise de dados eficientes e eficazes é uma questão crítica.

2.3 APLICATIVOS NA ÁREA DE AUDITORIA

Aplicativos de auditoria são rotinas analíticas formalizadas, executadas por uma ferramenta informatizada (Dai, Krahel e Vasarhelyi 2014). Cada aplicativo de auditoria geralmente executa um teste único de auditoria analítica, exigindo geralmente poucas interações do usuário. Os usuários só precisam carregar dados em aplicativos de auditoria para obter resultados sem muitas operações complicadas. Auditores podem até mesmo criar aplicativos de auditoria personalizadas que realizam tarefas especiais de auditoria. A popularidade



de aplicativos de auditoria tem crescido recentemente. Isto é, em parte, devido a desenvolvedores de software e provedores de serviços de auditoria, que têm feito esforços para criar aplicativos de auditoria². Alguns aplicativos no mercado, no entanto, são desenvolvidos especialmente para os cidadãos que pretendem analisar dados abertos do governo.

Aplicativos de auditoria podem estar entre as ferramentas favoritas para analisar dados abertos por várias razões. Uma delas é que aplicativos, em geral, podem ser operados facilmente com treinamento mínimo. Isso permite que várias partes interessadas, de várias áreas de escolaridade, possam analisar os dados com facilidade. O baixo custo dos aplicativos é outra razão pela qual elas são uma opção atraente. A maioria dos usuários são cidadãos ou empresas concorrentes e não auditores profissionais. Como resultado, é difícil para estes comprarem um software profissional de auditoria caro. Aplicativos de auditoria são substituídos com bom custo benefício que permitem aos usuários executar uma grande variedade de testes de auditoria com base em análise. Personalização é outra vantagem importante dos aplicativos, porque permite entender aplicativos para cumprir tarefas específicas do usuário. Os usuários podem criar aplicativos personalizados usando Kits de Desenvolvimento de Software profissional.

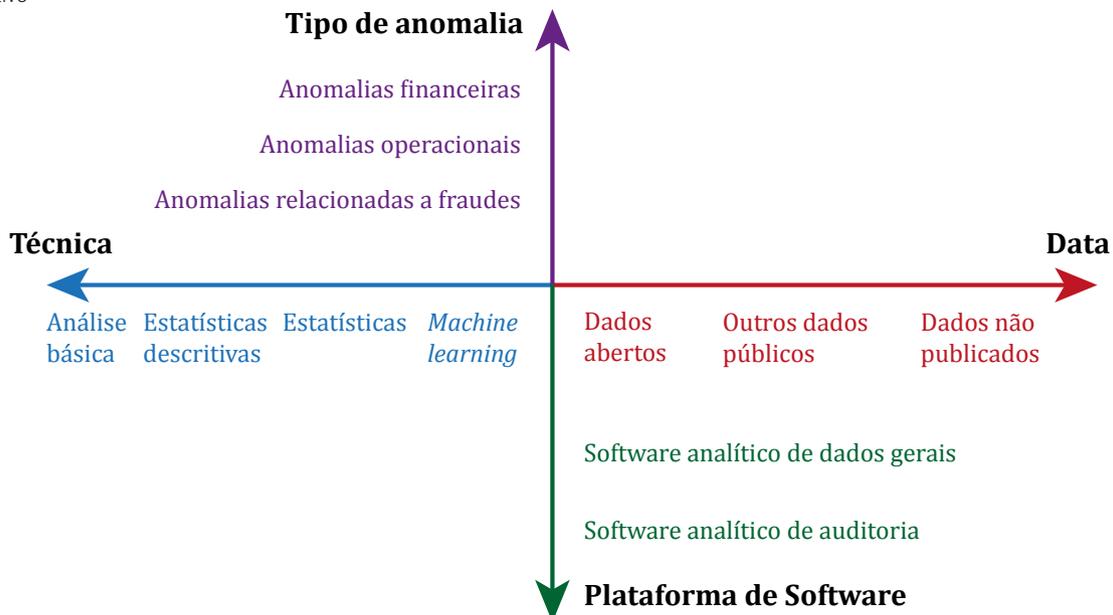
3. UMA ESTRUTURA PARA PROJETAR APLICATIVOS

É necessária orientação para projetar aplicativos eficientes e eficazes. Propomos uma estrutura que fornece orientação para design de aplicativos para auditorias de compras governamentais. A estrutura (mostrada na Figura 1) contém quatro dimensões: tipo de anomalia, tipo de dados, plataforma de software e técnica.

A primeira dimensão é o tipo de anomalia. Anomalias podem ser classificadas em três tipos: financeiro, operacional e relacionadas à fraude. Anomalias financeiras referem-se às anomalias de dados que podem afetar demonstrações ou relatórios financeiros governamentais. Estes podem incluir a falta de dados ou valores incorretos / desatualizados. Anomalias operacionais focam na ineficiência das operações e atividades governamentais. Estas incluem a compra de produtos de luxo para escritório (como uma cadeira de massagem extremamente cara), bens obviamente desnecessários (como joias) ou bens necessários com preços significativamente maiores do que a taxa de mercado. Anomalias relacionadas com fraude apresentam padrões de dados incomuns que podem ser resultados de fraudes. Por exemplo, se todos os licitantes oferecem um preço uniforme e se recusarem a negociar durante o processo de licitação do governo.

A segunda dimensão é a plataforma de software. Esta será usada para desenvolver e operar aplicativos de auditoria. Geralmente, existem dois tipos de plataformas: Software de análise de dados de auditoria (ADAS)³ e Software de análise de dados generalizados (GDAS)⁴. O

Figura 1:
Estrutura para projetar um aplicativo





ADAS contém funções de auditoria pré-programadas, o que torna a criação de aplicativos com base em funções mais fácil. O GDAS geralmente tem a capacidade de manipulação de dados de alto volume, como vários dados abertos. O GDAS também contém uma grande variedade de estatísticas e modelos de aprendizagem de máquina, tornando-o útil para desenvolvimento de aplicativos. A desvantagem desse tipo de software é que ele requer formação técnica ou treinamento dos usuários para compreender os modelos sofisticados. Como cada plataforma de software tem suas características especiais, os desenvolvedores podem criar várias versões de aplicativos que permite aos usuários executá-los em diferentes softwares.

A terceira dimensão está relacionada ao tipo de dados. Recentemente, muitos estudos têm discutido a utilização de "big data" para fins de garantia (Vasarhelyi, Kogan e Tuttle 2015; Cao, Chychyla e Stewart 2015; Yoon, Hoogduin e Zhang 2015; Alles 2015). À medida que mais países e organizações começam a abrir seus dados, tais dados abertos se moverão rapidamente para "big data" (O'Leary 2015). Além disso, outros dados públicos, tais como artigos de notícias, mídia social e dados coletados de máquinas ou vários sensores podem ser coletados. Todos esses dados podem ser combinados e usados para identificar anomalias ou respaldar investigações (Vasarhelyi et al. 2015). Ademais, certas informações inéditas, tais como políticas de operações internas de certas unidades governamentais, também podem fornecer esclarecimento quanto a identificação de atividades anormais governamentais. Quando integrado com dados públicos, essas informações inéditas podem

ser usadas para localizar processos arriscados de operações governamentais e identificar potenciais contratos ou transações fraudulentas. Os aplicativos devem ser desenvolvidos para facilitar as análises de uma grande variedade de dados.

A última dimensão está relacionada à técnica. Técnicas analíticas básicas podem ser usadas para identificar padrões incomuns. Tais técnicas incluem sumarização, consulta e correspondência de dados. Estatísticas descritivas podem fornecer uma visão global de dados. Estatísticas (por exemplo, regressão e séries temporais) e aprendizado de máquina (por exemplo, agrupamento e classificação) são técnicas avançadas que podem descobrir efetivamente padrões escondidos dentro de dados complexos. Os desenvolvedores devem tirar proveito dessas técnicas para criar aplicativos eficazes e eficientes

4. PROJETANDO APLICATIVOS PARA AUDITORIA DE DESPESAS PÚBLICAS

Propomos 29 aplicativos que podem ajudar a identificar potenciais anomalias nas despesas do governo. Essas anomalias incluem contratos suspeitos ou a utilização de fornecedores não qualificados. A Tabela 1 mostra os aplicativos propostos.

Os três primeiros aplicativos fornecem uma verificação preliminar sobre a confiabilidade e integridade dos dados de contratos governamentais. Eles também demonstram quaisquer padrões de dados especiais. Padrões de dados anormais, tais como valores ausentes e flutuação anormal, podem indicar riscos no processo de contratação. Esses aplicativos fornecem garantia prelimi-

Tabela 1:

Lista de aplicativos propostos

Nº	Objetivo do Aplicativo	Indicador de Anomalia	Dados	Técnicas
Aplicativos para dados incompletos / Verificação de integridade e Constatação de padrões de dados				
1	Verifica valores de contratos anormais	Números incomuns como 0, 0.01, 0.05	Dados de contrato	Consulta
2	Verifica a integralidade e integridade de dados	Ausência de fornecedores/licitação/datas/etc.	Dados de contrato	Consulta
3	Encontra padrões de dados especiais (usando painel de instrumentos)	Flutuação anormal de dados, números extremamente grandes ou pequenos, etc.	Dados de contrato	Estatística descritiva
Aplicativos para Fornecedores Suspeitos				
4	Verifica a informação do fornecedor	Fornecedores não existem no arquivo principal	Dados de contrato, dados de fornecedores	Comparação de dados
5	Verifica a qualificação do fornecedor	Fornecedores, ou sua empresa matriz/subsidiárias, estão na lista de empresas suspensas	Dados de contrato, Empresas suspensas	Comparação de dados
6	Verifica relacionamentos	Membros da família do fornecedor trabalham no governo	Informação de membros da família de fornecedores	Consulta
7	Verifica licitação dispensável	Fornecedores têm uma alta proporção de contratos que não passaram por processos normais de licitação	Dados de contrato	Consulta
8	Verifica ganhadores anormais de licitações	Algumas empresas sempre ganham ou todos os fornecedores ganham um número igual de licitações	Dados de contrato, dados de licitações	Sumarização
9	Verifica a distribuição regional de fornecedores	Fornecedores em uma área específica ganham a maioria dos contratos	Informação geográfica de fornecedores	Visualização
10	Verifica a combinação anormal do licitante	Os mesmos licitantes sempre ou nunca concorrem entre si	Dados de licitações	Sumarização
Aplicativos para Preços e Valor Inicial de Contrato Anormais				
11	Verifica valores anormais de contratos	Valores iniciais de contratos não estão em conformidade com a lei de Benford	Valores de contrato	Lei de Benford
12	Compara preços de contratos	Fornecedores propõem preços elevados em licitações governamentais do que o preço de mercado	Dados de preços	Comparação de dados, consulta
13	Identifica compras divididas	Contratos com mesmo fornecedor, datas e bens e/ou serviços	Dados de contratos	Deteção duplicada
14	Prevê e identifica preços vencedores anormais	Preços vencedores são maiores que preços previstos	Dados de contratos	Regressão
15	Identifica diferenças de preços anormais	Grande diferença entre o preço de oferta do vencedor e dos outros preços propostos	Informação de preços	Consulta
16	Verifica o desvio-padrão do valor de licitação	Todos os preços dos fornecedores parecem uniformes. Fornecedores recusam negociar preços.	Preços de licitação	Desvio-padrão
17	Verifica alterações de contrato	Os valores iniciais dos contratos mudaram muito	Dados de contrato	Comparação de dados
Aplicativos para Procedimento e Modalidade Anormais de Licitação				
18	Verificar a vigência do licitante	Muitos poucos fornecedores vigentes	Processo de licitação	Consulta
19	Detecta desistência de licitantes	Licitantes qualificados retiram propostas válidas inexplicavelmente	Processo de licitação, dados de fornecedores	Consulta, comparação de dados
20	Detecta propostas duplicadas	Empresas apresentam licitações com itens individuais idênticos e montantes fixos	Processo de licitação	Deteção duplicada
21	Detecta licitantes duplicados	Licitante com endereços, fac-símile, números de telefones ou outros dados demográficos duplicados	Informação demográfica de licitantes	Deteção duplicada
22	Verifica flutuação anormal do preço de oferta	Competidores anunciam o aumento do preço ao mesmo tempo, e no mesmo valor	Preços de licitação durante o processo licitatório	Deteção duplicada
Produtos Anormais /Implementação de Serviços				
23	Verifica produtos de luxo	Governo compra produtos de luxo	Dados de contrato	Mineração de texto
24	Verifica itens obviamente desnecessários	Governo compra muitos itens desnecessários como vale-presentes	Dados de contrato	Mineração de texto
25	Verifica custos excessivos	Custos excedem em muito as estimativas	Dados de contrato	Consulta
26	Verifica horas trabalhadas	Funcionários cobram mais horas do que o horário normal de trabalho	Faturas	Consulta
27	Verifica faturamentos duplicados	Faturas duplicadas para o mesmo produto ou serviço	Faturamento	Deteção duplicada
28	Verifica local de entrega anormal	O local de entrega não é um escritório, fábrica ou local de obra	Endereço de entrega	Visualização
29	Verifica informações geográficas de fatura	Empregados são faturados em vários locais de obras distantes, no mesmo dia	Faturas	Visualização

nar de nível de dados identificando contratos com valores ausentes ou anormais. Os aplicativos de 4 a 10, auxiliam na identificação de fornecedores suspeitos. Visto que fornecedores podem estar envolvidos em muitos tipos de fraudes, é necessário desenvolver aplicativos para monitorar o comportamento dos fornecedores durante os processos de licitação e contratação. Esses aplicativos são, portanto, projetadas para identificar fraudes tais como suborno, esquemas de comissões e manipulação de preços de licitação. Os aplicativos de 11 a 17 foram projetadas para identificar padrões de preços incomuns. Tendo em vista que contratos arriscados são normalmente associados com preços anormais, esses aplicativos podem chamar a atenção para contratos de alto risco. Os aplicativos de 18 a 22 podem ser utilizadas para analisar e monitorar o procedimento complexo de licitação. Estes aplicativos podem identificar rapidamente comportamentos suspeitos de licitação, que pode se caracterizar quando poucos licitantes válidos participam de uma licitação ou quando uma há uma retirada inexplicável de uma licitação do processo. Os aplicativos de 23 a 29 ajudam a identificar as compras de produtos

anormais ou desnecessários, bem como taxas de serviços não prestados. Isso inclui gastos excessivos em produtos para escritório ou utilização indevida de fundos públicos.

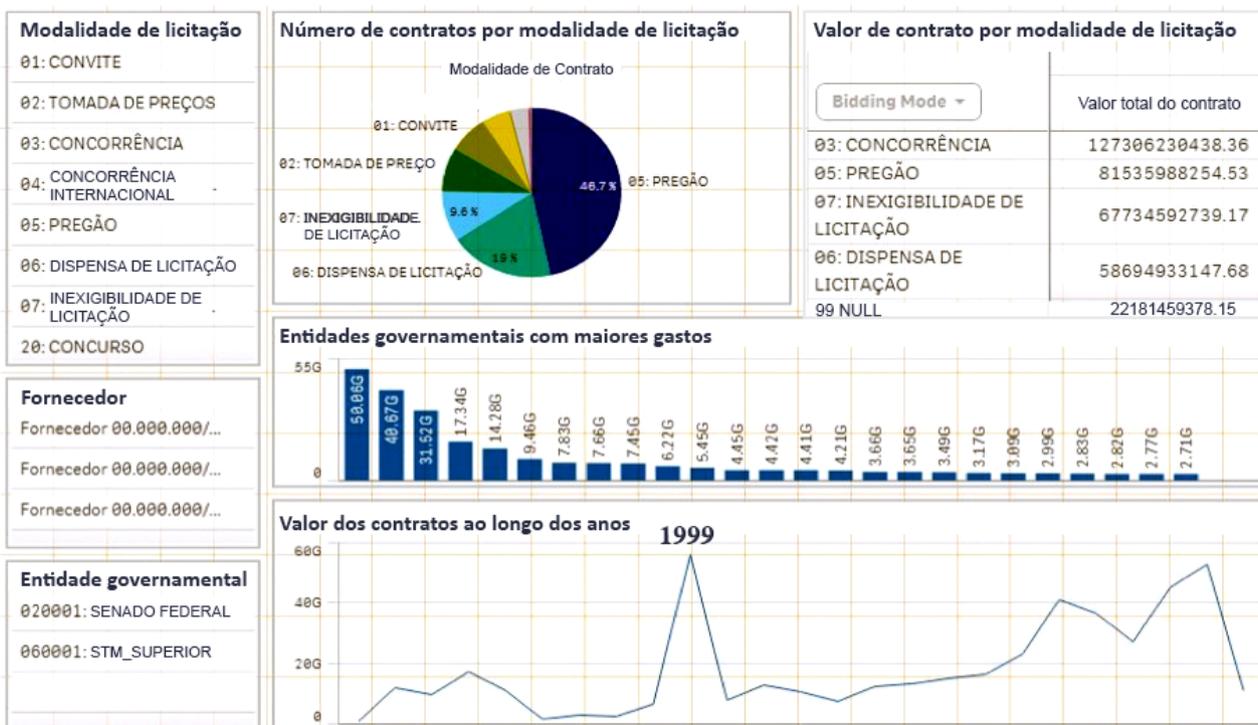
5. ILUSTRAÇÃO DE APLICATIVOS UTILIZANDO CONTRATOS FEDERAIS DE COMPRAS BRASILEIRAS

Para demonstrar a utilidade dos aplicativos de auditoria na auditoria de contratos de governo, coletamos dados de contratos e fornecedores do Sistema Integrado de Administração de Serviços Gerais (SIASG) e do Cadastro Nacional de Empresas Inidôneas e Suspensas (CEIS), de 1989 a 2014. O arquivo de contrato contém, principalmente, informações sobre entidades governamentais, fornecedores de bens e serviços, métodos de licitações, datas de início / término e valores iniciais de contratos. O arquivo de fornecedores registra previamente as empresas para participar de processos de licitação. Os dados nesse arquivo incluem o CNPJ das empresas e suas informações demográficas. O sistema CEIS grava o CNPJ das empresas que estão proibidas de vender produtos ou serviços para o

Figura 2:

Aplicativos de Auditoria para Análise Descritivas (Adaptado de Dai; Li, 2016)

Aplicativo de Análise Descritiva





governo federal brasileiro e inclui a data de início e término de suas sanções. Com base nos dados, desenvolvemos oito aplicativos para executar a análise descritiva, verificação de dados incompletos e de integridade e identificação de anomalias.

5.1 PAINEL DE ANÁLISE DESCRITIVA

Um aplicativo de auditoria⁵ é desenvolvido para realizar análises descritivas dos dados de contrato de compras do governo. A Figura 2 utiliza um painel para mostrar os resultados. O painel esquerdo lista os campos importantes no conjunto de dados. O painel direito usa diversos gráficos para mostrar a análise descritiva dos valores de contratos, oferecendo tipos de licitações e entidades governamentais. O gráfico mostra que 9,6% e 19% das compras foram associadas com a modalidade de licitação 06 (dispensa de licitação) ou 07 (inexigibilidade de licitação), respectivamente. A tabela resume o valor inicial de contratos em cada modalidade de licitação. De acordo com a tabela, contratos com modalidade de licitação 07 e 06 compõem o terceiro e quarto lugar, respectivamente, em termos de valores de contrato inicial. As informações do gráfico e da tabela indicam que os contratos com modalidade de licitação 07 ou 06 poderiam apresentar um alto risco de fraude. Isto decorre porque esquemas de comissões e suborno estão suscetíveis de ocorrer quando bens ou serviços são comprados sem processos licitatórios regulares. O histograma mostra os gastos de cada entidade de governo entre os anos de 1989 e 2014. As três entidades mais importantes do governo são atípicas pois gastaram muito mais do que o restante. O gráfico de

linhas mostra as alterações em valores totais de contrato ao longo do tempo. Verifica-se um pico em 1999, ano em que a primeira crise financeira atingiu o Brasil⁶. Isto indica uma preocupação de risco potencial nas despesas públicas.

5.2 DADOS INCOMPLETOS E VERIFICAÇÃO DE INTEGRIDADE DE DADOS

Os campos importantes nos dados de contrato são os fornecedores, a modalidade de licitação e as datas de início e término dos contratos. Portanto, a integridade desses dados críticos deve ser verificada antes de realizar análises avançadas. Três aplicativos⁷ foram criados para verificar a integridade de cada um dos três campos e denunciar contratos que tenham valores faltando. Os resultados mostram que 35.516 contratos (de 470.683) não têm informações de fornecedores, 16.167 contratos não informam a modalidade de licitação e as datas de início e término não são mostradas em 1.000 contratos. A ausência de valores nos campos críticos pode ser resultado de erros simples na alimentação de dados ou de atividades fraudulentas, como suborno ou esquemas de comissão. Mais testes de auditoria devem ser realizados para identificar e verificar o motivo pelo qual tais valores estão ausentes.

5.3 IDENTIFICAÇÃO DE ANOMALIAS

Contratos de alto risco geralmente estão associados com padrões de dados especiais que raramente ocorrem em situações normais. Por exemplo, um fornecedor pode ganhar



uma licitação ao oferecer um preço extremamente baixo no início, só para aumentar esse preço mais tarde. Outro exemplo seria uma empresa continuar vendendo bens e serviços após ter sido penalizada por um descumprimento de uma obrigação. Padrões anormais em contratos geralmente indicam comportamento fraudulento por parte dos fornecedores, entidades governamentais ou ambos. Tais fraudes podem causar enormes prejuízos aos governos. Esses contratos devem ser identificados e sinalizados para que uma investigação complementar seja conduzida. Nesse trabalho, quatro aplicativos foram desenvolvidos para identificar contratos que possuam valores iniciais ou fornecedores anormais.

Um aplicativo de auditoria⁸ foi desenvolvida para identificar contratos com valores iniciais extremamente baixos (menos que R\$ 0,1). Os resultados demonstram um total de 9.334 contratos que possuem valores iniciais inferiores a R\$ 0,1. Dentre esses contratos, 8.678 têm R\$ 0,00 como valor inicial, 625 possuem valores superiores a R\$ 0,00, mas menores que R\$ 0,05 e o valor inicial dos 31 contratos restantes estão na faixa de R\$ 0,05 a R\$ 0,10. Todos esses contratos devem ser sinalizados. Novos testes também devem ser realizados para analisar se os valores iniciais extremamente baixos são razoáveis.

Outro aplicativo⁹ foi desenvolvido para executar a análise da lei de Benford nos valores iniciais do contrato. A lei de Benford tem sido amplamente utilizada na contabilidade para detecção de fraudes (Nigrini 1999; Nigrini e Miller 2009). As frequências do primeiro ou segundo dígito dos valores iniciais dos contratos não devem exceder às sugeridas pela lei de Benford. Se este não for o caso, pode indicar

uma possível fraude envolvendo os contratos. Os resultados do aplicativo sugerem que há mais contratos que têm valores iniciais começando com "60", "79" e "80" do que o esperado. Isso indica possíveis riscos nesses contratos. A política interna do governo brasileiro permite procedimentos de licitação simplificados se o valor do contrato for inferior a R\$ 80.000. Compras diretas sem licitação são permitidas caso o valor não seja maior que R\$ 8.000. Os resultados da análise da lei de Benford indicaram que empresas e agências governamentais podem ter conspirado para reduzir o valor inicial do contrato para adaptar-se a esses limites que simplificam os processos de compras.

Analisar fornecedores é um aspecto importante na auditoria de contratos governamentais: auditores devem prestar mais atenção em compras feitas por empresas que possuem registro no CEIS. Isso ocorre porque o uso de tais fornecedores de alto risco pode resultar na quebra de contrato ou no fornecimento de produtos ou serviços de baixa qualidade. Contratos assinados com subsidiárias ou empresas matrizes das empresas suspensas também precisam ser analisadas cuidadosamente. Essas empresas podem subcontratar o serviço das empresas sob sanção. Um aplicativo de auditoria¹⁰ foi desenvolvido para identificar os contratos cujos fornecedores ou subsidiárias/matrizes estão registrados no CEIS. Os resultados mostram que um total de 25.100 contratos foram assinados com empresas ou suas subsidiárias/matrizes que estão, ou fizeram parte, da lista do CEIS. Por exemplo, o fornecedor "33.000.118" e as suas subsidiárias assinaram 1.717 contratos com entidades do governo brasileiro de 1989 a 2014. Isto ocorreu mesmo com o fornecedor tendo sido temporariamente suspenso por uma agência governamental específica de 15 de dezembro de 2014 a 14 de dezembro de 2016. Pode ocorrer que este fornecedor preste serviços a outras agências do governo; no entanto, os contratos assinados pela empresa e suas subsidiárias devem ser cuidadosamente analisados.

O sistema SIASG possui um arquivo de fornecedores independentes que registra empresas pré cadastradas em processos de licitação. Apesar desse registro não ser obrigatório, empresas que estão no arquivo de fornecedores tem um baixo risco de fraude em comparação às empresas não registradas. Isso acontece porque as empresas registradas passam por um processo de verificação durante o processo centralizado de registro. Dessa forma, há garantia de que irão cumprir os requisitos legais do processo de licitação governamental. Em contrapartida, contratos assinados com empresas que não foram registradas estão expostos a um grande risco de fraude. Ainda existe a possibilidade de empresas, que não satisfizeram todos os requisitos legais, ganharem contratos. Isso ocorre quando há conluio entre

as empresas e as agências governamentais. Um aplicativo foi desenvolvido¹¹ para ajudar a identificar contratos com fornecedores que não foram pré cadastrados no sistema. Os resultados mostram que 40.942 contratos foram assinados com fornecedores que não foram registrados no sistema. Mais testes de auditoria precisam ser executados, como, por exemplo, para analisar se as empresas cobram mais do que o preço de mercado e se há uma relação pessoal entre os gestores das empresas e o órgão do governo etc.

6. CONCLUSÃO

Visto que contratos públicos compõem entre 10% e 15% do PIB do país (OCDE 2015), auditorias de compras e a identificação de possíveis anomalias ou fraudes são questões importantes. Esse artigo discutiu o uso e os benefícios dos aplicativos para as auditorias de compras governamentais. Vinte e nove aplicativos específicos foram propostos para facilitar a auditoria dos gastos públicos. Oito aplicativos foram desenvolvidos para demonstrar a sua utilidade em auditorias de contratos. Esse artigo pode fornecer ideias sobre como criar aplicativos eficazes e como usar aplicativos em auditoria de contratos governamentais.

NOTAS

- 1 Este trabalho é baseado no artigo publicado no *Journal of Emerging Technologies in Accounting* (Dai and Li 2016).
- 2 Por exemplo, Caseware tem mais de 50 aplicativos de auditoria em um mercado on-line (Dai et al. 2014). QlikSense permite que usuários desenvolvam aplicativos de auditoria, criando um painel de controle personalizado. Outras empresas, como Forestpin TeamMate Analytics, também desenvolveram alguns produtos derivados de aplicativos.
- 3 Exemplos de ADAS incluem ACL e CaseWare IDEA.
- 4 Exemplos de GDAS incluem R, Weka, SPSS e SAS.
- 5 Esse aplicativo foi desenvolvido através de Qlik Sense
- 6 http://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/16274/Panel%2020-%20-%20Gabriel%20Palma%20_0.pdf?sequence=1.
- 7 Esse aplicativo foi desenvolvido usando Caseware IDEA.
- 8 Esse aplicativo foi desenvolvido usando Caseware IDEA.
- 9 Esse aplicativo foi desenvolvido usando Caseware IDEA.

10 Esse aplicativo foi criado pela SAS Enterprise.

11 Esse aplicativo foi criado pela SAS Enterprise.

REFERÊNCIAS

- Alles, M. G. 2015. Drivers of the Use and Facilitators and Obstacles of the Evolution of Big Data by the Audit Profession. *Accounting Horizons*, 29(2), pp.439-449.
- Auer, S. R., C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. 2007. DBpedia: A Nucleus for a Web of Open Data. *The Semantic Web Lecture Notes in Computer Science* 4825. p. 722. doi:10.1007/978-3-540-76298-0_52. ISBN 978-3-540-76297-3.
- Cao, M., R. Chychyla, and T. Stewart. 2015. Big Data analytics in financial statement audits. *Accounting Horizons*, 29(2), pp.423-429.
- Dai, J., J. P. Krahel, and M. A. Vasarhelyi. 2014. Which Audit App(s) Should Auditors Use? An Exploratory Study of Using Recommender Systems for Audit App Selection. Working Paper.
- Dai, J., and Q. Li. 2016. Designing Audit Apps for Armchair Auditors to Analyze Government Procurement Contracts. *Journal of Emerging Technologies in Accounting*. Forth coming.
- Data.gov. 2016. <https://www.data.gov/open-gov/>.
- Economist. 2015. Open government data: Out of the box. Available at: <http://www.economist.com/node/21678833/print>.
- Nigrini, M. J. 1999. I've Got Your Number: How a mathematical phenomenon can help CPAs uncover fraud and other irregularities. *Journal of Accountancy*.
- Nigrini, M., and S. J. Miller. 2009. Data diagnostics using second order tests of Benford's Law. *Auditing: A Journal of Practice and Theory* 28 (2): 305-324.
- OECD. 2015. "Size of public procurement", in *Government at a Glance 2015*, OECD Publishing, Paris. Available at: http://dx.doi.org/10.1787/gov_glance-2015-42-en.
- O'Leary, D. E. 2015. Armchair Auditors: Crowdsourcing Analysis of Government Expenditures. *Journal of Emerging Technologies in Accounting*.
- Robbins Geller Rudman & Dowd LLP. 2016. Government Procurement Fraud. <http://www.whistleblower-lawfirm.com/types-of-fraud-Government-Procurement-Fraud.html>.
- Vasarhelyi, M. A., A. Kogan, and B. M. Tuttle. 2015. Big data in accounting: An overview. *Accounting Horizons*, 29(2), pp.381-396.
- WTO. 2015a. WTO and Government Procurement. Available at: https://www.wto.org/english/tratop_e/gproc_e/gproc_e.htm.
- Yoon, K., L. Hoogduin, and L. Zhang. 2015. Big Data as complementary audit evidence. *Accounting Horizons*, 29(2), pp.431-438.

Modelagem de dados geográficos para definição de corredores alternativos no rodoanel da região metropolitana de Belo Horizonte: cenários comparativos



José Irley Ferreira Júnior
é geógrafo com especializações e mestrado nas áreas de Geociências e Meio Ambiente, além de consultor autônomo em geotecnologias e transportes.



Leise Kelli de Oliveira
é matemática, professora associada da Universidade Federal de Minas Gerais e co-autora do livro "Logística Urbana: Fundamentos e Aplicações".



Rodrigo Affonso de Albuquerque Nóbrega
é engenheiro cartógrafo e doutor em Engenharia de Transportes, além de professor adjunto da Universidade Federal de Minas Gerais.



RESUMO

Dentre as iniciativas propostas pela Organização da Este trabalho tem como objetivo apresentar a metodologia utilizada no cálculo de diferentes alternativas de corredores para viabilidade do segmento sul do Rodanel da Região Metropolitana de Belo Horizonte (RMBH) computada a partir de cenários preditivos desenvolvidos por meio de geoprocessamento e análise multicritério. A área de estudo engloba preocupações extremas de naturezas física, biológica, econômica, social e logística. Mesmo contando com plena presença de unidades de conservação, áreas de recarga de água e terrenos com topografia muito acidentada, a região tem sofrido fortes pressões antrópicas, principalmente com o crescimento da mancha urbana, da instalação de indústrias e entrepostos de carga, além da atividade minerária. Do ponto de vista logístico, a região é estratégica por concentrar as conexões entre as rodovias BR-040 e BR-381, que interligam a RMBH ao Rio de Janeiro e à São Paulo, respectivamente. Embora com edital e termo de referência lançados em 2011, o estudo de viabilidade técnica, econômica e ambiental, e o projeto executivo do segmento sul do rodanel não foram concluídos ou apresentados. Nesse sentido, e com o intuito de promover elementos para análise, controle e discussão, a proposta deste trabalho foi produzir material capaz de qualificar e quantificar diferentes alternativas de corredores para o desenvolvimento do traçado na infraestrutura de transportes. O modelo desenvolvido recorreu às in-

formações presentes no edital e no termo de referência do projeto, e os dados utilizados foram todos oficiais e de domínio público. A análise multicritério foi implantada em ambiente de sistema de informações geográficas utilizando a técnica AHP em níveis hierárquicos de decisão. Foram produzidos quatro cenários que refletem interesses distintos e concorrentes: biofísico, restrições ambientais, socioeconômico e mercadológico/logístico. Em cada cenário foram computadas as superfícies de esforço e os corredores de viabilidade. Os corredores foram comparados quanto a extensão, declividade, área urbana, unidades de conservação e área vegetada. O trabalho mostrou grande potencial de aplicação em controle externo. Os resultados mostraram que os cenários preditivos podem ser utilizados para fomentar análises qualitativas e quantitativas quanto à viabilidade de projetos lineares de infraestrutura, ainda que em fase de publicação do edital.

Palavras-chave: Geoprocessamento. Análise multicritério. Regras de decisão. Transparência. Avaliação de obras públicas.

1. INTRODUÇÃO

Embora com longa evolução histórica, dificuldades existem quanto ao desenvolvimento de projetos de infraestrutura viária no Brasil, sobretudo em fases de planejamento e implementação. De acordo com Nóbrega (2013), há inadequações em grande parte dos projetos

quanto à fase de planejamento, e a falta de transparência acerca dos dados e métodos empregados nas análises causa problemas de natureza técnica e orçamentária nas obras, refletindo o aumento do custo e dos prazos.

Para haver um retorno efetivo da atividade de planejamento em transportes, de maneira transparente tanto na esfera pública quanto privada, é imprescindível que haja um pensamento sistêmico dos projetos. Assim, é vital que os profissionais envolvidos tomem decisões em conjunto e considerem a interação das variáveis presentes em todas as fases do processo. Esse raciocínio é o princípio da análise multicritério, com o suporte da técnica Analytic Hierarchy Process (AHP) (Saaty, 1995). Este modelo tem sido explorado em Sistemas de Informações Geográficas (SIG) que permitem a modelagem espacial das variáveis em um processo de tomada de decisão.

Segundo o DNIT (BRASIL, 2012), a proposta de construção do rodoanel caracteriza o projeto em três segmentos: sul, norte e leste. O objeto deste trabalho é investigar, a partir das ferramentas de análise supracitadas, corredores de viabilidade para o “Anel Sul”, trecho cujo estudo de viabilidade encontra-se em desenvolvimento e sem quaisquer resultados prévios apresentados. Neste contexto, esse artigo apresenta os resultados da modelagem geográfica de dados para definir corredores alternativos que representem simultaneamente áreas de maior viabilidade econômica, técnica e ambiental para a implantação dessa infraestrutura viária. Foram utilizados, além do edital público e do termo de referência (BRASIL, 2012), dados geográficos de domínio público dos quais foram extraídas as variáveis ambientais, mercadológicas, logísticas e socioeconômicas utilizadas no modelo.

2. ANÁLISE MULTICRITÉRIO E MODELAGEM DO SIG EM TRANSPORTES

A demanda por metodologias para modernizar o planejamento do transporte é notória e o geoprocessamento tem sido peça-chave para integrar, de forma coordenada, as inúmeras variáveis espaciais deste processo. A inserção do SIG como suporte ao planejamento de corredores de transportes demanda grande volume de dados. De acordo com Longley et al. (2013), um SIG é caracterizado por um conjunto de construtores para representar objetos, ferramentas e processos em ambiente computadorizado, normalmente operando na forma de modelos de dados geográficos.

Um dos modelos que tem sido discutido e utilizado juntamente com o SIG para apoio à tomada de decisão é a análise multicritério. Esta propicia a integração

estruturada das variáveis geográficas, das opiniões dos atores envolvidos, mesmo que distintas ou divergentes, e das ponderações nos pesos das variáveis nas regras de tomada de decisão, com o objetivo de reproduzir cenários diagnósticos e prognósticos. Para subsidiar a análise multicritério, desenvolvem-se metodologias para aperfeiçoar o fluxo de trabalho, e estas são implantadas na modelagem do SIG. Cita-se nesse contexto a AHP, técnica desenvolvida por Saaty (1995) para permitir que a subjetividade advinda das decisões humanas seja minimizada com a aplicação de regras matemáticas no processo de atribuições de peso às variáveis. De acordo com Sadasivuni et al. (2009), esta técnica é aplicada como um método de comparação de variáveis para análise multicritério e se utiliza da modelagem matemática para determinar prioridades, algo importante no tocante à organização das disparidades de valores, opiniões e interesses dos agentes envolvidos no planejamento de corredores de transporte.

Se por um lado o método AHP utiliza valores pareados como dados de entrada, por outro lado suas informações de saída correspondem a um ranking numérico que elenca, ordena e atribui pesos às preferências. No SIG, seu emprego mais frequente tem sido na composição dos valores atribuídos aos pixels de mapas digitais em formato matricial (NOBREGA et al., 2009).

Embora avanços significativos na contextualização geográfica dos processos de tomada de decisão em transportes tenham sido alcançados nas décadas de 1990 e 2000, a metodologia AHP, acoplada ao SIG, tornou-se alvo de interesse em projetos práticos de corredores de viabilidade há apenas alguns anos. No Brasil, o uso combinado de SIG e AHP em projetos de planejamento de corredores de transportes não está restrito à Academia. Trabalhos recentes aplicados ao planejamento de ferrovias foram desenvolvidos no âmbito federal junto a gestores de transportes e fiscalização (BERBERIAN et al., 2015). Essas iniciativas revelam o interesse dos gestores e técnicos em transporte pela modernização do processo de planejamento. Os resultados têm comprovado o potencial do geoprocessamento em catalisar não só a enorme gama de variáveis envolvidas no planejamento de transportes, mas também em promover meios de modelar soluções perante a complexidade das políticas públicas, ambientais e de transportes envolvidas no processo.

3. ÁREA DE ESTUDO

A RMBH, composta por 34 municípios, é um polo gerador de fluxos por concentrar altas demandas de transporte para suprir as atividades comerciais, in-

dustriais, minerárias e de serviços. É interceptada por importantes rodovias federais como a BR-040, BR-381, BR-262 e BR-356, pelas quais trafegam também veículos de pequeno, médio e grande porte em circulação interna. Visando a mitigar os problemas de congestionamento e segurança relativos ao excesso de veículos que obrigatoriamente cruzam essas áreas urbanas, foi proposto um rodoanel de forma a oferecer uma alternativa externa à RMBH. De acordo com o DNIT (BRASIL, 2012), o Anel Sul é parte integrante do projeto de um complexo viário envolvendo o Rodoanel Norte (Trevo da Krupp-Ravena, com 67 km de extensão) e o Anel Leste (Olhos D'Água-Sabará com 22 km de extensão). A região proposta para a alça sul do rodoanel é caracterizada por densidade populacional elevada, atividades industriais, áreas vulneráveis e de proteção ambiental. De acordo com o edital e termo de referência, os pontos localizados no Contorno de Betim e no Jardim Canadá,

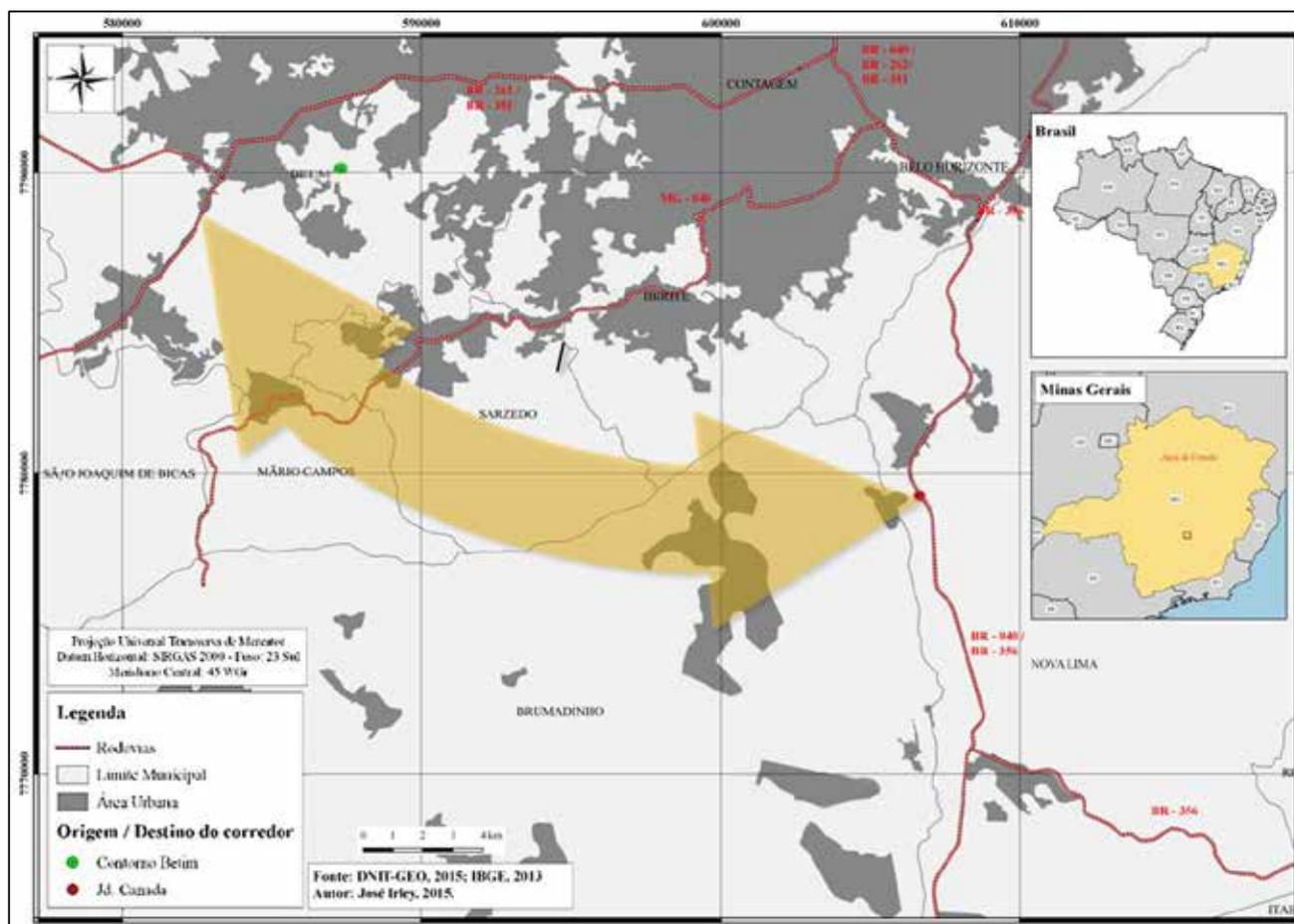
no município de Nova Lima, foram eleitos pelo DNIT para origem e destino do Anel Sul. A Figura 1 ilustra a proposta, interligando os fluxos da BR-040 e da BR-381, respectivas saídas para Rio de Janeiro e São Paulo.

4. METODOLOGIA

Para o desenvolvimento deste trabalho foram necessárias a coleta e a construção da base de dados, tanto tabulares quanto geoespaciais. No que tange à documentação técnica, foram utilizados como norteador para a modelagem de corredores os “Termos de referência para o estudo do traçado e elaboração do projeto executivo de engenharia para o contorno rodoviário sul da região metropolitana de Belo Horizonte BR-040/MG” (BRASIL, 2012). Este documento se caracteriza como Anexo I – Projeto básico para contratação de serviço segundo a alínea I, § 2º, Art. 7º, da Lei nº 8.666 de 21/06/93, cujo número de processo é o 50600.032686/2011-

Figura1:

Localização da área de estudo



78 (BRASIL, 2012). Estes termos orientam quanto à determinação de variáveis em seu próprio texto, assim como indicam demais manuais e documentos técnicos para consulta, como as Instruções de Serviço (IS) e Escopos Básicos (EB), além de terem sido utilizados como diretrizes para a modelagem geográfica deste trabalho.

As bases geográficas (Tabela 1) que compõem deste estudo foram adquiridas mediante contatos telefônicas e visitas. Para facilitar a organização, os dados foram separados em subconjuntos, conforme a descrição dos termos de referência do Anel Sul na RMBH, informando que o estudo de traçado deverá considerar simultaneamente “questões ambientais, culturais, sociais, comunitárias, geográficas, de engenharia e financeiras envolvidas no estudo do empreendimento” (BRASIL, 2012). A arquitetura do modelo seguiu as orientações de Nóbrega (2014).

4.1 TRATAMENTO E PROCESSAMENTO DOS DADOS

Um procedimento essencial que antecedeu o processamento dos dados geográficos foi a padronização do sistema de coordenadas, posto que o simples emprego da projeção cartográfica correta pode evitar inconsistências nas medições de um projeto linear de engenharia. Neste estudo, todos os dados foram reprojatados para o sistema geodésico Sirgas 2000 com projeção UTM-23S.

Na sequência, devido à necessidade de emprego de dados em formato matricial para a análise multicriterial utilizada no modelo, os dados originais em formato vetorial e tabular (Tabela 1) foram convertidos para o formato matricial. A transformação de dados vetoriais (discretos) em dados matriciais (contínuos) possibilita a realização da álgebra de mapas. A Figura 2 ilustra um exemplo de transformação de dados de formato vetorial para matricial e sua respectiva integração com outros dados matriciais para a composição da superfície de custo acumulado por meio da álgebra de mapas.

4.2 ANÁLISE MULTICRITERIAL, SUPERFÍCIES DE ESFORÇO E CORREDORES DE VIABILIDADE

Esta etapa consistiu na aplicação da técnica AHP para padronização das variáveis e construção da superfície de custo acumulado. O emprego se respaldou em três níveis: intravariáveis, intervariáveis e intergrupos, sendo este último o responsável pela integração da superfície de custo final. O cerne desse processo consiste na atribuição de índices de importância a variáveis do modelo que exigem conhecimento multidisciplinar.

No nível intravariável, cada dado de entrada foi analisado para que fossem convertidos em planos de informação do modelo. A documentação técnica do projeto

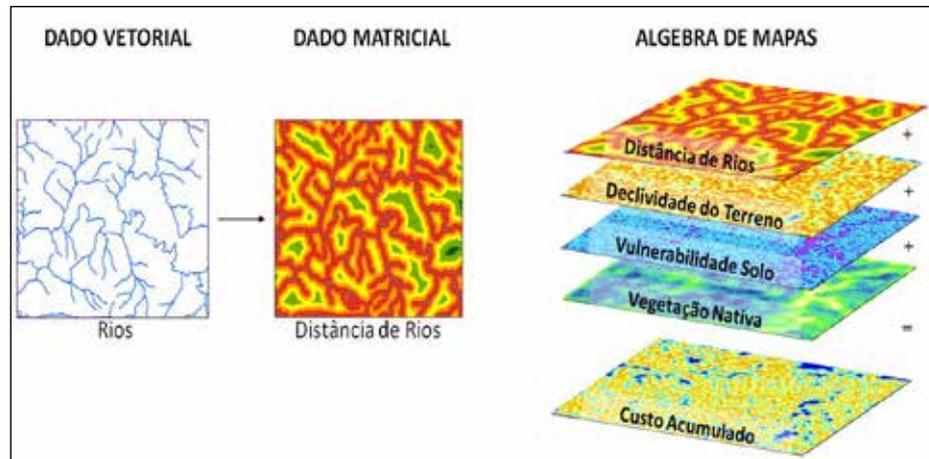
Tabela 1:

Organização da base de dados

Dado	Escala	Fonte	Data	Tipo	Subconjunto
Áreas urbanizadas	1:50000	ZEE-MG (2009)	2009	Vetor	Socioeconômico
População	1:500000	IBGE (2010)	2010	Tabela	
Assentamentos rurais	1:10000	INCRA (2015)	2015	Vetor	
Patrimônio arqueológico – Distância	1:50000	IPHAN (2015)	2015	Tabela	
Patrimônio histórico, artístico e cultural – Distância	1:50000	Prefeituras (2015)	2015	Tabela	
Rodovias – Densidade	1:10000	DNIT – GEO/Vetorização	2015	Vetor	Mercadológico e suporte logístico
Vias urbanas – Densidade	1:50000	Open Street Map	2015	Vetor	
Redes de gás – Distância	1:10000	GASMIG	2015	Vetor	
Linhas de transmissão – Distância	1:10000	CEMIG	2015	Vetor	
Recursos minerais	1:1000000	Geodiversidade – CPRM	2010	Vetor	
Interesse mineral	1:1000000	DNPM – SIGMINE	2015	Vetor	Biofísico
Vegetação remanescente	1:150000	Landsat 8 – NDVI	2015	Raster	
Nascentes – Distância das nascentes	1:1000000	IGAM	2014	Vetor	
Hidrografia – Densidade e distância da hidrografia	1:1000000	IGAM	2014	Vetor	
MDS - Modelo Digital de Terreno (Declividade)	1:10000	IGTEC	2009	Raster	
Vulnerabilidade à Erosão	1:1500000	ZEE-MG	2009	Vetor	
Risco de erosão – Filito, cárstico e movimento de massa	1:1000000	Geodiversidade - CPRM	2010	Vetor	
Espelho d'água	1:1000000	Vetorização	2015	Vetor	Restrições ambientais
Unidade de conservação – Proteção integral	1:50000	ZEE-MG	2009	Vetor	
Unidade de conservação – Uso sustentável	1:50000	ZEE-MG	2009	Vetor	
Cavernas – Distância	1:50000	SECAV	2015	Vetor	

Figura2:

Ilustração do tratamento e processamento de dados empregados no modelo



foi consultada para verificar como cada variável poderia ser explorada e quais os graus de importância de cada classe presente no dado. Tomam-se como exemplo as distâncias euclidianas dos cursos d'água na Figura 2: enquanto o dado vetorial informa apenas a presença ou não de rio, o dado matricial informa o quão distante ele está. Essas distâncias foram categorizadas e ponderadas seguindo critérios que indiretamente refletem na possível presença de mata ciliar, no alto custo para travessia ou mesmo em solos moles e colapsáveis. Procedeu-se com a conversão de dados vetor-matriz e o respectivo emprego da técnica AHP para cada variável, conforme ilustra a Figura 3 (superior), gerando assim planos de informação que foram integrados por grupo para o desenvolvimento do segundo nível da análise multicriterial, como proposto em Nóbrega (2013).

Para o nível intervariáveis, os planos de informação foram organizados em 4 grupos: restrições ambientais, biofísico, mercadológico e suporte logístico e socioeconômico. As ponderações entre os diferentes planos de informação por grupo foram resultantes de consultas a especialistas. Foram adotadas comparações pareadas para evitar inconsistências nos resultados, como descrito em Sadasivuni et al. (2009). Como resultado foram geradas superfícies de esforço acumulado que correspondem a mapas em formato matricial onde cada célula é representada pelo valor calculado de seu respectivo esforço (ou custo de implantação) acumulado das variáveis que participaram da composição em cada grupo. A Figura 3 (centro) ilustra esse processo para as variáveis de um grupo Biofísico. O processo foi reproduzido para os demais grupos, resultando em quatro superfícies de esforço acumulado que serviram

como entrada para o terceiro nível do processo AHP – intergrupos.

A Figura 3 (inferior) ilustra a integração dos grupos em quatro diferentes configurações de valores, por sua vez obtidos por meio de perspectivas distintas que visaram focar a viabilidade dos corredores na preservação de interesses próprios. Cada perspectiva adotada foi intencionalmente focada em defender interesses ambientais, redução de custos de engenharia, atendimento a demanda mercadológica/logística, ou em minimizar impactos socioeconômicos negativos. Como resultante foram produzidos quatro cenários distintos, cada qual conservador em seus interesses, para que com isso pudessem ser computadas as alternativas de corredores.

Uma vez computadas as superfícies integradas de esforço para cada cenário, estas foram utilizadas como base para simular computacionalmente o esforço previsto para conectar os pontos de origem e destino, pontos inicial e final do projeto, localizados na BR-040 e na BR-381, respectivamente. Esse cálculo é feito em duas etapas, na qual inicialmente são computados os custos de afastamento (esforço vs. distância) dos dois pontos extremos, para então integrar os dois mapas resultantes em uma superfície final que revela o corredor de menos esforço, conseqüentemente o de maior viabilidade segundo a perspectiva do cenário adotado (Figura 4).

5. RESULTADOS

Todo o desenvolvimento metodológico teve como objetivo a geração de corredores de transporte, cujos resultados estão apresentados na Figura 5. Assim,

Figura 3: Regras de decisão intravariáveis, intervariáveis e intergrupos utilizadas no modelo

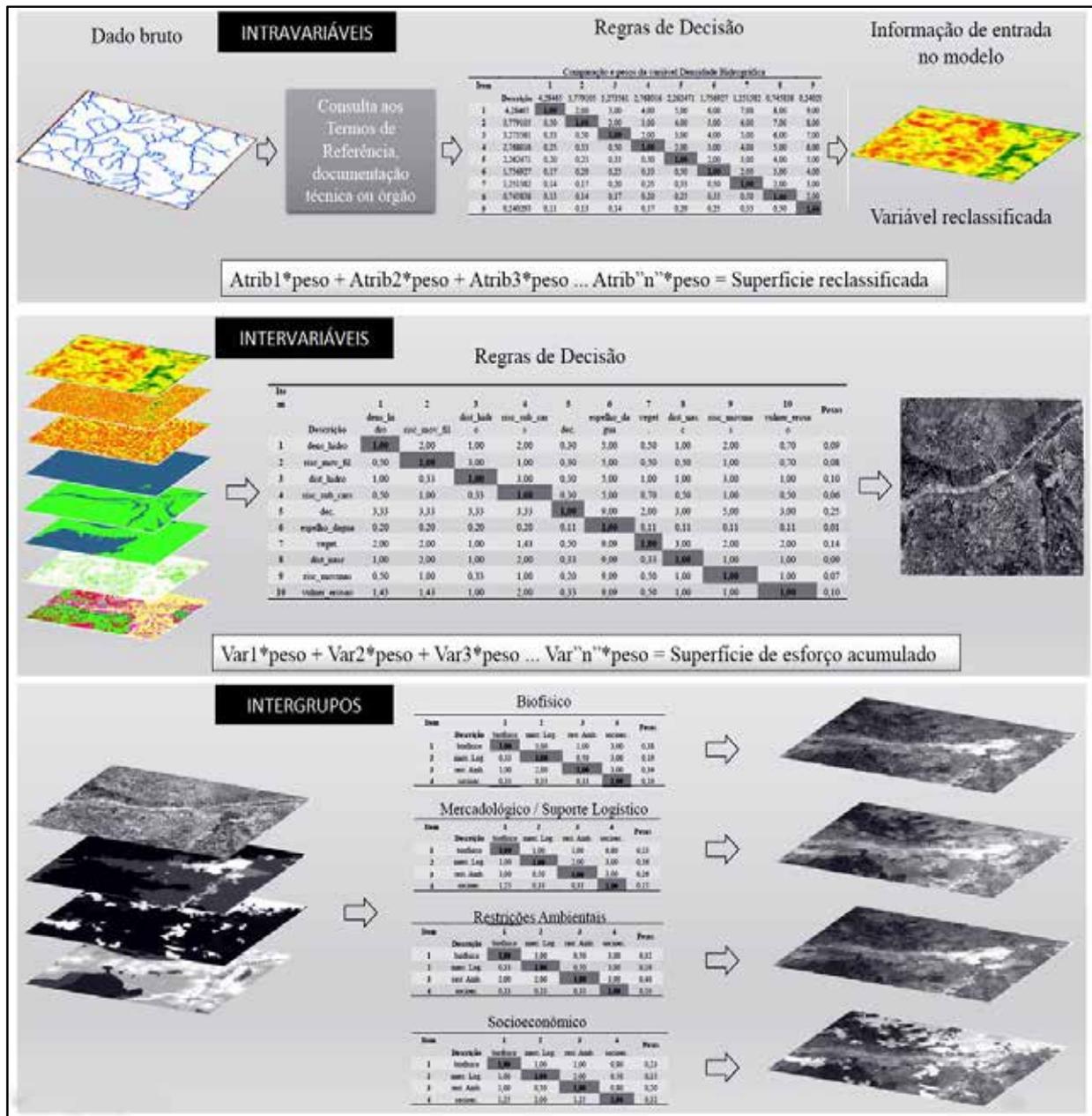
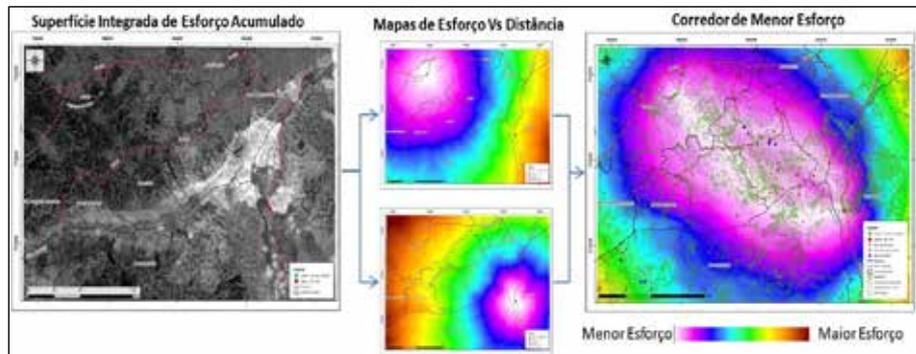


Figura 4: Ilustração do processo de cálculo de um corredor de viabilidade



foram produzidos quatro cenários: restrições ambientais, biofísico, socioeconômico e mercadológico e suporte logístico. Estes cenários respeitaram a origem e destino presentes nos termos de referência.

A partir do desenho dos corredores foi possível calcular métricas comparativas como extensão do projeto e interseção dos corredores para com outras bases de dados, a fim de quantificar valores para avaliar o impacto e comparar as alternativas. Para demonstrar esta análise de sensibilidade, foram escolhidas quatro variáveis que geralmente são preponderantes em avaliação de alternativas de corredores viários: declividade, área urbana, unidades de conservação e área vegetada. Contudo, a metodologia pode ser aplicada para quantificar o número de residências a serem afetadas, de rios a serem transpostos ou mesmo para monetizar o impacto de cada corredor alternativo, a depender da disponibilidade de dados presentes na área de estudo.

Com relação à extensão das alternativas, o cenário ambiental produziu um corredor com 24,8 km, o cenário biofísico um corredor com 25 km, o mercadológico com 25,2 km e o cenário socioeconômico produziu duas alternativas com 25,2 km e 23,7 km de extensão, respectivamente.

Quanto à análise de declividade do terreno, os resultados mostraram que, inevitavelmente, os traçados dos corredores deverão interceptar áreas de alta declividade, fato justificado pela natureza geomorfológica da região. Entretanto, no modelo se considerou o alto esforço para transpor essas áreas e a construção dos corredores evitou áreas com rampas acentuadas. Uma síntese da declividade média por corredor para os quatro cenários avaliados mostrou que o cenário biofísico apresentou menor média (13,7%), seguido pelo cenário mercadológico/logístico (13,9%) e o de restrições ambientais (14,1%). O cenário socioeconômico apresentou a maior média (14,6%).

Figura 5:

Corredores computados por cada cenário

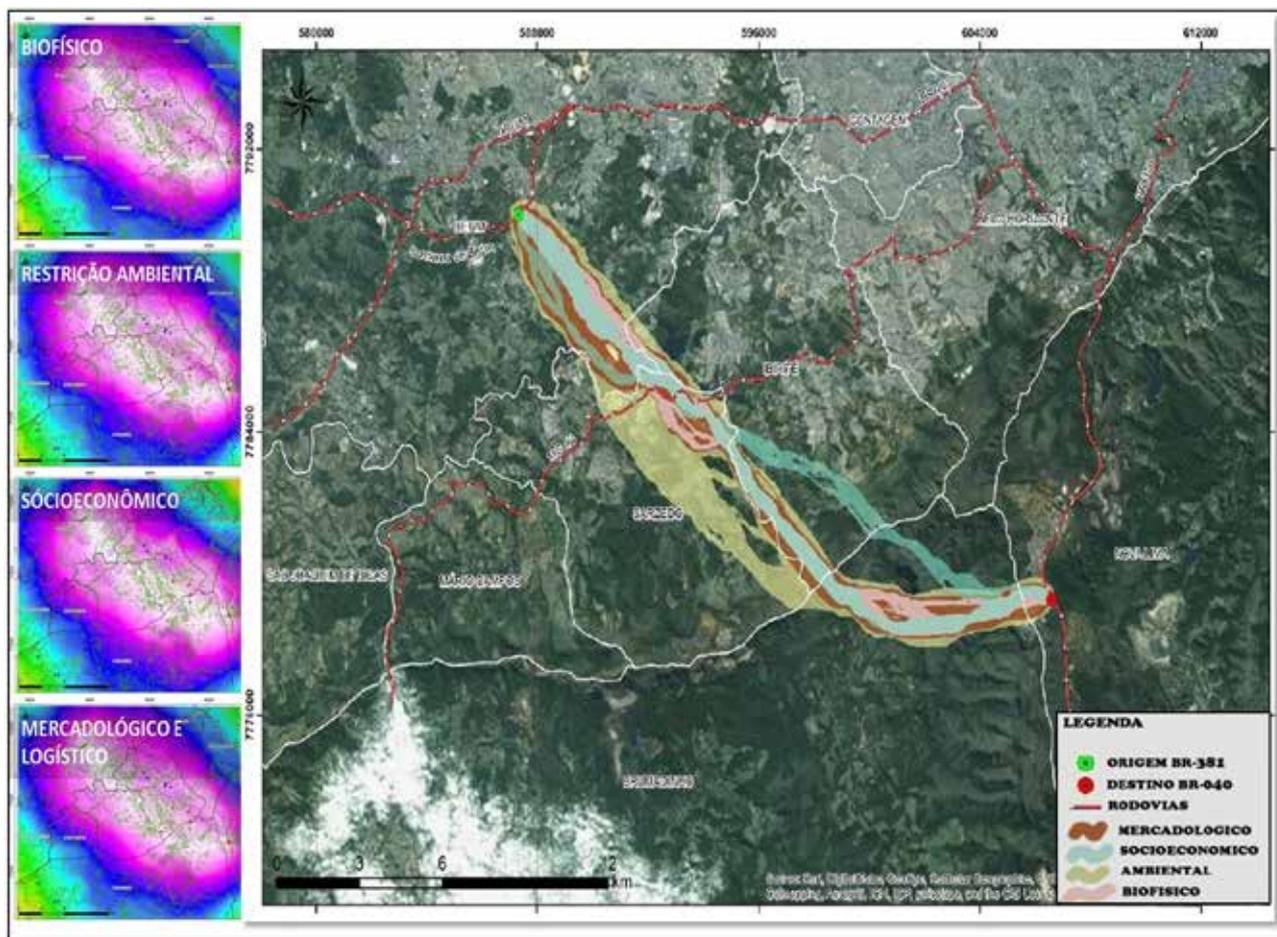


Gráfico 1:

Métricas de impacto para áreas urbanas, vegetadas e pertencentes a unidades de conservação interceptadas pelos corredores dos cenários analisados



Para as variáveis área urbana, unidades de conservação e vegetação, as áreas afetadas podem ser visualizadas no Gráfico 1. Em relação ao cenário biofísico, para o qual a ponderação das variáveis vegetação e unidades de conservação foi maior, observou-se que quando se proporcionou liberdade ao modelo para construir o corredor, os resultados se mostraram melhores, correspondendo a menores áreas impactadas. No subconjunto socioeconômico, o qual evidencia a preferência pelo distanciamento das áreas urbanas, observou-se que o resultado alcançou o menor impacto dentre todos os cenários avaliados.

Contudo, as análises revelam que existem diversas possibilidades de uso quando se identifica o que sofrerá intervenção e quanto. O modelo permite aperfeiçoar as análises de risco e otimizar o preenchimento da matriz de impacto ambiental utilizada para avaliar as alternativas do projeto.

Outras duas propostas foram desenvolvidas considerando os cenários biofísico e socioeconômico, tendo como origem o ponto instituído pelo DNIT na BR-381 e como destino a BR-040, sem a definição de uma localidade específica (Figura 7). A ideia foi verificar se o local na BR-040 apresentado no edital corresponde ao local de maior viabilidade segundo os critérios adotados nesta investigação. Embora não quantificadas, as análises preliminares indicam que o ponto de origem apresentado no edital do projeto é justificado quando considerado a necessidade de distanciar o rodovial de áreas urbanas existentes, todavia o modelo apontou alternativas de conexão

entre o Rodoanel e a BR-040 que podem produzir menores impactos ambientais e possivelmente menores custos de engenharia por interceptarem áreas mais planas.

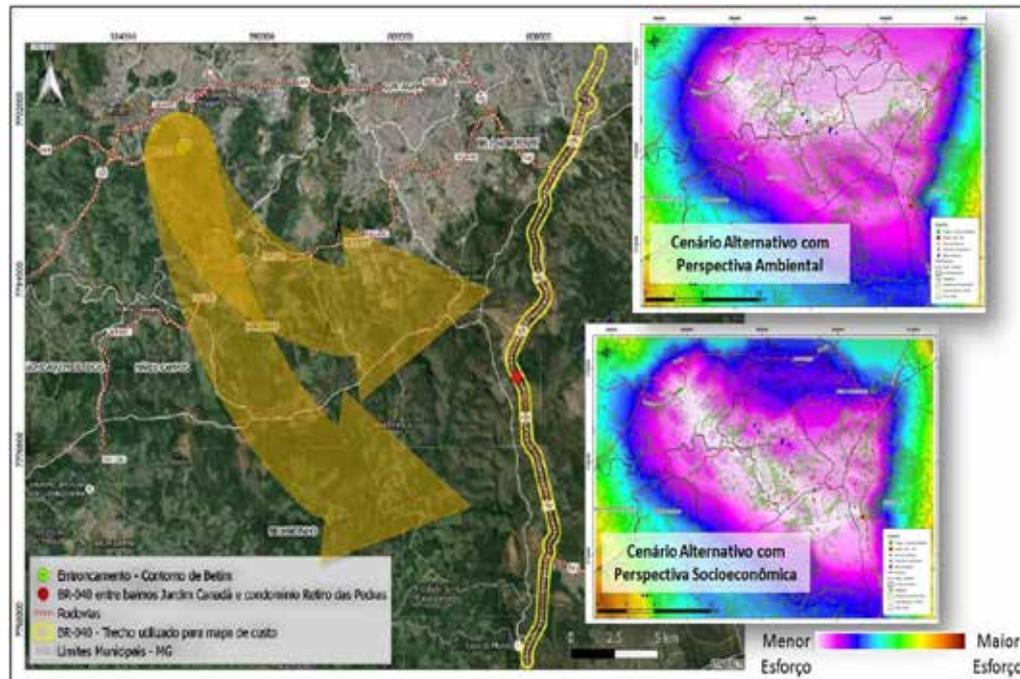
6. CONCLUSÃO

Na realização desse trabalho, deparamos com diversos desafios a respeito da interpretação dos termos de referência e da documentação técnica e processamento de dados. Todavia o objetivo principal de obter traçados alternativos para o Rodoanel Sul da região metropolitana de Belo Horizonte a partir de perspectivas distintas, realísticas e confrontantes foi atingido. Foram gerados corredores considerando diferentes variáveis, baseados no emprego de geotecnologias e análise multicriterial. Os resultados permitiram uma interação harmônica entre a aplicação desta técnica e a modelagem de dados geográficos. A capacidade operacional do modelo em processar áreas extensas com profundo detalhamento contextual das análises, bem como a otimização do tempo e da capacidade analítica dos atores envolvidos no modelo contribuem de forma positiva para a modernização do processo de planejamento de transportes.

O trabalho piloto de análise multicritério espacial mostrou grande potencial de aplicação em controle externo. Acredita-se que, de posse do modelo e de seu conhecimento operacional, as ações de fiscalização e controle externo poderão ser desenvolvidas com maior agilidade e acurácia.

Figura 7:

Processamento alternativo para testar a aderência dos pontos de origem e destino publicados no edital do projeto



Os corredores resultantes desse trabalho ainda não podem ser comparados ao traçado oficial do Rodoanel da RMBH, posto que este não apresenta a fase de estudo de viabilidade concluída. Espera-se que métricas comparativas possam ser quantificadas e analisadas a partir do término do projeto por parte do DNIT ou de sua apresentação para a comunidade.

7. AGRADECIMENTOS

Os autores agradecem ao CNPq pelo auxílio no desenvolvimento da pesquisa.

REFERÊNCIAS

- BERBERIAN, C. F. Q. et al. O uso de geotecnologias como uma nova ferramenta para o controle externo. *Revista do Tribunal de Contas da União, Brasília, DF*, n. 133, p. 40-53, 2015.
- BRASIL. DNIT – Departamento Nacional de Infraestrutura de Transportes. Coordenação Geral de Cadastro e Licitações. Termos de referência para o estudo do traçado e elaboração do projeto executivo de engenharia para o contorno rodoviário sul da região metropolitana de Belo Horizonte BR-040/MG. DNIT, Belo Horizonte, edital nº 0240/2012-00, processo nº 50600.032686/2011-78, p. 79-157, 2012.

LONGLEY, P. A. et al. *Sistemas e ciência da informação geográfica*. 3. ed. Porto Alegre: Bookman, 2013.

NOBREGA, R. A. A. *Ferrovia Norte-Sul: estudo dos traçados alternativos para escoamento eficiente da produção*. In: ENCONTRO REGIONAL DA 3ª CÂMARA DE COORDENAÇÃO E REVISÃO DO MINISTÉRIO PÚBLICO FEDERAL – REGIÕES SUL/SUDESTE, 2, 2013, Florianópolis. Anais... Florianópolis, 17 out. 2013.

_____. *Relatório de fechamento de projeto de extensão: ligação Pato Branco/PR – Cruz Alta/RS*. Belo Horizonte: UFMG, 2014.

NOBREGA, R. A. A. et al. Bridging decision-making process and environmental needs in corridor planning. *Management of Environmental Quality: an International Journal, Mississippi*, v. 20, n. 6, p. 622-637, 2009.

SAATY, T. L. *Transport planning with multiple criteria: the analytic hierarchy process applications and progress review*. *Journal of Advanced Transportation, New York*, v. 29, n. 1, p. 81-126, 1995.

SADASIVUNI, R. et al. *A Transportation corridor case study for multi-criteria decision analysis*. In: ASPRS ANNUAL CONFERENCE, 2009, Maryland. Anais... Maryland, 9-13 mar. 2009.

O potencial de dados de sensoriamento remoto na fiscalização de obras públicas



Osmar Abílio de Carvalho Júnior

é geólogo e doutor em Geologia pela Universidade de Brasília, onde atua como professor.



Roberto Arnaldo Trancoso Gomes

é geógrafo e doutor em Geografia pela Universidade Federal do Rio de Janeiro, além de professor na Universidade de Brasília.



Renato Fontes Guimarães

é engenheiro cartógrafo com doutorado em Geologia, além de professor na Universidade de Brasília.



RESUMO

Este artigo analisa as potenciais aplicações de técnicas de sensoriamento remoto para a fiscalização e o monitoramento de obras de infraestrutura. Atualmente, existe uma ampla disponibilidade de dados de sensoriamento remoto provenientes de diferentes sensores e plataformas, constituindo uma fonte de informação rápida e útil para descrever a composição dos elementos presentes na superfície e as suas mudanças ao longo do tempo. O aprimoramento dos sensores remotos com o aumento da resolução espacial, temporal e espectral embarcados em diferentes plataformas (satélites, aeronaves e veículos aéreos não tripulados) tem ampliado as perspectivas de estudos e aplicações dos dados de sensoriamento remoto, incluindo-se o monitoramento de infraestruturas públicas em construção ou em concessões. A extração de modelos digitais de elevação a partir de sensores remotos também constitui um importante atributo para descrever as feições de obras de infraestrutura. Os estudos mais comumente encontrados na literatura são referentes às feições urbanas e rodovias, sendo ainda pouco relatado estudos sobre fiscalização e monitoramento de obras, o que consiste em amplo campo de pesquisa e inovação. Inúmeras técnicas de detecção de mudança têm sido propostas e avaliadas para diferentes ambientes e alvos, devendo em cada tipo de ambiente e alvo realizar comparações e análise de acurácia para definir o melhor procedimento a ser

adotado. Pesquisas específicas para cada tipo de obra devem ser realizadas demonstrando o potencial real da detecção remota para a fiscalização em ambientes urbanos ou rurais.

Palavras-chave: Sensoriamento Remoto, detecção de mudança, modelo digital de elevação, classificação espectral

1. INTRODUÇÃO

A fiscalização efetiva das obras públicas é um fator preponderante para a minimização dos gastos públicos. Como exemplificação, auditorias realizadas pelo Fiscobras (plano anual de fiscalização de obras do TCU) no estado do Rio Grande do Norte durante o período 2011-2012 proporcionaram uma economia aos cofres públicos de R\$ 119.529.497,78 (SOUZA; BATISTA, 2013). O Fiscobras no ano de 2015 realizou 97 auditorias, totalizando um volume de recursos fiscalizados de R\$ 31 bilhões, em que 61 obras (62,9%) evidenciaram indícios de irregularidades graves (BRASIL, 2015).

No entanto, a dimensão continental do território brasileiro dificulta o processo tradicional de fiscalização, que exige uma atuação constante de profissionais no local. A necessidade crescente de infraestruturas para o crescimento econômico de longo prazo do país aliado ao alto grau de irregularidades torna premente o aprimoramento tecnológico para obter uma fiscalização contínua

das obras públicas em execução ou concessão em todo o território nacional (MIRANDA; MATOS, 2015; VITAL et al., 2015). No processo de fiscalização é fundamental obter informações precisas sobre a evolução das obras públicas (construção, reforma, fabricação, recuperação ou ampliação de bem público), buscando detectar eventuais inconsistências ou inexistência dos elementos presentes no projeto básico e nas especificações técnicas. Nessa conjuntura, o sensoriamento remoto pode ser uma importante ferramenta, proporcionando um monitoramento periódico de extensas áreas com baixo custo. Nos últimos anos, extensos esforços de pesquisa têm sido feitos para a detecção de mudanças utilizando imagens de sensoriamento remoto em diferentes cenários: (a) urbanos (HEGAZY; KALLOP, 2015; SUN et al., 2013); (b) agrícolas (MENKE et al., 2009; OLIVEIRA et al., 2014); e (c) áreas naturais e de preservação ambiental (COPPIN; BAUER, 1996; YADAV; KAPOOR; SARMA, 2012). Portanto, o sensoriamento remoto tem sido amplamente utilizado para avaliar a dinâmica espacial da superfície terrestre e a eficácia do planejamento territorial. Especificamente, estudos sobre a fiscalização de obras públicas usando sensoriamento remoto são pouco relatados nos periódicos científicos. A ausência de pesquisa para aplicação dessa técnica demonstra um extenso campo para inovações científicas, contendo relevância econômica e retorno imediato para a sociedade. O acervo de imagens multitemporais possibilita fornecer subsídios para a comparação do que foi planejado

em relação ao que foi executado em cada obra, tanto na dimensão espacial como temporal. Essa abordagem permite que se compare o cronograma pré-estabelecido da obra com o andamento real, permitindo reconstruir de forma total ou parcial o histórico financeiro e executivo.

Contudo, o emprego de imagens para a fiscalização de obras públicas apresentam dificuldades similares aos descritos no mapeamento de feições de áreas urbanas, tais como (CAVALLI et al., 2008; WENTZ et al., 2014): (a) os espectros dos pixels de ambientes urbanos são constituídos por misturas espectrais devido à alta heterogeneidade desses ambientes, podendo gerar confusões entre as classes no processo de classificação, e (b) as estruturas físicas das classes urbanas variam espacialmente, considerando diferentes composições de telhados, pavimentos e formas arquitetônicas. Assim, as obras de infraestrutura (edificações, pontes, estradas, hidrovias, ferrovias, entre outros) são constituídas por diferentes materiais (asfalto, pinturas, concreto, metal, vidro, telhas, vegetação e solo) que estão combinados em diferentes proporções (JENSEN; COWEN, 1999). Por exemplo, duas pontes podem ter comportamentos diferenciados tanto pelo tipo de materiais utilizados como pelas suas diferenças arquitetônicas. Além disso, outro fator de complexidade é que durante a evolução de uma obra ocorre uma intensa modificação dos elementos e padrões de construção. Portanto, o acompanhamento de uma obra por sensoriamento remoto exige a obtenção de informações constantes e detalhadas, além da elaboração



de uma metodologia especializada de processamento digital de imagens. Diferentes técnicas para a detecção de feições urbanas consolidadas ou em construções podem ser utilizadas. Normalmente, o método de interpretação visual é considerado o de maior acurácia, mas também é o mais demorado e caro. Alternativa para interpretação visual é utilizar classificações supervisionadas, não supervisionadas ou abordagens de sistema especialista sem supervisão e baseada no conhecimento (JAT; GARG; KHARE, 2006). Além disso, técnicas de detecção de mudança devem ser adotadas, considerando as diferentes abordagens de pré-classificação ou pós-classificação.

No presente trabalho são discutidas as questões relacionadas ao emprego do sensoriamento remoto na fiscalização das obras públicas. A definição das imagens a serem utilizadas é avaliada considerando os avanços ocorridos na melhoria dos atributos espaciais, espectrais, temporais e a na extração de dados altimétricos. Os progressos obtidos com o aumento das diferentes resoluções das imagens resultam em novos métodos para o tratamento e análise dos dados, com implicações na eficiência da fiscalização por sensoriamento remoto. Além disso, o presente trabalho apresenta uma revisão dos principais métodos de detecção de mudança, por ser um conjunto de procedimentos básicos que permite acompanhar as alterações no decorrer da obra e avaliar o seu ajuste ao projeto inicial.

2. CARACTERÍSTICAS DA RESOLUÇÃO TEMPORAL, ESPECTRAL E ESPACIAL DOS SENSORES REMOTOS NA FISCALIZAÇÃO DE OBRAS

2.1 RESOLUÇÃO ESPACIAL

A definição das classes referentes às obras de engenharia por sensoriamento remoto e suas implicações no cumprimento legal é altamente dependente da resolução espacial (tamanho do pixel), espectral (número de bandas espectrais) e temporal (período de revisita do mesmo local na superfície terrestre) da imagem. Os três fatores são importantes, mas prioritariamente a imagem deve conter uma alta resolução espacial para que possam ser individualizados os objetos presentes em uma obra. Não adianta ter uma alta resolução espectral se em um pixel existem diferentes elementos urbanos e comportamentos espectrais misturados. Normalmente, a identificação de um objeto urbano em uma imagem precisa ter uma representação mínima de quatro pixels (COWEN et al., 1995; JENSEN; COWEN, 1999). Conforme Small (2003), a resolução espacial mínima para



captar as estruturas urbanas é de 5 metros, também aplicável para as obras de engenharia.

As imagens de alta resolução espacial em ambientes urbanos permitem utilizar os elementos básicos de interpretação (tom, cor, textura, forma, tamanho, orientação, padrão, sombra, local e situação dos objetos na paisagem urbana) para identificar e julgar o seu significado. Atualmente, diferentes imagens de sensores orbitais de alta resolução espacial (inferior a 4 metros) estão disponíveis no mercado, destacando: GeoEye-1 (0,46m), WorldView-1 e 2 (0,46 m), WorldView-3 e 4 (0,31 m), Pleiades-1A e 1B (0,5 m), Kompsat-3A (0,55 m) e 3 (0,7 m), QuickBird (0,65), Gaofen-2 (0,8 m), TripleSat (0,8 m), Ikonos (0,82m), SkySat-1 e 2 (0,9m) e Spot-6 e 7 (1,5 m) (Tabela 1). O aumento da oferta de imagens de alta resolução espacial provenientes de satélites comerciais proporcionou um crescimento de técnicas de processamento digital de imagens visando estudos de infraestrutura, redes rodoviárias e elementos urbanos.

Uma importante inovação no mapeamento de áreas urbanas usando imagens de alta resolução espacial é o emprego da análise de imagem baseada em objetos geoespaciais (Geobia), que se diferencia dos tradicionais métodos baseados em pixels. Na Geobia, a imagem é segmentada em regiões relativamente homogêneas (objetos de imagem) antes da classificação (BLASCHKE, 2010; MYINT et al., 2011). Assim, a classificação utiliza como unidade básica os segmentos e seus atributos em vez dos pixels. O alto grau de variabilidade espectral dentro de uma

Tabela 1:

Descrição dos principais satélites orbitais. As imagens disponíveis gratuitamente são demarcadas com um asterisco (*)

SATÉLITES DISPONÍVEIS	RESOLUÇÃO ESPACIAL	RESOLUÇÃO TEMPORAL	RESOLUÇÃO ESPECTRAL
ALOS	2,5 m	Variável	1 banda pancromática
	10 m	Variável	3 bandas no visível
	10 a 100 m	Variável	1 banda no infravermelho
CARTOSAT	2,5 m	5 dias	1 banda pancromática
CBERS*	5 m	Variável	1 banda (pancromática)
	10 m	Variável	2 bandas no visível
			1 banda no infravermelho
	20 m	26 dias	3 bandas no visível
	40 m	26 dias	1 banda no infravermelho
			1 banda pancromática
	80 m	26 dias	2 bandas no infravermelho
1 banda no termal			
64 m	5 dias	3 bandas no visível	
EROS	0,7 m	Programado	1 banda no infravermelho
FORMOSAT	2 m	Programado	1 banda (pancromática)
	8 m	Programado	3 bandas no visível
			1 banda no infravermelho
GAOFEN	0,8 m	5 dias	1 banda (pancromática)
	3,2 m	5 dias	3 bandas no visível
			1 banda no infravermelho
GEOEYE	0,5 m	Programado	1 banda (pancromática)
	2 m	Programado	3 bandas no visível
IKONOS	1 m	Variável	1 banda no infravermelho
			1 banda pancromática
	4 m	Variável	3 bandas no visível
KAZEOSAT-1	1 m	Programado	1 banda no infravermelho
	4 m	Programado	3 bandas no visível
KOMPSAT 2	1 m	Programado	1 banda pancromática
	4 m	Programado	3 bandas no visível
KOMPSAT 3	0,7 m	Diária (possível)	1 banda no infravermelho
	2,8 m	Diária (possível)	3 bandas no visível
KOMPSAT 3A	0,55 m	Diária (possível)	1 banda no infravermelho
	2,2 m	Diária (possível)	3 bandas no visível
	5,5 m	Diária (possível)	2 bandas no infravermelho
LANDSAT 5*	30 m	16 dias	3 bandas no visível
	120 m	16 dias	3 bandas no infravermelho
LANDSAT 7*	15 m	16 dias	1 banda no termal
	30 m	16 dias	1 banda pancromática
			3 bandas no visível
	60 m	16 dias	3 bandas no infravermelho
LANDSAT 8	15 m	16 dias	1 banda no termal
	30 m	16 dias	1 banda pancromática
			4 bandas no visível
			3 bandas no infravermelho
100 m	16 dias	1 banda aerossol	
			1 banda cirrus
			2 bandas no termal

SATÉLITES DISPONÍVEIS	RESOLUÇÃO ESPACIAL	RESOLUÇÃO TEMPORAL	RESOLUÇÃO ESPECTRAL
PLEIADES	0,5 m	Diária (possível)	1 banda pancromática
	2 m	Diária (possível)	3 bandas no visível 1 banda no infravermelho
RAPIDEYE	5 m	Programado	4 bandas no visível 1 banda no infravermelho
SENTINEL 1*	5 a 20 m	12 dias	Diferentes bandas no radar
SENTINEL 2*	10 m	10 dias com possibilidade de 5 dias	3 bandas no visível 1 banda no infravermelho
	20 m		4 bandas <i>red edge</i> 2 bandas no infravermelho
	60 m		1 banda aerossol 1 banda cirrus 1 banda vapor d'água
SPOT	1,5 m	Diária (possível)	1 banda pancromática
	6 m	Diária (possível)	3 bandas no visível 1 banda no infravermelho
SKYSAT	1,1 m	Programado	1 banda pancromática (possibilidade de gerar um vídeo de 90 segundos)
	2 m	Programado	3 bandas no visível 1 banda no infravermelho
TERRASAR-X	0,25 a 40 m	Programado	Diferentes bandas no radar
TH-01	2 m	Programado	1 banda pancromática
	10 m	Programado	3 bandas no visível 1 banda no infravermelho
TRIPLESAT	1 m	Programado	1 banda pancromática
	4 m	Programado	3 bandas no visível 1 banda no infravermelho
WORLDVIEW	0,3 m	Programado	1 banda pancromática 1 banda aerossol
	1,24 m	Programado	4 bandas no visível 1 banda <i>red edge</i> 2 bandas no infravermelho
	3,7 m	Programado	8 bandas no infravermelho
	30 m	Programado	12 bandas Cavis (nuvens, aerossóis, vapor, gelo e neve)
TERRA/AQUA (Sensor-MODIS)*	250 m	1-2 dias	1 banda no visível 1 banda no infravermelho
	500 m		2 bandas no visível 3 bandas no infravermelho
	1000 m		7 bandas no visível 16 bandas no infravermelho 6 bandas no termal
TERRA (Sensor ASTER)*	15 m	Variável	2 bandas no visível 1 banda no infravermelho
	30 m		6 bandas no infravermelho
	90 m		5 bandas no termal

classe (sombras, ângulo do sol, as lacunas na copa das árvores etc.) pode dificultar a classificação baseada em pixel e favorecer técnicas baseadas em objetos que são representados por valores médios do segmento (YU et al., 2006).

Um problema da classificação baseada em objetos é a dependência da etapa de segmentação, que pode gerar segmentos em excesso ou em quantidade reduzida em relação às feições do terreno (LIU; XIA, 2010). Normalmente, a falta de segmentos é considerada pior do que o excesso (KIM; MADDEN; WARNER, 2009). A minimização desse tipo de erro pode ser obtida através de sucessivos testes de segmentações antes da classificação (TRIAS-SANZ; STAMON; LOUCHET, 2008) e por análise de acurácia dos segmentos (DORREN; MAIER; SEIJMONSBERGEN, 2003; KIM; MADDEN; XU, 2010).

2.2 RESOLUÇÃO ESPECTRAL

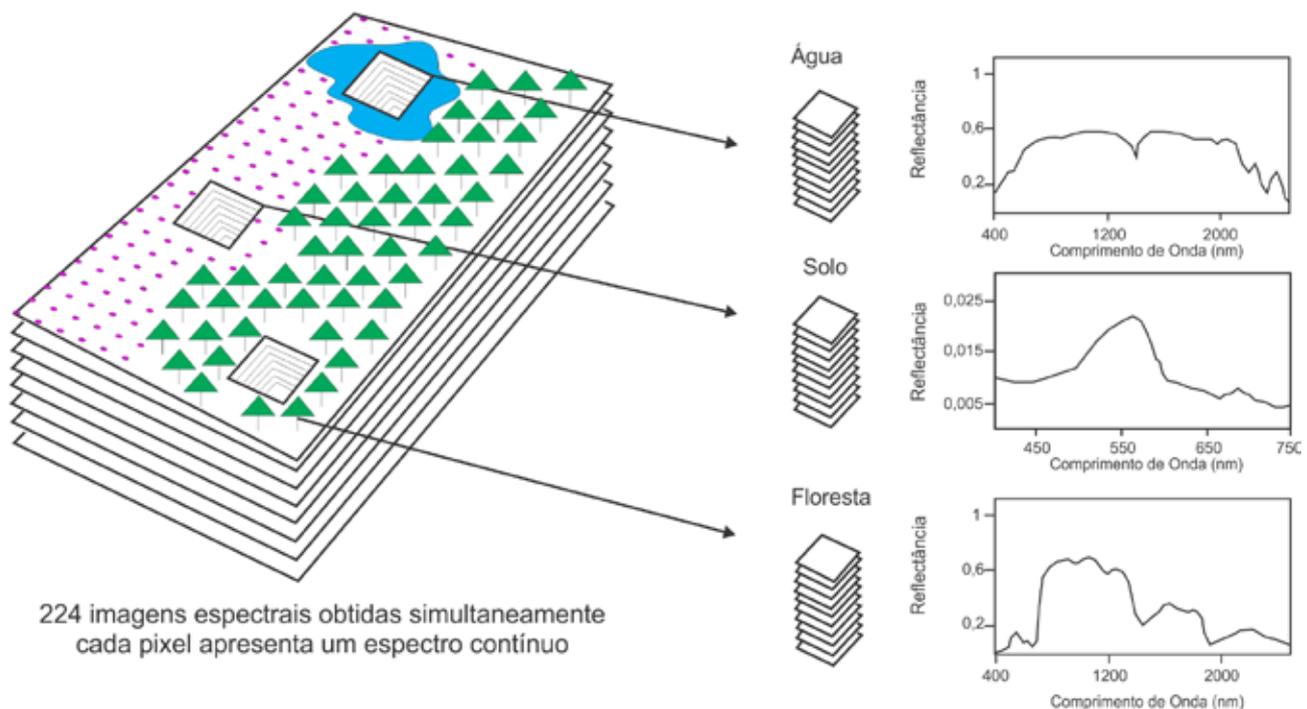
O conhecimento pormenorizado das características espectrais permite uma acurada identificação dos elementos superficiais. Nesse propósito, diversos estudos desenvolvem bibliotecas espectrais específicas

para diferentes alvos a partir de espectrômetros de campo ou laboratório para servir de suporte à classificação, tais como elementos urbanos (BEN-DOR; LEVIN; SARON, 2001; HEROLD et al., 2004), minerais (CLARK et al., 2007), plantações (RAO; GARG; GHOSH, 2007) e áreas alagadas (ZOMER; TRABUCCO; USTIN, 2009).

O advento das imagens hiperspectrais (**Figura 1**), caracterizadas por conter centenas de bandas espectrais estreitas e contínuas no espectro solar refletido, permitiu o desenvolvimento de novas técnicas que aprimoraram a detecção e quantificação dos materiais (CARVALHO JÚNIOR et al., 2003). A classificação espectral busca converter os sinais espectrais que refletem as coberturas urbanas em categorias que representam a natureza física da superfície. Além disso, essas imagens permitem uma análise espectral subpixel com a estimativa de abundância dos materiais superficiais contidos no pixel a partir da análise linear de mistura (PHINN et al., 2002; WU; MURRAY, 2003) e sua variação Multiple Endmember Spectral Mixture Models (MESMA) (DEMARCHI et al., 2012; FRANKE et al., 2009). Essa abordagem possibilita tratar o problema de heterogeneidade espectral dentro

Figura 1:

Concepção do sensor hiperspectral AVIRIS, segundo a qual a alta resolução espectral torna a informação de um dado pixel próxima à obtida por meio de medições realizadas em laboratório e/ou campo. (Modificado de GREEN et al., 1998).





de uma mesma classe, que é uma das principais dificuldades em áreas urbanas (HEIDEN; SEGL; KAUFMANN, 2007). Um mesmo tipo de classe pode ser caracterizado por diversos materiais distintos espectralmente ou ter composições diferenciadas devido à deterioração ao longo do tempo (DEMARCHI et al., 2012). Por exemplo, os telhados podem ser constituídos por diferentes materiais e, além disso, modificar o seu comportamento espectral devido ao acúmulo de fungos e sujeira.

Normalmente, as imagens hiperespectrais mais utilizadas nos estudos urbanos são do sensor Hyperion a bordo do satélite EO-1 e os provenientes de sensores embarcados em aeronaves, como o Airborne Visible Infrared Imaging Spectrometer (AVIRIS), Compact Airborne Spectrographic Imager (CASI) e o Hyperspectral Mapper (HyMap).

2.3 RESOLUÇÃO TEMPORAL

O advento de sensores orbitais com alta resolução temporal tem promovido uma nova abordagem na classificação de alvos superficiais, que consideram aspectos de mudanças cíclicas ou com trajetórias conhecidas de um evento ou de um objeto que podem ser descritas e identificadas por uma assinatura temporal. A similaridade espectral entre os diferentes tipos de vegetação, caracterizadas por serem constituídas pelas mesmas feições de absorção, pode apresentar dificuldades para serem distinguidas a partir de uma única imagem no tempo. No entanto, séries temporais de sensoriamento remoto permitem estabelecer uma assinatura fenológica característica, identificando diferentes tipos de ecossistemas naturais

(ABADE et al., 2015), fitofisionomias (CARVALHO JÚNIOR; HERMUCHE; GUIMARÃES, 2006; CARVALHO JÚNIOR et al., 2008) e plantios (COUTO JÚNIOR; CARVALHO JÚNIOR; MARTINS, 2012; SAKAMOTO et al., 2005). Essa abordagem permite detectar eventos naturais singulares, tais como fogo (CARVALHO JÚNIOR et al. 2015) e inundações (AIRES et al., 2014).

Na fiscalização de obras de engenharia, é fundamental avaliar a resolução temporal dos sensores remotos. A obra progride através de um calendário programado que permite estabelecer quais sensores possuem adequadas resoluções temporais para a sua fiscalização. As condições climáticas (interferência atmosférica e cobertura de nuvem) também afetam a aquisição de imagens, podendo ser críticas para sensores ópticos em algumas localidades, como na região amazônica. Nesse caso, deve-se buscar satélites de alta frequência de revisita para a seleção de imagens ou adotar sensores de imagem radar.

Os sensores de alta resolução temporal possuem normalmente baixa resolução espacial, como é o caso do sensor MODIS e o Advanced Very High Resolution Radiometer (AVHRR). No entanto, muitas missões utilizam uma constelação de satélites idênticos que orbitam a Terra de forma sincronizada, permitindo uma sucessão de imagens com alto grau de revisita e alta resolução espacial, como é o caso dos seguintes programas orbitais: Rapideye (5 satélites); TripletSat (3 satélites); Pleiades (2 satélites) e o Spot 6-7 (2 satélites). Apesar de séries temporais constituídas com imagens com mesmas especificações facilitar o desenvolvimento de métodos automatizados, estudos conciliando

imagens de diferentes sensores tornam-se um desafio de pesquisa.

2.4 MODELOS DIGITAIS DE ELEVAÇÃO

Os modelos digitais de elevação (MDE) são representações em 3D da superfície terrestre que reproduzem as feições naturais e antrópicas da paisagem. A construção desses modelos inicia-se a partir de fotografias aéreas analógicas e passa a ser desenvolvida por outras tecnologias, tais como: fotografia aérea digital, sensor ótico orbital de alta resolução, radar interferométrico e radar laser aerotransportado. Atualmente, veículos aéreos não tripulados (VANT) carregam a bordo esses diferentes sensores que elaboram MDEs. Recentemente, os sensores de radar laser móvel podem ser acoplados em qualquer plataforma (um barco, um automóvel etc.).

Os sensores óticos a bordo de satélites mapeiam a superfície terrestre em dois pontos de vista distintos, possibilitando a extração de MDEs. Dentre eles destacam-se satélites com diferentes resoluções espaciais, desenvolvidos por agências espaciais, tais como a série Spot, IRS, Cartosat-1, Alos-Prism, WorldView-2, QuickBird-2, Ikonos-2, Aster, entre muitos outros. A aquisição dos dados pode ser ao longo de uma única linha de varredura (Spot) ou adquirido paralelamente com uma área de superposição (QuickBird) (POLI; TOUTIN, 2012). As imagens orbitais vêm preenchendo a lacuna na produção de MDEs precisos, porém câmeras aéreas de última geração a bordo de aviões, como a ADS 80, produzem atualmente precisão compatível com imagens de sensores remotos de alta resolução, como o Worldview (HOBI; GINZLER, 2012).

Os MDEs elaborados por Interferometric Synthetic Aperture Radar (InSAR) são obtidos a partir do

retorno das diferenças de fase das ondas para o satélite. Dentre os MDEs provenientes desse método, destacam-se o Radarsat 1 e 2, Sentinel 1 e SRTM, sendo este último o mais utilizado mundialmente (JARIHANI et al., 2015; MUSA; POPESCU; MYNETT, 2015). Esses produtos têm a vantagem de mapear de forma contínua, mesmo com a cobertura de nuvens. A **Figura 2** mostra um exemplo de MDE de uma área urbana (A) e uma fotografia aérea sobreposta a este MDE (B).

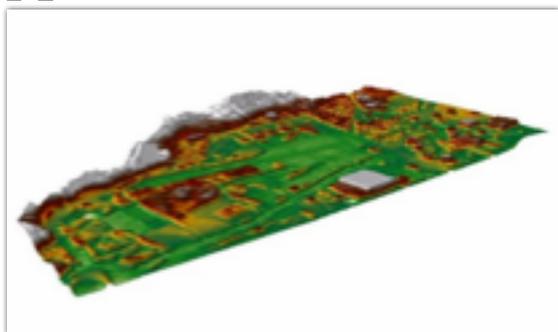
A produção de MDEs a partir de radar laser, que mede a distância entre um sensor e um alvo baseado na metade do tempo gasto entre a emissão e detecção do pulso (BALTSAVIAS, 1999), já possui algumas décadas. Porém, grandes avanços tecnológicos nos últimos anos possibilitaram identificar tênues mudanças de elevação a partir de densa nuvem de pontos, possibilitando mapear e distinguir objetos com pequenas variações de textura (MENG; CURRIT; ZHAO, 2010). Os sensores laser podem ser acoplados em aeronaves, satélites e veículos aéreos não tripulados. Recentemente foi desenvolvido o sistema móvel de varredura a laser que adquire dados em 2D e 3D e que são implantados em qualquer plataforma móvel terrestre ou náutica (PUENTE et al., 2013). Essa tecnologia apresenta grandes vantagens, tais como: (a) diminuição de custo devido à captura de dados em alta velocidade; (b) alta densidade de pontos, garantindo um levantamento planialtimétrico completo que diminui o número de dados questionáveis; e (c) visualização 3D que permite verificar se os objetos mapeados correspondem às condições do mundo real.

Os MDEs são amplamente utilizados na detecção e descrição de feições urbanas (KIM; NEVATIA, 2004). No entanto, as obstruções no ambiente urbano denso ain-

Figura 2:

MDE de área urbana (A) e Fotografia aérea sobreposta ao MDE (B).

A



B



da são um grande obstáculo para o mapeamento, sendo necessárias metodologias para suprir a falta de informação e insuficiência de textura para identificar as feições (DURUPT; TAILLANDIER, 2006). Novas tecnologias que utilizam radar laser, como o Light Detection and Ranging (Lidar), estão sendo desenvolvidas para reprodução em 3D das feições antrópicas com alta resolução, porém a sua cobertura é limitada, e a aquisição de dados e o processamento demandam um alto custo (BAUGH et al., 2013).

3. PROCESSAMENTO DIGITAL DE IMAGENS E DETECÇÃO DE MUDANÇA

As técnicas de detecção de mudança usando sensoriamento remoto podem fornecer importantes informações para o monitoramento de obras de engenharia e infraestrutura, como: (a) a área e a taxa em que as construções são desenvolvidas; (b) a distribuição e a relação espacial dos tipos de mudanças avaliando o desempenho relativo das atividades da construção civil com o cronograma previsto; (c) a definição da trajetória de mudança estabelecendo a sucessão dos eventos e as eventuais paralizações; e (d) a elaboração de uma representação cartográfica que auxilie e evidencie os problemas existentes espacialmente, favorecendo as ações de fiscalização. Para tanto, uma análise pormenorizada deve ser feita em relação aos métodos existentes, efetuando testes e combinações de procedimentos para obter o melhor tratamento dos dados a partir de uma análise de acurácia.

3.1 PRÉ-PROCESSAMENTO

Normalmente, na detecção de mudança são descritas duas etapas de pré-processamento: (a) correção geométrica; e (b) calibração radiométrica (ALMUTAIRI; WARNER, 2010). A correção geométrica permite a justaposição dos pixels de uma mesma área no terreno e de tempos diferentes, eliminando a existência de alguma distorção sistemática. O registro geométrico incorreto danifica a precisão da detecção de mudança, gerando falsos artefatos artificiais que não correspondem às feições superficiais. Assim, as imagens devem ter uma perfeita sobreposição, possuindo uma raiz de erro médio quadrático de no máximo 0,2 pixels para alcançar um erro de apenas 10% (TOWNSHEND et al., 1992; DAI; KHORRAM, 1999).

A calibração radiométrica remove as alterações causadas por fatores externos, tais como: mudanças na calibração do sensor ao longo do tempo, ângulo solar, variabilidade na distância Terra-Sol e interferências at-



mosféricas. Esses métodos são subdivididos em dois tipos: absoluta e relativa.

A correção radiométrica absoluta utiliza códigos de transferência radiativa para obter a reflectância na superfície da Terra, eliminando as interferências atmosféricas. As principais interferências atmosféricas são provenientes de dois efeitos: (a) espalhamento (difusão ou dispersão) que muda a direção de propagação da radiação solar por interação elástica com particulados, principalmente aerossóis; e (b) absorção atmosférica, com perda efetiva de energia em comprimentos de onda específicos, principalmente provenientes de sete gases: vapor d'água (H₂O), dióxido de carbono (CO₂), ozônio (O₃), óxido nítrico (N₂O), monóxido de carbono (CO), metano (CH₄) e oxigênio (O₂) (ZULLO JÚNIOR, 1994). No entanto, a correção dos efeitos atmosféricos é mais efetiva usando sensores hiperspectrais que contem bandas exclusivas para os diferentes gases e efetua a correção para cada pixel. No caso das imagens multispectrais, são adotados valores constantes para toda a imagem.

As correções relativas normalizam os valores digitais para uma escala comum, na qual feições invariantes entre as duas imagens são ajustadas para uma única referência, assumindo que esses pixels são linearmente correlacionados (DU; TEILLET; CIHLAR, 2002). Portanto, a questão principal torna-se a obtenção das feições invariantes para efetuar a regressão linear entre as imagens temporais, podendo ser determinadas por inspeção visual ou computacionalmente. Vários métodos automatizados foram propostos para encontrar os pontos invariantes e efetuar a normalização entre as



imagens, destacando: regressão linear robusta (HEO; FITZHUGH, 2000) densidade no diagrama de frequência (SONG et al., 2001; CHEN; VIERLING; DEERING, 2005); medidas espectrais de distância e similaridades (CARVALHO JÚNIOR et al., 2013), medidas espectrais entre componentes canônicas (CANTY; NIELSEN; SCHMIDT, 2004); e zonas de não mudança envolta da linha de regressão (ELVIDGE et al., 1995) ou de componentes principais (DU et al., 2002). Com o propósito de aglutinar todas essas técnicas, Carvalho Júnior et al. (2013) propuseram um método sequencial para a determinação dos pontos invariantes composto pelos seguintes métodos: (a) medidas espectrais nas imagens temporais originais ou nas componentes canônicas; (b) densidade do diagrama de dispersão; e (c) regressão robusta. O presente método está presente no programa Abílio.

3.2 DETECÇÃO DE MUDANÇA

Diferentes métodos foram propostos para a detecção de mudança a partir de imagens de sensoriamento remoto, havendo importantes revisões bibliográficas sobre o tema (COPPIN et al., 2001; HALL; HAY, 2003; LAMBIN, 1999; LU et al., 2003; SINGH, 1989; TEWKESBURY et al., 2015).

Os algoritmos de detecção de mudança são constituídos de duas etapas de processamento: (a) classificação; e (b) detecção de mudança. Uma classificação dos métodos de detecção de mudança considera se a etapa de detecção de mudança precede ou é após a etapa de classificação, sendo denominados de (YUAN et al.,

2005): (a) pré-classificação, na qual inicialmente é gerada uma nova imagem realçando as feições de mudança e depois realiza-se a etapa de classificação; e (b) pós-classificação, na qual inicialmente realiza-se uma classificação independente para cada tempo e em seguida efetua-se a extração e quantificação das áreas de mudança a partir da tabulação cruzada entre as imagens temporais.

Diferentes métodos de pré-classificação foram propostos: (a) operações algébricas de subtração e divisão de imagens multitemporais (COPPIN et al., 2001; FRANKLIN et al., 2003; SKAKUN; WULDER; FRANKLIN, 2003); (b) análise de vetor de mudança (CARVALHO JÚNIOR et al., 2011; JOHNSON; KASISCHKE, 1998); (c) mistura espectral (ADAMS et al., 1995); e (d) diversas transformações lineares, como a análise de componentes principais (BYRNE et al., 1980; FUNG e LEDREW, 1987), análise de correspondência (CAKIR et al., 2006), análise canônica (NIELSEN; CONRADSEN; SIMPSON, 1998) e Tasseled-Cap (HEALEY et al., 2005). Os métodos de pré-classificação, apesar de serem efetivos na localização das mudanças, muitas vezes são difíceis de precisar a natureza da mudança, necessitando de outra etapa de classificação.

Dessa forma, o método de pós-classificação é o mais amplamente usado nos estudos de detecção de mudança em ambientes urbanos, por ser efetivo na descrição da magnitude, localização e natureza das alterações ocorridas (HARDIN; JACKSON; OTTERSTROM, 2007). As principais vantagens do método de pós-classificação são: (a) a independência no processo de classificação das imagens temporais compensa as variações das condi-

ções atmosféricas, mudanças fenológicas e umidade de solo; (b) o processo de atualização dos dados é simples, beneficiando o monitoramento; (c) possibilita comparar dados de sensores com diferentes tipos de resoluções; e (d) permite individualizar as diferentes categorias de mudança, não ficando restrito em apenas realçar as feições de mudanças (COPPIN et al., 2001; MENKE et al., 2009). Em contraposição, as duas principais desvantagens desse método são: (a) normalmente não é completamente automático, o que torna o método moroso; e (b) a precisão da detecção das mudanças depende da acurácia da classificação em cada tempo, o que pode facilitar a propagação de erros (YUAN et al., 2005; MENKE et al., 2009).

Conforme Silva et al. (2012), a classificação direta sobre as diversas bandas espectro-temporais não se enquadra nos métodos de pós-classificação e pré-classificação, uma vez que as etapas de classificação e de detecção de mudança são efetuadas sincronicamente. Nesse tipo de classificação são usados normalmente a análise de grupos (WEISMILLER et al., 1977) e as redes neurais artificiais (DAI; KORRAM, 1999).

4. CONCLUSÃO

No presente trabalho foram revistas as principais potencialidades e desafios para o emprego do sensoriamento remoto na fiscalização de obras dentro das atividades do TCU, focando os seguintes aspectos: (a) qualidade dos atributos da imagem; (b) etapas de processamento digital de imagens; e (c) possíveis adequações e esforços necessários para quantificar e entender as etapas das obras públicas. A fiscalização de obras públicas a partir de sensoriamento remoto é um processo complexo, com algum grau de interferências entre classes e uma forte componente de mudanças espaçotemporais. Apenas uma representação contínua no tempo dos elementos da obra permite uma análise e fiscalização precisas. Para esse propósito, medições repetitivas dos componentes espectrais e espaciais da superfície da Terra devem ser obtidas preferencialmente em alta resolução. Cada atributo fornece um tipo específico de informações sobre a obra, devendo ser combinados para obter uma descrição detalhada dos processos superficiais. Diferentes modelos de detecção de mudanças devem ser testados, considerando as condições de circunvizinhança e de ambientes.

A fiscalização e o monitoramento a partir de técnicas de geoprocessamento e sensoriamento remoto permitem monitorar diferentes obras simultaneamente, praticamente em tempo real. Essa nova abordagem necessita de um conjunto de investigações metodológicas,

a fim de reforçar a relação entre o projeto original da obra e a magnitude de mudança detectada na imagem. O avanço de modelos especialistas para a detecção de mudança de forma automatizada ou semiautomatizada das obras públicas permitirá estabelecer um sistema de alerta com priorização de vistorias em campo. Nesse propósito, deve-se ter um esforço de pesquisa computacional para o desenvolvimento de um conjunto de técnicas de detecção de padrões que permita aperfeiçoar o gerenciamento das fases da obra reduzindo as incertezas. Esse arsenal tecnológico adequado aos diferentes alvos permitirá estabelecer estratégias para coibir a ação de possíveis fraudes ou atrasos no cronograma.

O fator-chave para obter um mapeamento bem-sucedido das obras de engenharia é utilizar as imagens de alta resolução (espectral, espacial e temporal) e o MDE de forma integrada com a resposta pretendida descrita pelo projeto executivo, constando todos os fatores específicos à atividade de execução. Diferentemente de outros estudos de sensoriamento remoto em áreas urbanas, no presente caso tem-se previamente a locação espacial e as mudanças pretendidas, o que permite uma nova abordagem para o desenvolvimento de técnicas automatizadas considerando um modelo prévio.

REFERÊNCIAS

ABADE, N. A. et al. Comparative Analysis of MODIS Time-Series Classification Using Support Vector Machines and Methods Based upon Distance and Similarity Measures in the Brazilian Cerrado-Caatinga Boundary. *Remote Sensing*, [S.l.], v. 7, n. 9, p. 12160-12191, 2015.



- ADAMS, J. B. et al. Classification of Multispectral Images Based on Fractions of Endmembers: Application to Land-Cover Change in the Brazilian Amazon. *Remote Sensing of Environment*, [S.l.], v. 52, p. 137-154, 1995.
- AIRES, F. et al. Characterization and Space-Time Downscaling of the Inundation Extent over the Inner Niger Delta Using GIEMS and MODIS Data. *Journal of Hydrometeorology*, Washington, DC, v. 15, n. 1, p. 171-192, 2014.
- ALMUTAIRI, A.; WARNER, T. A. Change Detection Accuracy and Image Properties: A Study Using Simulated Data. *Remote Sensing*, [S.l.], v. 2, p. 1508-1529, 2010.
- BALTSAVIAS, E. P. Airborne Laser Scanning: Basic Relations and Formulas. *ISPRS Journal of Photogrammetry and Remote Sensing*, [S.l.], v. 54, p. 199-214, 1999.
- BAUGH, C. A. et al. SRTM Vegetation Removal and Hydrodynamic Modeling Accuracy. *Water Resource Research*, Washington, DC, v. 49, p. 5276-5289, 2013.
- BEN-DOR, E.; LEVIN, N.; SAARONI, H. A Spectral Based Recognition of the Urban Environment Using the Visible and Near-Infrared Spectral Region (0.4-1.1 μm): A Case Study Over Tel-Aviv. *International Journal of Remote Sensing*, Londres, v. 22, n. 11, p. 2193-2218, 2001.
- BLASCHKE, T. Object Based Image Analysis for Remote Sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, [S.l.], v. 65, n. 1, p. 2-16, 2010.
- BRASIL. Tribunal de Contas da União. Fiscobras 2015. 2015. Disponível em: <<http://portal.tcu.gov.br/lumis/portal/file/fileDownload.jsp?fileId=8A8182A250C885960150CD7694B146CC&inline=1>>. Acesso em: 20 out. 2016.
- BYRNE, G. F.; CRAPPER, P. F.; MAYO, K. K. Monitoring Land Cover Change by Principal Component Analysis of Multitemporal Landsat Data. *Remote Sensing of Environment*, [S.l.], v. 10, p. 175-184, 1980.
- CAKIR, H. I.; KHORRAM, S.; NELSON, S. A. C. Correspondence Analysis for Detecting Land Cover Change. *Remote Sensing of Environment*, [S.l.], v. 102, n. 3-4, p. 306-317, 2006.
- CANTY, M. J.; NIELSEN, A. A.; SCHMIDT, M. Automatic Radiometric Normalization of Multitemporal Satellite Imagery. *Remote Sensing of Environment*, [S.l.], v. 91, p. 441-451, 2004.
- CARVALHO JÚNIOR, A. O. et al. A New Approach to Change Vector Analysis Using Distance and Similarity Measures. *Remote Sensing*, [S.l.], v. 3, p. 2473-2493, 2011.
- _____. Análise de imagens hiperespectrais pelo método Multiple Endmember Spectral Mixture Analysis (MESMA) em depósito supergênico de níquel. *Brazilian Journal of Geology*, São Paulo, v. 33, n. 1, p. 63-74, 2003.
- _____. Classificação de padrões de savana usando assinaturas temporais NDVI do sensor MODIS no Parque Nacional Chapada dos Veadeiros. *Revista Brasileira de Geofísica*, [S.l.], v. 26, n. 4, p. 505-517, 2008.
- _____. Standardized Time-Series And Interannual Phenological Deviation: New Techniques for Burned-Area Detection Using Long-Term MODIS-NBR Dataset. *Remote Sensing*, [S.l.], v. 7, n. 6, p. 6950-6985, 2015.
- _____. Radiometric Normalization of Temporal Images Combining Automatic Detection of Pseudo-Invariant Features from the Distance and Similarity Spectral Measures, Density Scatterplot Analysis, and Robust Regression. *Remote Sensing*, [S.l.], v. 5, n. 6, p. 2763-2794, 2013.
- CAVALLI, R. M. et al. Hyperspectral Sensor Data Capability for Retrieving Complex Urban Land Cover in Comparison with Multispectral Data: Venice City Case Study (Italy). *Sensors*, [S.l.], v. 8, n. 5, p. 3299-3320, 2008.
- CHEN, X.; VIERLING, L.; DEERING, D. A Simple and Effective Radiometric Correction Method to Improve Landscape Change Detection across Sensors and across Time. *Remote Sensing of Environment*, [S.l.], v. 98, p. 63-79, 2005.
- CLARK, R. N. et al. USGS Digital Spectral Library Splib06a. US Geological Survey, Digital Data Series, [S.l.], n. 231, 2007.
- COPPIN, P. et al. Operational Monitoring of Green Biomass Change for Forest Management. *Photogrammetric Engineering and Remote Sensing*, Bethesda, v. 67, p. 603-611, 2001.
- COPPIN, P. R.; BAUER, M. E. Digital Change Detection in Forest Ecosystems with Remote Sensing Imagery. *Remote Sensing Reviews*, Londres, v. 13, n. 3-4, p. 207-234, 1996.
- COUTO JÚNIOR, A. F.; CARVALHO JÚNIOR, O. A.; MARTINS, E. S. Séries temporais MODIS aplicadas em sucessão de culturas de soja (*Glycine max* (L.) Merrill) e milho (*Zeamays* L.) em sistema de plantio direto. *Revista Brasileira de Cartografia*, Rio de Janeiro, v. 64, n. 3, p. 405-418, 2012.

- COWEN, D.J. et al. The Design and Implementation of an Integrated Gis For Environmental Applications. *Photogrammetric Engineering and Remote Sensing*, Bethesda, v. 61, p. 1393-1404, 1995.
- DAI, X. L.; KHORRAM, S. Remotely Sensed Change Detection Based on Artificial Neural Networks. *Photogrammetric Engineering and Remote Sensing*, Bethesda, v. 65, n. 10, p. 1187-1194, 1999.
- _____. The Effects of Image Misregistration on the Accuracy of Remotely Sensed Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, [S.l.], v. 36, p. 1566-1577, 1998.
- CARVALHO JÚNIOR, O. A., HERMUCHE, P. M., GUIMARÃES, R. F. Identificação regional da floresta estacional decidual na bacia do Rio Paranã a partir da análise multitemporal de imagens MODIS. *Revista Brasileira de Geofísica*, [S.l.], v. 24, n. 3, p. 319-332, 2006.
- DEMARCHI, L. et al. Multiple Endmember Unmixing of CHRIS/Proba Imagery for Mapping impervious Surfaces in Urban and Suburban Environments. *IEEE Transactions on Geoscience and Remote Sensing*, [S.l.], v. 50, n. 9, p. 3409-3424, 2012.
- DORREN, L.; MAIER, B.; SEIJMONSBERGEN, A. Improved Landsat-Based Forest Mapping in Steep Mountainous Terrain Using Object-Based Classification. *Forest Ecology and Management*, [S.l.], v. 183, n. 1-3, p. 31-46, 2003.
- DU, Y.; TEILLET, P. M.; CIHLAR, J. Radiometric Normalization of Multitemporal High-Resolution Satellite Images with Quality Control for Land Cover Change Detection. *Remote Sensing of Environment*, [S.l.], v. 82, p. 123-134, 2002.
- DURUPT, M.; TAILLANDIER, F. Automatic Building Reconstruction from a Digital Elevation Model and Cadastral Data: An Operational Approach. 2006. Disponível em: <http://www.isprs.org/proceedings/XXXVI/part3/singlepapers/O_14.pdf>. Acesso em: 25 nov. 2016.
- ELVIDGE, C. D. et al. Relative Radiometric Normalization of Landsat Multispectral Scanner (MSS) Data Using an Automatic Scattergram-Controlled Regression. *Photogrammetric Engineering and Remote Sensing*, Bethesda, v. 61, p. 1255-1260, 1995.
- FRANKE, J. et al. Hierarchical Multiple Endmember Spectral Mixture Analysis (MESMA) of Hyperspectral Imagery for Urban Environments. *Remote Sensing of Environment*, [S.l.], v. 113, n. 8, p. 1712-1723, 2009.
- FRANKLIN, S. E. et al. Mountain Pine Beetle Red-Attack Forest Damage Classification Using Stratified Landsat TM Data in British Columbia, Canada. *Photogrammetric Engineering and Remote Sensing*, Bethesda, v. 69, n. 3 p. 283-288, 2003.
- FUNG, T.; LE DREW, E. Application of Principal Components Analysis to Change Detection. *Photogrammetric Engineering and Remote Sensing*, Bethesda, v. 53, n. 12, p. 1649-1658, 1987.
- GREEN, R. O. et al. Imaging Spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). *Remote Sensing of Environment*, [S.l.], v. 65, p. 227-248, 1998.
- HALL, O.; HAY, G. J. A Multiscale Object-Specific Approach to Digital Change Detection. *International Journal of Applied Earth Observation and Geoinformation*, [S.l.], v. 4, p. 311-327, 2003.
- HARDIN, P. J.; JACKSON, M. W.; OTTERSTROM, S. M. Mapping, Measuring, and Modeling Urban Growth. In: JENSEN, R. R.; GATRELL J. D.; MCLEAN D. (Org.). *Geo-Spatial Technologies in Urban Environments: Policy, Practice and Pixels*. 2. ed. Heidelberg: Springer-Verlag, 2007. p. 141-176.
- HEALEY, S. P. et al. Comparison of Tasseled Cap-based Landsat Data Structures for Use in Forest Disturbance Detection. *Remote Sensing of Environment*, [S.l.], v. 97, n. 3, p. 301-310, 2005.
- HEGAZY, I. R.; KALOOP, M. R. Monitoring Urban Growth and Land Use Change Detection with GIS and Remote Sensing Techniques in Daqahlia Governorate Egypt. *International Journal of Sustainable Built Environment*, [S.l.], v. 4, p. 117-124, 2015.
- HEIDEN, U.; SEGL, K.; KAUFMANN, H. Determination of Robust Spectral Features for Identification of Urban Surface Materials in Hyperspectral Remote Sensing Data. *Remote Sensing of Environment*, [S.l.], v. 111, n. 4, p. 537-552, 2007.
- HEO, J.; FITZHUGH, T. W. A Standardized Radiometric Normalization Method for Change Detection Using Remotely Sensed Imagery. *Photogrammetric Engineering and Remote Sensing*, Bethesda, v. 66, n. 2, p. 173-181, 2000.
- HEROLD, M. et al. Spectrometry for Urban Area Remote Sensing – Development and Analysis of a Spectral Library from 350 to 2400 nm. *Remote Sensing of Environment*, [S.l.], v. 91, n. 3-4, p. 304-319, 2004.
- HOBBI, M. L.; GINZLER, C. Accuracy Assessment of Digital Surface Models Based on WorldView-2 and ADS80 Stereo Remote Sensing Data. *Sensors*, [S.l.], v. 12, p. 6347-6368, 2012.

- JARIHANI, A. A. et al. Satellite-Derived Digital Elevation Model (DEM) Selection, Preparation and Correction for Hydrodynamic Modelling in Large, Low-Gradient and Data-Sparse Catchments. *Journal of Hydrology, [S.l.]*, v. 524, p. 489-506, 2015.
- JAT, M. K.; GARG, P. K.; KHARE, D. Monitoring and Modelling of Urban Sprawl Using Remote Sensing and GIS Techniques. *International Journal of Applied Earth Observation and Geoinformation, [S.l.]*, v. 10, n. 1, p. 26-43, 2008.
- JENSEN, J. R. et al. An Evaluation of Coastwatch Change Detection Protocol in South Carolina. *Photogrammetric Engineering and Remote Sensing, Bethesda*, v. 59, n. 4, p. 519-525, 1993.
- JENSEN, J. R.; COWEN, D. C. Remote Sensing of Urban/Suburban Infrastructure and Socio-Economic Attributes. *Photogrammetric Engineering and Remote Sensing, Bethesda*, v. 65, p. 611-622, 1999.
- JOHNSON, R. D.; KASISCHKE, E. S. Change Vector Analysis: A Technique for the Multispectral Monitoring of Land Cover and Condition. *International Journal of Remote Sensing, Londres*, v. 19, p. 3, n. 411-426, 1998.
- KIM, M.; MADDEN, M.; XU, B. GEOBIA Vegetation Mapping in Great Smoky Mountains National Park with Spectral and Non-Spectral Ancillary Information. *Photogrammetric Engineering and Remote Sensing, Bethesda*, v. 76, n. 2, p. 137-149, 2010.
- KIM, M.; MADDEN, M.; WARNER, T. Forest Type Mapping Using Object-Specific Texture Measures from Multispectral IKONOS Imagery: Segmentation Quality and Image Classification Issues. *Photogrammetric Engineering and Remote Sensing, Bethesda*, v. 75, n. 7, p. 819-830, 2009.
- KIM, Z. W.; NEVATIA, R. Automatic Description of Complex Buildings from Multiple Images. *Computer Vision Image Understanding*, v. 96, p. 60-95, 2004.
- LAMBIN, E. F. Monitoring Forest Degradation in Tropical Regions by Remote Sensing: Some Methodological Issues. *Global Ecology and Biogeography, [S.l.]*, v. 8, n. 3-4, p. 191-198, 1999.
- LIU, D.; XIA, F. Assessing Object-Based Classification: Advantages and Limitations. *Remote Sensing Letters, Melbourne*, v. 1, n. 4, p. 187-194, 2010.
- LU, D. et al. Change Detection Techniques. *International Journal of Remote Sensing, Londres*, v. 25, n. 12, p. 2365-2407, 2003.
- MENG, X.; CURRIT, N.; ZHAO, K. Ground Filtering Algorithms for Airborne LiDAR Data: A Review of Critical Issues. *Remote Sensing, [S.l.]*, v. 2, p. 833-860, 2010.
- MENKE, A. B. et al. Análise das mudanças do uso agrícola da terra a partir de dados de sensoriamento remoto multitemporal no município de Luis Eduardo Magalhães (BA-Brasil). *Sociedade & Natureza, Uberlândia*, v. 21, n. 3, p. 315-326, 2009.
- MIRANDA, A. C. O.; MATOS, C. R. Potencial uso do BIM na fiscalização de obras públicas. *Revista do Tribunal de Contas da União, Brasília, DF*, v. 133, p. 22-31, 2015.
- MUSA, Z. N.; POPESCU, I.; MYNETT, A. A Review of Applications of Satellite SAR, Optical, Altimetry and DEM Data for Surface Water Modelling, Mapping and Parameter Estimation. *Hydrology and Earth System Sciences, [S.l.]*, v. 19, p. 3755-3769, 2015.
- MYINT, S. et al. Per-Pixel vs. Object-Based Classification of Urban Land Cover Extraction Using High Spatial Resolution Imagery. *Remote Sensing of Environment, [S.l.]*, v. 115, n. 5, p. 1145-1161, 2011.
- NIELSEN, A. A.; CONRADSEN, K.; SIMPSON, J. J. Multivariate Alteration Detection (MAD) and MAF Postprocessing in Multispectral, Bitemporal Image Data: New Approaches to Change Detection Studies. *Remote Sensing of Environment, [S.l.]*, v. 64, p. 1-19, 1998.
- OLIVEIRA, S. N. et al. Detecção de mudança do uso e cobertura da terra usando o método de pós-classificação na fronteira agrícola do oeste da Bahia sobre o Grupo Urucua durante o período 1988-2011. *Revista Brasileira de Cartografia, Rio de Janeiro*, v. 66, n. 5, p. 1157-1176, 2014.
- PHINN, S. R. et al. Monitoring the Composition of Urban Environments Based on the Vegetation-Impervious Surface-Soil (VIS) Model by Subpixel Analysis Techniques. *International Journal of Remote Sensing, Londres*, v. 23, n. 20, p. 4131-4153, 2002.
- POLI, D.; TOUTIN, T. Review of Developments in Geometric Modelling for High-Resolution Satellite Pushbroom Sensors. *Photogrammetric Record, [S.l.]*, v. 27, p. 58-73, 2012.
- PUENTE, I. et al. Review of Mobile Mapping and Surveying Technologies. *Measurement: Journal of the International Measurement Confederation, [S.l.]*, v. 46, n. 7, p. 2127-2145, 2013.

- RAO, N. R.; GARG, P. K.; GHOSH, S. K. Development of an Agricultural Crops Spectral Library and Classification of Crops at Cultivar Level Using Hyperspectral Data. *Precision Agriculture*, [S.l.], v. 8, n. 4-5, p. 173-185, 2007.
- SAKAMOTO, T. et al. A Crop Phenology Detection Method Using Time-Series MODIS Data. *Remote sensing of environment*, [S.l.], v. 96, n. 3, p. 366-374, 2005.
- SINGH, A. Digital Change Detection Techniques Using Remotely-Sensed Data. *International Journal of Remote Sensing*, Londres, v. 10, p. 989-1003, 1989.
- SILVA, N. C. et al. Change Detection Software Using Self-Organizing Feature Maps. *Revista Brasileira de Geofísica*, [S.l.], v. 30, n. 4, p. 505-518, 2012.
- SKAKUN, R. S.; WULDER, M. A.; FRANKLIN, S. E. Sensitivity of the Thematic Mapper Enhanced Wetness Difference Index to Detect Mountain Pine Beetle Red-Attack Damage. *Remote Sensing of Environment*, [S.l.], v. 86, p. 433-443, 2003.
- SMALL, C. High Resolution Spectral Mixture Analysis of Urban Reflectance. *Remote Sensing of Environment*, [S.l.], v. 88, p. 170-186, 2003.
- SONG, C. et al. Classification and Change Detection Using Landsat TM Data: When and How to Correct Atmospheric Effects? *Remote Sensing of Environment*, [S.l.], v. 75, p. 230-244, 2001.
- SOUZA, I. V. N.; BATISTA, H. M. Estudo dos benefícios econômicos gerados pelas fiscalizações de obras públicas, realizadas pelo Tribunal de Contas da União, no estado do Rio Grande do Norte, no período de 2011 e 2012. 2013. 67 f. (Trabalho de Conclusão de Curso) – Departamento de Ciências Contábeis da Universidade Federal do Rio Grande do Norte, Natal, 2013.
- SUN, C. et al. Quantifying Different Types of Urban Growth and the Change Dynamic in Guangzhou Using Multi-Temporal Remote Sensing Data. *International Journal of Applied Earth Observation and Geoinformation*, [S.l.], v. 21, p. 409-417, 2013.
- TEWKESBURY, A. P. et al. A Critical Synthesis of Remotely Sensed Optical Image Change Detection Techniques. *Remote Sensing of Environment*, [S.l.], v. 160, p. 1-14, 2015.
- THOMAS, N.; HENDRIX, C.; CONGALTON, R. G. A Comparison of Urban Mapping Methods Using High-Resolution Digital Imagery. *Photogrammetric Engineering and Remote Sensing*, Bethesda, v. 69, n. 9, p. 963-972, 2003.
- TOWNSHEND, J. R. G. et al. The Impact of Misregistration on Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, [S.l.], v. 30, n. 5, p. 1054-1060, 1992.
- TRIAS-SANZ, R., STAMON, G., LOUCHET, J. Using Colour, Texture, and Hierarchical Segmentation for High-Resolution Remote Sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, [S.l.], v. 63, n. 2, p. 156-168, 2008.
- VITAL, A. L. F. et al. Fiscobras: uma obra em construção. *Revista do Tribunal de Contas da União*, Brasília, DF, v. 133, p. 32-39, 2015.
- WEISMILLER, R. A. et al. Change Detection in Coastal Zone Environments. *Photogrammetric Engineering and Remote Sensing*, Bethesda, v. 43, p. 1533-1539, 1977.
- WENTZ, E. A. et al. Supporting Global Environmental Change Research: A Review of Trends and Knowledge Gaps in Urban Remote Sensing. *Remote Sensing*, [S.l.], v. 6, n. 5, p. 3879-3905, 2014.
- WU, C.; MURRAY, T. A. Estimating Impervious Surface Distribution by Spectral Mixture Analysis. *Remote Sensing of Environment*, [S.l.], v. 84, n. 4, p. 493-505, 2003.
- YADAV, P. K.; KAPOOR, M.; SARMA, K. Land Use Land Cover Mapping, Change Detection and Conflict Analysis of Nagzira-Navegaon Corridor, Central India Using Geospatial Technology. *International Journal of Remote Sensing and GIS*, Delhi, v. 1, n. 2, p. 90-98, 2012.
- YU, Q. et al. Object-Based Detailed Vegetation Classification with Airborne High Spatial Resolution Remote Sensing Imagery. *Photogrammetric Engineering and Remote Sensing*, Bethesda, v. 72, n. 7, p. 799-811, 2006.
- YUAN, F. et al. Land Cover Classification and Change Analysis of the Twin Cities (Minnesota) Metropolitan Area by Multitemporal Landsat Remote Sensing. *Remote Sensing of Environment*, [S.l.], v. 98, n. 2, p. 317-328, 2005.
- ZOMER, R. J.; TRABUCCO, A.; USTIN, S. L. Building Spectral Libraries for Wetlands Land Cover Classification and Hyperspectral Remote Sensing. *Journal of Environmental Management*, [S.l.], v. 90, n. 7, p. 2170-2177, 2009.
- ZULLO JÚNIOR, J. Correção atmosférica de imagens de satélite e aplicações. 1994. 191 p. Tese (Doutorado em Engenharia Elétrica) – Universidade Estadual de Campinas, Campinas, 1994.

Implementação de um geocatálogo para auxílio na localização e recuperação de dados geográficos abertos



Drausio Gomes dos Santos

é servidor do Tribunal de Contas da União, mestre em Computação Aplicada pela Universidade de Brasília, especialista em Gestão da Tecnologia da Informação pelo Cin/UFPE, tecnólogo em Processamento de Dados pela FATEC/SP e bacharel em Geografia pela USP.



Alexandre Zaghetto

é mestre e doutor em Engenharia Elétrica pela Universidade de Brasília (UnB), possui graduação em Engenharia Elétrica, com ênfase em Eletrônica, pela Universidade Federal do Rio de Janeiro, além de professor adjunto no Dept. de Ciência da Computação da UnB e líder do Grupo de Pesquisa em Sistemas Biométricos (BiTGroup).

RESUMO

Este artigo trata da construção de um geocatálogo que utiliza a semântica e ontologias para auxílio no processo de descoberta de recursos geográficos em serviços de uma infraestrutura aberta de dados espaciais. Sua arquitetura incorpora um banco de dados não relacional orientado a grafos e é passível de implementação em ambiente de nuvem. Para sua concepção foi utilizada a metodologia de desenvolvimento de sistemas SERVUS, especializada em arquitetura orientada a serviços geoespaciais.

Palavras-chave: Engenharia de Software. Geocatálogo. SERVUS. IDE. Semântica. Dados abertos.

1. INTRODUÇÃO

Tem sido um desafio para os órgãos da administração pública obter uma visão espacial do resultado das ações governamentais de forma simplificada e atualizada, uma vez que os dados são descentralizados, produzidos por fontes variadas e em momentos diferentes. O governo tem tomado algumas medidas na direção da padronização, otimização de recursos e integração de dados, sendo uma delas a criação da Infraestrutura Nacional de Dados Espaciais do Brasil (INDE-BR).



A INDE-BR consiste em uma iniciativa de criação de uma arquitetura para disseminação e regramento do uso de dados geográficos no âmbito da administração pública. Cabe ressaltar um dos objetivos definidos na sua concepção, o de

“evitar a duplicidade de ações e o desperdício de recursos na obtenção de dados aeroespaciais pelos órgãos da administração pública, por meio da divulgação dos metadados relativos a esses dados disponíveis nas entidades e nos órgãos públicos das esferas federal, estadual, distrital e municipal” (BRASIL, 2008).

Camboim (2013) ressalta a iniciativa governamental de criação da Infraestrutura Nacional de Dados Abertos (INDA) e do Portal Brasileiro de Dados Abertos como uma estratégia de adoção de Dados Interligados Abertos (*Linked Open Data*). Esse autor propõe uma arquitetura capaz de tornar os dados da INDE-BR disponíveis na INDA de forma integrada. Essa proposta de arquitetura destaca-se pelo uso de camadas semânticas e ontologias, fazendo uma junção de técnicas da web semântica com dados geoespaciais, ingressando no campo da geosseântica.

As arquiteturas das IDE atuais são, via de regra, orientadas a serviço e adotam os padrões tecnológicos estabelecidos pela *Open Geospatial*

Consortium (OGC). Alguns desses padrões definem o funcionamento dos geosserviços, que permitem a obtenção de mapas cartográficos, mapas temáticos, dados geográficos e metadados.

Um dos componentes essenciais numa arquitetura de IDE é o catálogo de serviços. A metodologia SERVUS de desenvolvimento de sistemas (USLÄNDER, 2010), feita especialmente para tratar do uso e composição de serviços geoespaciais, propõe a utilização de um geocatálogo semântico para auxílio na localização de recursos geográficos em rede.

Para estudar esse aspecto, este documento foi dividido da seguinte forma: na Seção 2 são apresentados alguns trabalhos correlatos; na Seção 3 é apresentada a fundamentação teórica, tratando dos temas IDE e metodologia SERVUS; na Seção 4 é apresentado o modelo proposto, detalhes do geocatálogo e um exemplo de utilização com a metodologia SERVUS e, por fim, na Seção 5 temos as conclusões.

2. TRABALHOS CORRELATOS

Andrade e Baptista (2011) trilharam um caminho que incluía busca semântica, listagem de grande quantidade de informações oriundas de fontes diversas, estabelecimento de métricas e geração de um ranking de recursos disponibilizados numa

infraestrutura de dados espaciais. Os autores partiram do pressuposto de que o uso de ontologias seria o caminho para tornar mais precisas as buscas e facilitar o processo de automatização. Como contribuições, tivemos o estabelecimento de uma métrica combinada e a construção de uma rede semântica.

Daltio e Carvalho (2012) propuseram a construção de um *framework* para a recuperação semântica de dados espaciais. O *framework* era baseado em um processo de anotação semântica dos recursos geográficos e na utilização de um serviço de gerenciamento de ontologias. Cabe ressaltar neste trabalho a importância dada ao processo de seleção das ontologias.

Gimenez, Tanaka e Baião (2013) apresentam uma proposta de integração semântica para a INDE-BR usando geo-ontologias de domínio incorporadas em camadas semânticas atuando sobre as IDEs.

3. FUNDAMENTAÇÃO TEÓRICA

3.1 INFRAESTRUTURA DE DADOS ESPACIAIS (IDE)

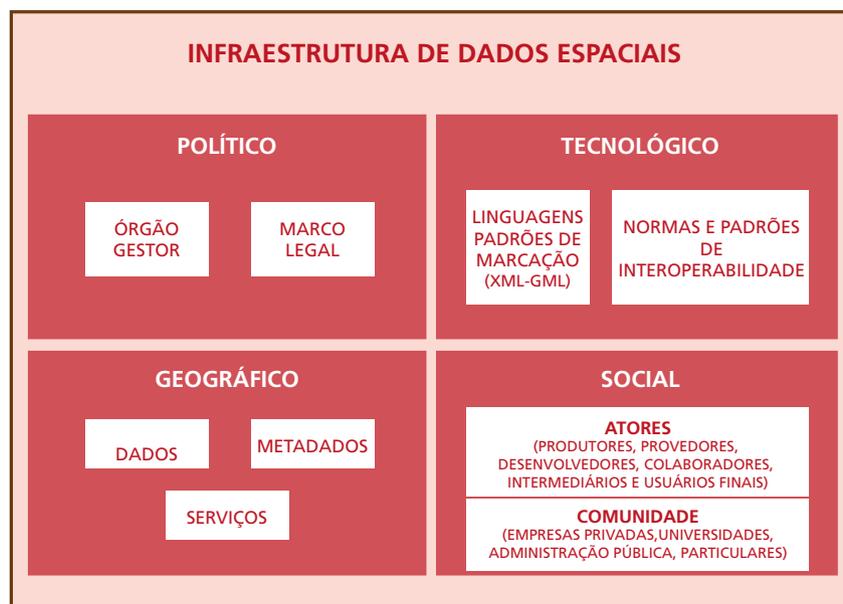
Bernabé-Poveda e López Vázquez (2012, p. 57-59) conceituam Infraestrutura de Dados Espaciais (IDE) como uma infraestrutura necessária para o acesso, compartilhamento, troca, combinação e

análise dos dados geográficos, de uma forma padrão e interoperável. Também consideram a necessidade de que esses dados estejam disponíveis em rede, por meio de um conjunto de sistemas que utilizem protocolos e interfaces padrões, propiciando a criação de aplicações que possam ser vistas pelo usuário como um único sistema. IDE pode ser classificada como uma estrutura de componentes tecnológicos, geográficos, sociais e políticos conforme ilustrado na Figura 1.

O componente tecnológico é representado por uma arquitetura baseada em padrões de interoperabilidade e capaz de compartilhar dados e informações geográficas. As linguagens de marcação XML e GML têm um papel fundamental nesse componente. O componente social é representado por um conjunto de atores, dentre eles, produtores de dados, provedores de serviços, usuários, desenvolvedores de software e responsáveis por padrões e normas, além de uma grande comunidade formada por empresas privadas, governo, universidades e a sociedade em geral. O componente geográfico é representado pelos dados, seus metadados e os geosserviços. O componente político é representado pelas pessoas e pelo órgão responsável por estabelecer o marco regulatório e as regras de funcionamento.

Figura 1:

Componentes de uma IDE



OGC *Web Services Common Standard* é um padrão tecnológico que propõe uma interface comum para um conjunto de geosserviços: *Web Map Service* (WMS), *Web Feature Service* (WFS) e *Web Coverage Service* (WCS). Nessa proposta busca-se padronizar as formas de acesso, operações, parâmetros obrigatórios e opcionais, além de estruturas de dados, a fim de diminuir os esforços de interoperabilidade. Uma das operações definidas é a *GetCapabilities* que recupera os metadados referentes às capacidades dos geosserviços (GREENWOOD; WHITE-SIDE, 2010).

WMS consiste num serviço que produz mapas dinâmicos, em que as informações geográficas são organizadas em camadas (*layers*). O serviço WFS trabalha com entidades geográficas com dados discretos ou vetoriais em formato GML, representando atributos e geometrias. WCS é o serviço que dá suporte a recuperação de dados espaciais como coberturas.

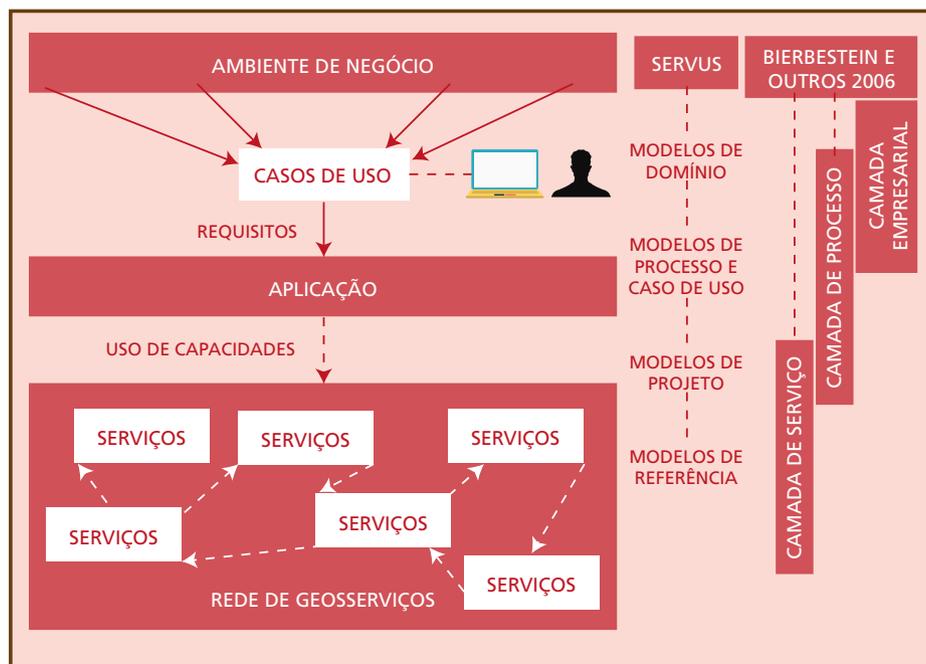
Um catálogo de serviço permite a localização de dados ou serviços geográficos através de uma gama de operações, dentre elas a operação *GetRecords* que recupera um conjunto de registros de metadados (OGC, 2016).

3.2 METODOLOGIA DE DESENVOLVIMENTO SERVUS

A sigla SERVUS é oriunda de dois termos da engenharia de software: “serviço” (*SERVice*) e “caso de uso” (*USE case*). O problema central que a metodologia se propõe a resolver é –considerando-se um conjunto de casos de uso que atendam aos requisitos de negócio e demandem capacidade de recursos em uma rede de serviços geoespaciais – como descobrir e associar recursos requisitados a recursos ofertados, compondo uma aplicação que supra os requisitos de qualidade, funcionais e informacionais. Dois outros elementos compõem esse cenário: descrições semânticas em uma rede de recursos semânticos e condições inerentes (*side conditions*) ao desenvolvimento de sistemas ambientais (processos de descoberta de recursos, uso de padrões OGC, combinações de recursos) (USLÄNDER, 2010).

Segundo Usländer (2010, p. 90), “SERVUS compreende o projeto de sistemas de informação ambientais como uma atividade iterativa de descoberta e associação: capacidades disponíveis são descobertas e associadas a requisitos do usuário formulados como casos de uso”.

Figura 2:
Hierarquia de modelos SERVUS



Fonte: Adaptado de Usländer (2010, p. 92)

A metodologia é composta por uma linguagem de modelagem, um processo de desenvolvimento e uma arquitetura de referência. Durante o processo de desenvolvimento são criados artefatos tais como: um modelo de domínio, um modelo de processos/casos de uso e um modelo de projeto. O modelo de projeto é o principal deles e expressa os requisitos e as capacidades sob a forma de recursos. Os modelos gerados são baseados nas camadas de abstração propostas por Bieberstein (2006). A Figura 2 exibe o mapeamento entre requisitos e capacidades, os modelos produzidos e a relação com as camadas de abstração.

O modelo de domínio da metodologia SERVUS representa o domínio temático do problema a ser resolvido. Ele formalmente define a parte do mundo que compreende o universo de discurso entre o usuário e o projetista do sistema, isto é, compreende o conhecimento compartilhado acerca do domínio da aplicação. Tipicamente esses conhecimentos compartilhados são representados pela especificação de uma ontologia (USLÄNDER, 2010).

O modelo de referência é composto por um *framework* arquitetural, responsável pelas orientações e regras de como especificar o sistema, e por um modelo conceitual. O modelo conceitual de SERVUS é um metamodelo segundo os preceitos de *Model Driven Architecture* (MDA) e *Meta-Object*

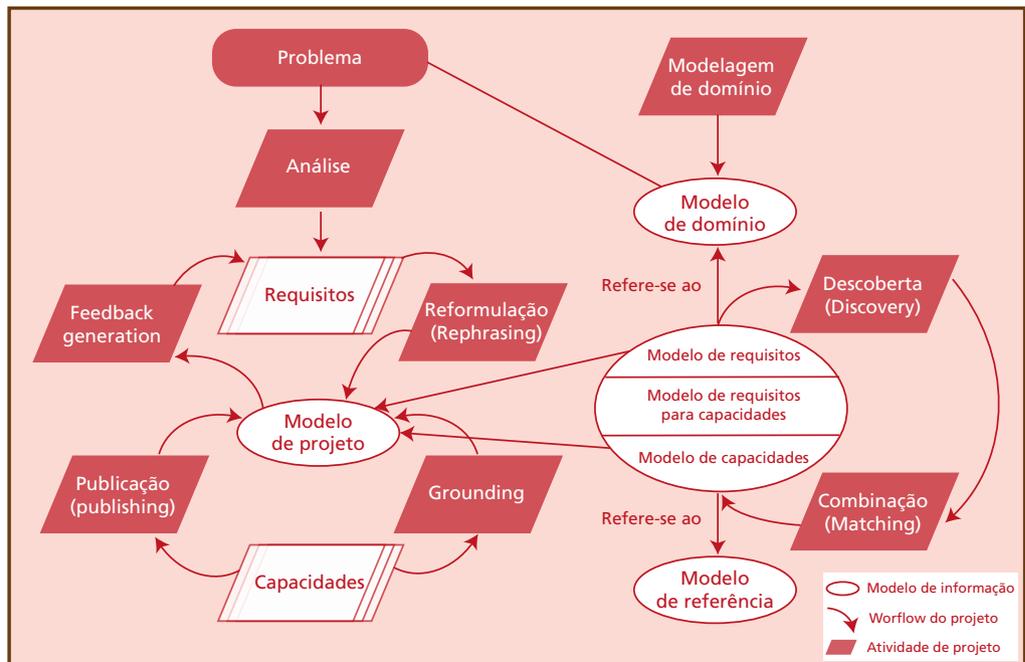
Facility (MOF). As principais metaclasses são: feição (*feature*), *interface*, serviço e recurso. O modelo conceitual é composto por três subconjuntos de metaclasses, o primeiro está associado à visão de serviço, o segundo à visão de informação e o terceiro à visão de recurso. O modelo de projeto pode ser visto como uma composição de três modelos: o de requisitos (REQ's), o de capacidades (CAP's) e o de mapeamento entre os dois primeiros (REQ2CAP). A Figura 3 apresenta um quadro geral dos modelos e atividades da etapa de projeto.

Segue a seguir uma descrição das principais atividades do processo de desenvolvimento.

a. Publishing: atividade, manual ou automatizada, de busca de capacidades na rede de serviços geoespaciais e de incorporação ao modelo de capacidades. As capacidades são traduzidas para recursos ofertados, ou seja, disponíveis na rede de geosserviços;

b. Rephrasing: atividade responsável por traduzir os requisitos, de forma a extrair um conjunto de recursos necessários para realização dos casos de uso, os recursos podem ser associados aos conceitos de ontologia e em seguida carregados na rede de recursos semânticos, compondo o modelo de requisitos;

Figura 3: Modelos e atividades de projeto



Fonte: Adaptado de Usländer (2010, p. 99)

- c. Discovery:** atividade responsável por selecionar, no rol de capacidades levantadas na fase publishing, um conjunto que atenda aos recursos solicitados levantados na fase de rephrasing, para cada recurso requisitado pode existir um conjunto de recursos ofertados que atendam às necessidades;
- d. Matching:** atividade responsável por associar requisitos a capacidades, ou seja, recursos requisitados a recursos ofertados, avalia dentre as capacidades candidatas, elencadas na etapa de discovery, as mais adequadas aos requisitos. O resultado final dessa etapa é o modelo de mapeamento requisitos/capacidades;
- e. Grounding:** atividade que disponibiliza uma nova capacidade na rede de serviços geoespaciais.

SERVUS propõe a criação de um catálogo semântico que compõe uma arquitetura de implementação e um ambiente de projeto. O papel do catálogo semântico é o de se comunicar com os geosserviços da rede e processar as consultas por recursos. Funciona como uma extensão semântica que possibilita consultas baseadas em ontologias e a avaliação da proximidade semântica dos resultados obtidos. No

que se refere ao suporte à metodologia, o catálogo semântico responde pelas atividades de *harvesting*, que consiste na coleta de metadados dos serviços, e de *publishing*, publicação dos recursos disponíveis para uma rede de recursos semânticos.

6. MODELO PROPOSTO

O trabalho foi desenvolvido conforme as seguintes etapas: revisão de literatura, com a realização de estudos sobre IDEs, metodologia de desenvolvimento SOA, metodologia SERVUS, banco de dados orientado a grafos, geocatálogo e busca semântica; levantamento de requisitos baseado nos geocatálogos estudados, principalmente no proposto pela metodologia SERVUS, nas necessidades dos *stakeholders* da área de meio ambiente e na infraestrutura tecnológica atual do órgão; definição da arquitetura de alto nível, considerando-se, dentre outros aspectos, fontes de dados, arquiteturas de IDEs e geosserviços da INDE-BR; desenvolvimento do protótipo, incluindo a definição das linguagens de desenvolvimentos, dos componentes, das bibliotecas de software e do ambiente de desenvolvimento; implantação com definição do ambiente e instalação do protótipo; elaboração de um cenário de utilização do geocatálogo típico de auditoria em meio ambiente, validação do protótipo e, por fim, avaliação do resultados (SANTOS, 2016).

Figura 4:
Arquitetura de implementação

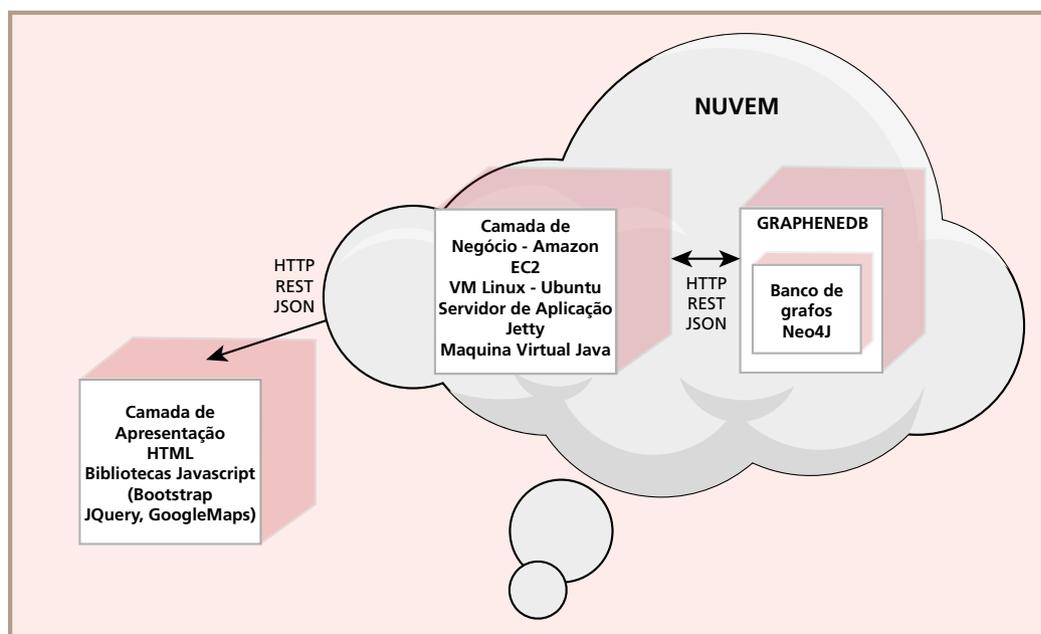
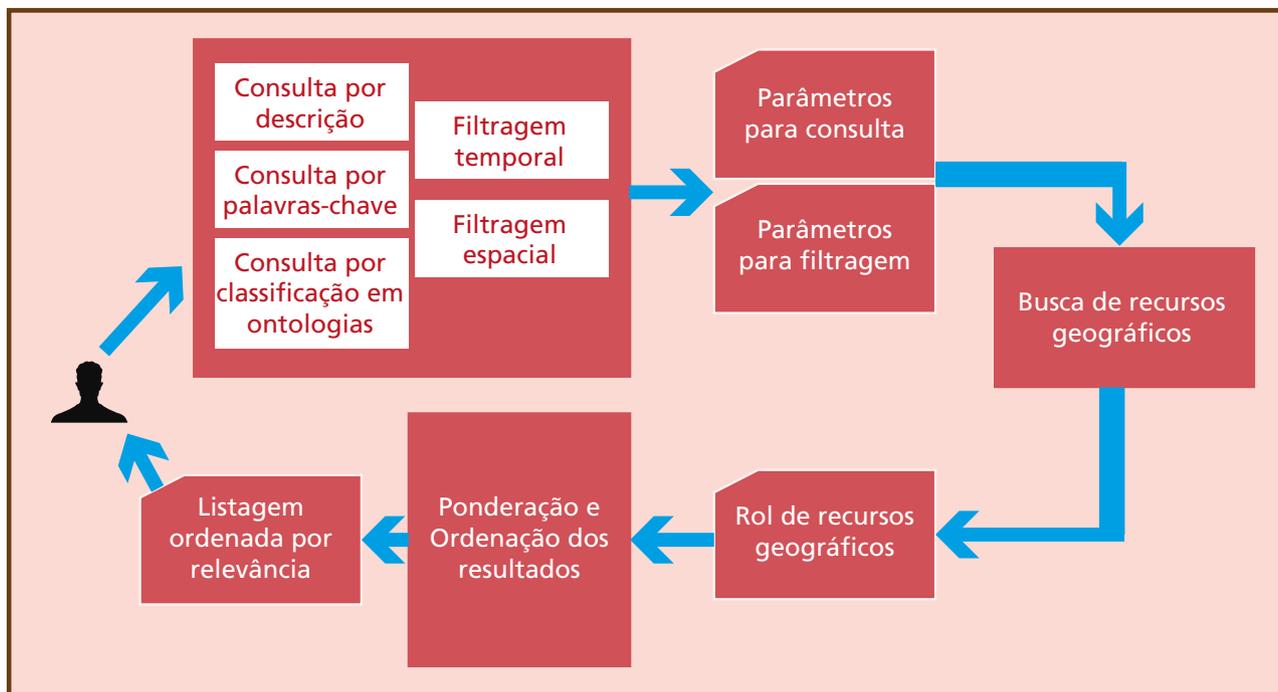


Figura 5:

Modelo de busca do catálogo



A arquitetura escolhida para implementação do protótipo do geocatálogo foi a de uma aplicação web de três camadas. A camada de apresentação foi desenvolvida em HTML5 utilizando bibliotecas *JavaScript: Bootstrap 3.3.5*, *jQuery 2.1.0*, *Google Maps 3.0* e o componente *bootstrap-slider.js*. A camada de negócio foi implementada em Java versão 1.7 e a aplicação empacotada no formato de *web archive – WAR*. As requisições da camada de apresentação são feitas pelo protocolo *http* e acessam um servidor padrão REST (*Representa-*

tional State Transfer) que retorna documentos no formato JSON (*JavaScript Object Notation*). A camada de negócio foi hospedada num serviço de nuvem EC2 da *Amazon*. A camada de persistência foi implementada através de um banco de dados NoSQL de grafos, o Neo4J 2.3.1. A hospedagem da camada de persistência foi realizada na página web GrapheneDB, que implementa os bancos Neo4J num ambiente de nuvem. A comunicação entre camada de negócio e camada de persistência também ocorre por meio de interfaces REST e no

Tabela 1:

Cálculo do índice de relevância

Termos de pesquisa	Campo com termo localizado	Peso	Peso Máximo
Descrição: "vegetação" - (peso 8)	Nome - (peso 8)	64	216
Ontologia VCE (termo raiz): "vegetação" - (peso 6)	Descrição - (peso 6)	36	162
Palavra-chave: "agrícola" - (peso 5)	Resumo - (peso 5)	25	135
	Peso total	125	513
	Índice de relevância calculado (125/513)	0,24	

formato JSON. A Figura 4 representa a arquitetura implementada.

As consultas podem ser realizadas por descrição (texto livre), palavras-chave e por termos das ontologias. Para cada um dos itens anteriores as frases ou termos são divididos em *tokens* que serão os termos utilizados na busca. Podem ser realizadas filtragens por coordenadas geográficas (filtro espacial) e por ano (temporal). Para cada par ordenado, composto pelo termo e o campo pesquisado, é atribuído um peso que será a base para o procedi-

mento de geração do ranking. A Figura 5 apresenta o modelo de busca do geocatálogo.

O resultado efetivo de uma busca depende da qualidade dos metadados e do valor de negócio atribuído a cada campo. Para lidar com esse aspecto foi elaborado um esquema de ponderação dos campos de pesquisa e dos campos de metadados e o estabelecimento de um índice de relevância que serve de base para o processo de geração do ranking de resultados. A Tabela 1 apresenta um exemplo de cálculo do índice de relevância.

Figura 6:
Tela de busca do geocatálogo

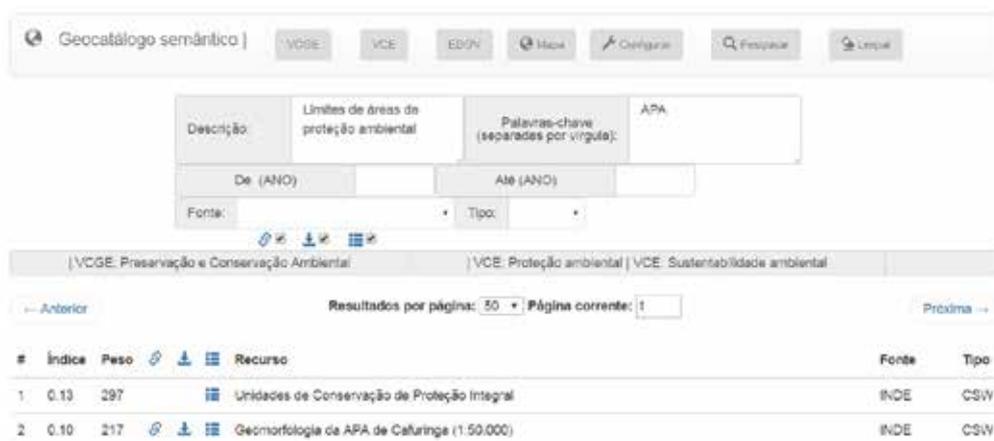


Figura 7:
Áreas de proteção ambiental importadas para ferramenta QGIS



Os registros dos recursos são recuperados mediante requisições REST para o banco de dados remoto, compostas pelos comandos na linguagem *Cypher*. Os principais tipos de interação são: consulta a ontologias, criação de relacionamentos entre os recursos e geração do ranking dos recursos geográficos. A Figura 6 exibe o ranking na forma de uma listagem com os links para visualização do recurso, download e visualização dos metadados.

Para validação do uso do geocatálogo foram utilizadas as etapas da metodologia SERVUS, utilizamos como base um problema relacionado a uma auditoria na área ambiental, que assim apresentamos: como obter os dados necessários, a nível brasileiro, para o levantamento da localização de áreas de proteção ambiental, utilizando como base os geosserviços disponíveis na rede de infraestrutura de dados espaciais nacional, de instituições e órgãos públicos?

O resultado final foi a localização e visualização dessas áreas. A Figura 7 apresenta as áreas de proteção ambiental localizadas pelo geocatálogo, cujos dados foram baixados de serviços WFS e importados para a ferramenta Quantum GIS.

7. CONCLUSÃO

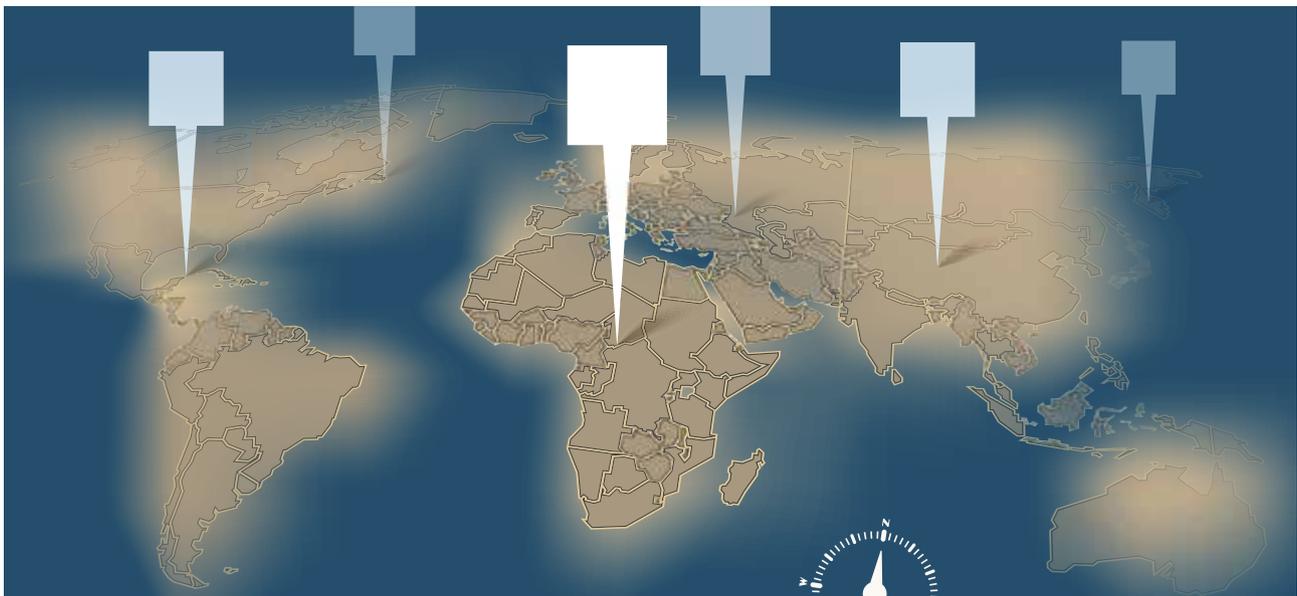
Este artigo abordou aspectos da implementação de um geocatálogo que possibilita a utilização de ontologias como parâmetros de pesquisa. As consultas são realizadas nos metadados de geosserviços de infraestruturas de dados espaciais. Foram aborda-

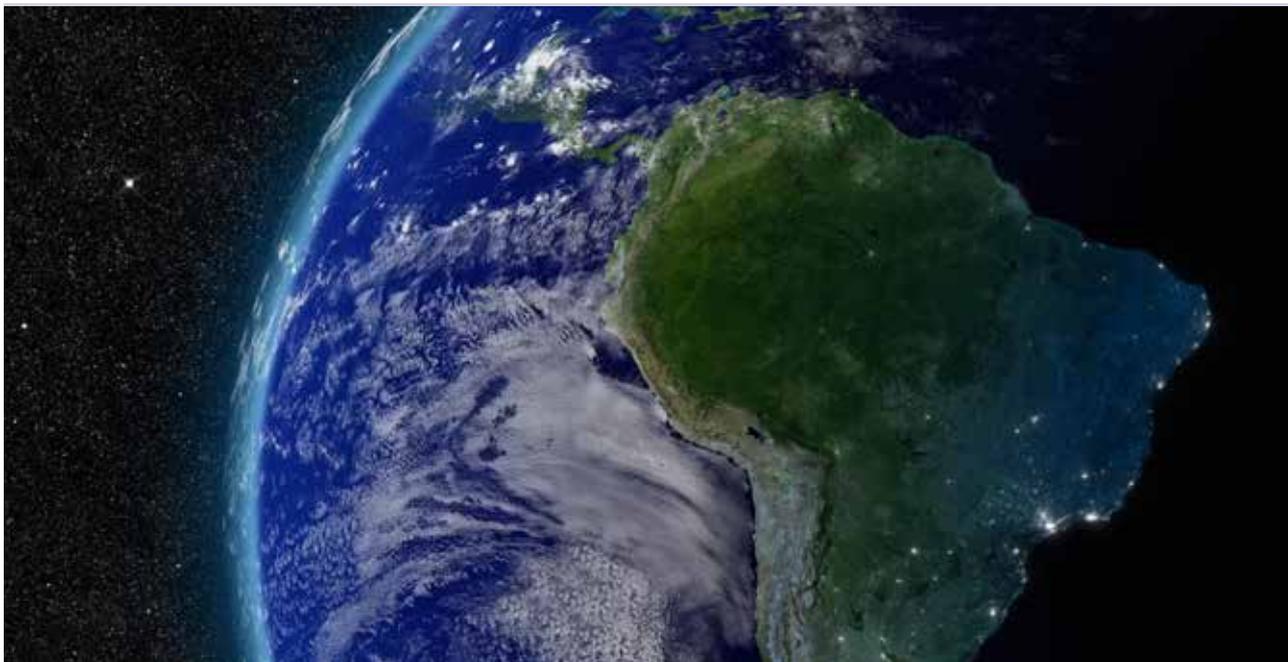
dos aspectos da implantação tais como: arquitetura e modelo de dados. No que se refere à arquitetura, o geocatálogo foi implementado como uma aplicação web de três camadas, passível de implantação em ambiente de nuvem. O modelo de dados foi implementado em banco de dados orientado a grafos.

O geocatálogo se mostrou viável como ferramenta de busca de recursos geográficos e de apoio à metodologia SERVUS, possibilitando a atuação nas etapas de *publishing* e *matching*. Podem ser consideradas como contribuições a utilização da linguagem *Cypher* no manuseio de rede de recursos, ontologias e metamodelo, de forma integrada, em banco de dados não relacionais orientado a grafos. O estabelecimento de mecanismos de parametrização das consultas, mediante ponderação de parâmetros e campos de metadados e o estabelecimento de um índice de relevância, são contribuições para o campo da busca em bases de metadados. Como trabalhos futuros, podemos considerar a automatização do processo *harvesting* e a integração com uma infraestrutura de dados espaciais corporativa, de forma que o geocatálogo atue conjuntamente com os geoportais.

REFERÊNCIAS

ANDRADE, F. G.; BAPTISTA, C. S. Using semantic similarity to improve information discovery in spatial data infrastructures. *Journal of Information and Data Management*, v. 2, n. 2, p. 181, 2011.





BERNABÉ-POVEDA, M. Á.; LÓPEZ-VÁZQUEZ, C. M. (Eds.). Fundamentos de las infraestructuras de datos espaciales (IDE). Madrid: Universidad Politécnica de Madrid, 2012. 593p.

BIEBERSTEIN, N. Service-oriented architecture compass: business value, planning, and enterprise roadmap. Indiana: FT Press, 2006.

BRASIL. Decreto n. 6.666, de 27 de novembro de 2008. Institui, no âmbito do Poder Executivo Federal, a Infraestrutura Nacional de Dados Espaciais – INDE, e dá outras providências. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2008/Decreto/D6666.htm>. Acesso em: 22 out. 2016.

CAMBOIM, S. F. Arquitetura para integração de dados interligados abertos à INDE-BR. 2013. 140f. Tese (Doutorado em Ciências Geodésicas) – Universidade Federal do Paraná, Curitiba, 2013.

DALTIO, J.; CARVALHO, C. A. Um framework para recuperação semântica de dados espaciais. In: BRAZILIAN SYMPOSIUM ON GEOINFORMATICS, 13, 2012, Campos do Jordão. Proceedings... São José dos Campos: MCTI/INPE, 2012. p. 60-65.

GIMENEZ, P. J. A.; TANAKA, A. K; BAIÃO, F. Uma proposta de integração semântica para a Infraestrutura Nacional de Dados Espaciais usando geo-ontologias. In: WORKSHOP DE COMPUTAÇÃO APLICADA AO GOVERNO ELETRÔNICO (WCGE), 2013, João Pessoa. Anais... João Pessoa: UFPB, 2013. v. 5. p. 25-32.

GREENWOOD, J.; WHITESIDE, A. OGC Web Services Common Standard. 2010. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.248.6351&rep=rep1&type=pdf>>. Acesso em: 22 out. 2016.

OGC. Catalogue Services 3.0 Specification: HTTP Protocol Binding. 2010. Disponível em: <<http://docs.opengeospatial.org/is/12-176r7/12-176r7.html>>. Acesso em: 22 out. 2016.

SANTOS, D. Implementação de um geocatálogo utilizando banco de dados orientado a grafos para apoio à metodologia SERVUS. 2016. 82f. Dissertação (Mestrado em Computação Aplicada) – Universidade de Brasília, Brasília, 2016. Disponível em: <http://alexandre.zaghetto.com/wp-content/uploads/2016/07/Disserta%C3%A7%C3%A3o_Vers%C3%A3o_FinalParaE-di%C3%A7%C3%A3o.pdf>. Acesso em: 22 out. 2016.

USLÄNDER, T. Service-oriented design of environmental information systems. Vol. 5. Karlsruhe: KIT Scientific, 2010.

O porquê de governança de dados em organizações de controle



Ricardo Dantas Stumpf

é servidor do Tribunal de Contas da União, lotado na Secretaria de Soluções de TI, onde atua na área de Gestão Técnica de Dados. É, também, bacharel em Ciências da Computação pela Universidade de Brasília.

RESUMO

Existem diferentes níveis de governança dentro de uma organização, seja lidando com estratégia, com TI, com dados em geral ou dados abertos. Dados são ativos gerenciáveis, seja para o funcionamento saudável da instituição, seja para análises de dados sofisticadas. Este artigo trata de definir a utilidade da governança dos dados nesse contexto, assim como elementos básicos para sua implantação, com um viés para organizações de controle.

Palavras-chave: Governança de dados. Gestão de dados. Ciência de dados.

1. INTRODUÇÃO

O que é importante em uma organização? Pessoas, orçamento, equipamentos e imóveis, dados... É difícil sequer pensar no funcionamento de organizações caso falte algum desses elementos, o que prova a crucial importância de todos eles. Vários desses são **ativos** essenciais à maioria das entidades e, como tais, devem ser bem **geridos** e **governados**, de modo a maximizar o seu valor. Em síntese, gerir é garantir que fazemos certo as coisas, enquanto governar é garantir que fazemos as coisas certas (WODZINSKI et al, 2015).

E agora que o TCU está investindo fortemente em análise de dados, cuidar dos dados tornou-se ainda



mais crítico, como pode-se inferir do trecho do relatório trimestral abaixo:

[C]hegou-se ao consenso dos três principais desafios que o tribunal deverá superar para obter bons resultados com a análise de dados: o desafio técnico, especialmente vinculado à qualidade dos dados; o regulatório, atinente às restrições legais e normativas; e o cultural, concernente aos aspectos comportamentais das pessoas envolvidas (RELATÓRIO TRIMESTRAL TCU 4º TRIMESTRE/2014, p. 101).

A 2ª edição do guia de Governança do TCU (2015) traz um resumo da interação entre a gestão e a governança:

“São funções da governança:

- a. definir o direcionamento estratégico;
- b. supervisionar a gestão;
- c. envolver as partes interessadas;

Figura 1:

TCU - Referencial básico de Governança aplicável a órgãos e entidades da Administração Pública, 2ª edição, p. 32



- d. gerenciar riscos estratégicos;
- e. gerenciar conflitos internos;
- f. auditar e avaliar o sistema de gestão e controle; e
- g. promover a *accountability* (prestação de contas e responsabilidade) e a transparência” (TCU, REFERENCIAL BÁSICO DE GOVERNANÇA APLICÁVEL A ÓRGÃOS E ENTIDADES DA ADMINISTRAÇÃO PÚBLICA, 2ª edição, p. 31).

E essa lógica vale tanto para o nível estratégico, como para os diversos níveis e facetas mais focadas da instituição. Em torno da governança e gestão corporativas, gravitam, dentre outras, a governança e gestão de tecnologia da informação e, em torno desta, a governança e gestão de dados produzidos e custodiados. Mesmo a governança e gestão de dados também se subdivide nas organizações mais complexas.

2. DADO E INFORMAÇÃO

É comum usarmos dado e informação como sinônimos, assim como conhecimento e competência. A rigor, porém, representam conceitos diferentes, segun-

do o professor Valdemar Setzer (2015), da Universidade de São Paulo.

Dado é “uma sequência de símbolos quantificados ou quantificáveis”, enquanto informação é “uma abstração na mente de alguém”.

Como exemplo simples, a sequência de símbolos **880**. É um **dado**, digital ou não. Se alguém reconhece 880 numa lista de salários, então temos a **informação** de um salário numérico com 880 unidades.

Já **conhecimento** exige uma abstração e experimentação pessoal (SETZER, 2015). Um funcionário de uma empresa sabe, da experiência, que se trata de R\$880,00, valor do salário mínimo em outubro de 2016 atribuído a um colaborador recém-contratado. O número 880 passou de dado a informação e, então, a conhecimento. Finalmente, um auditor tem **competência** para interagir e atuar sobre esse conhecimento durante uma fiscalização.

Esse processo acontece diariamente em ambientes profissionais: “processo” é um documento, fluxo de trabalho repetível, instrumento jurídico, código de computador em execução ou saliência de ossos? Depende.

É aí que a **Governança de dados** entra, ao utilizar o conhecimento e competência das pessoas para definir políticas, responsabilidades, glossários, metadados, fluxos de trabalho dos dados em movimento,

Figura 2:

Diferentes níveis de governança “gravitando” simultaneamente

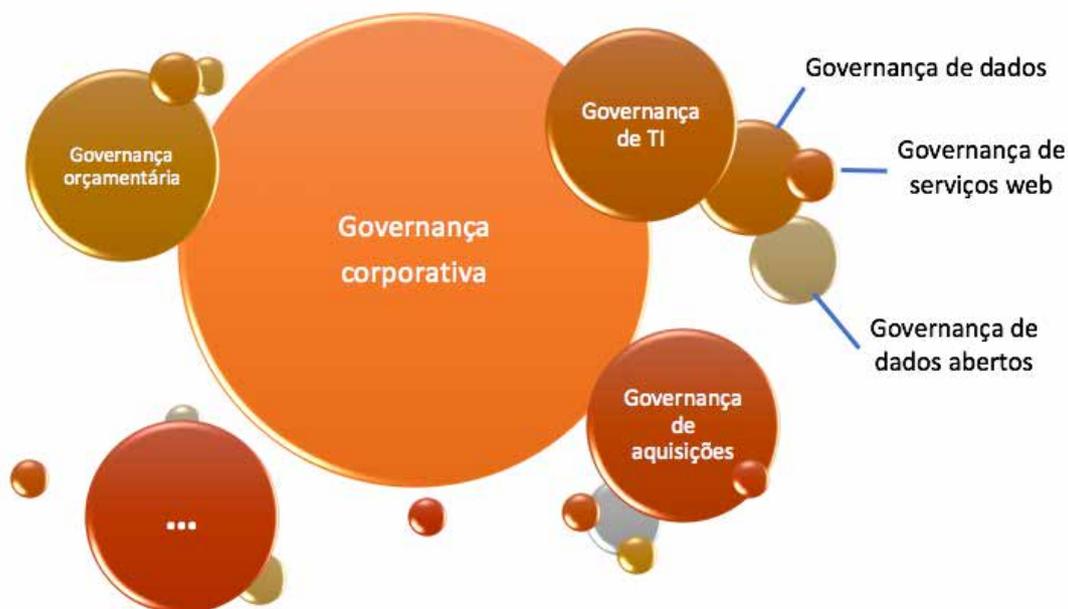


Figura 3:

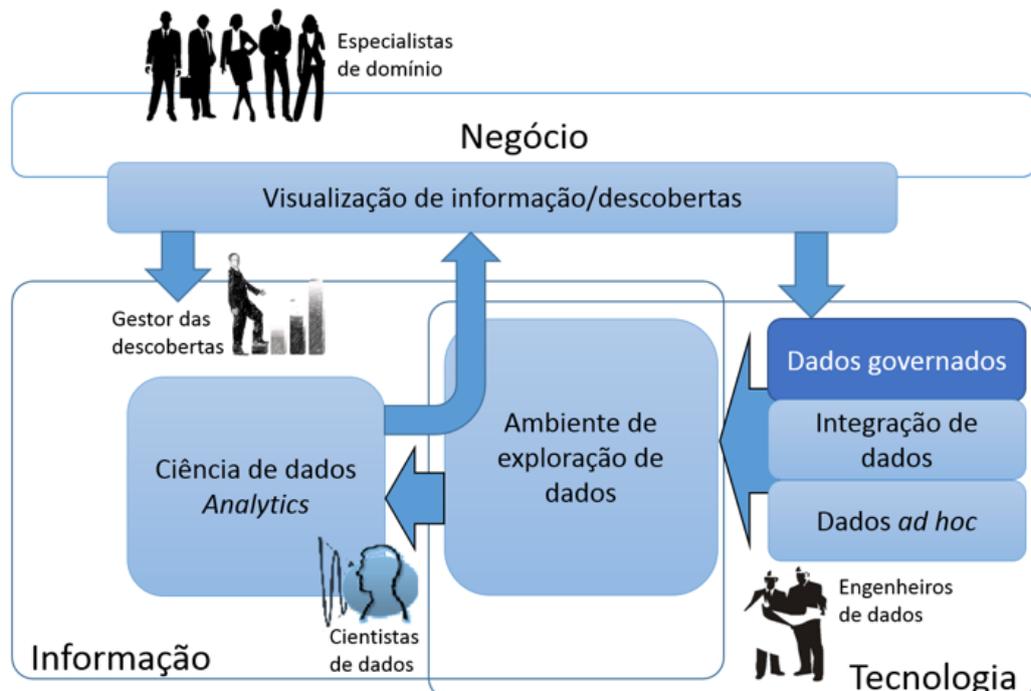
Exemplo de como símbolos relacionam-se com diversos conceitos



monitoramento e linhagem de dados que permitam à organização melhorar a qualidade, descoberta e entendimento desses, de modo a simplificar a extração de informações e conhecimentos, resultando em melhores análises e decisões de negócio.

Figura 4:

Modelo de ciclo de dados, adaptado (HENDERYCKX, 2016)



3. NOVO PETRÓLEO É DIGITAL

Dados estão sendo chamados de “novo petróleo” (VANIAN, 2016), mas sempre foram importantes e presentes, mesmo antes do formato digital. Contudo, nosso foco está nos dados digitais, que estão crescendo em tal ritmo e variedade que têm sido associados a “tsunamis”, “avalanches” e “tempestades” de dados, ou *Big Data*. Os dados internos, normalmente os mais alinhados com os processos de trabalho internos da corporação, são chamados de *Small Data* (HENDERYCKX, 2016).

Uma tendência nas empresas é investir cada vez mais em análise e visualização de dados, de modo a vislumbrar mais oportunidades de negócio e a aumentar a eficiência operacional. Da mesma forma, órgãos públicos, em especial os de controle, estão coletando grandes quantidades de dados de seus parceiros e auditados, além do *feedback* da sociedade nas redes sociais.

Jan Henderyckx, presidente da DAMA Benelux, apresentou um modelo genérico para o ciclo dos dados (figura 4):

Nesse modelo, observe que os dados que vêm das áreas de negócio precisam ser bem governados. Isso não significa que os dados externos custodiados não precisam de governança, mas as preocupações primárias podem ser outras. Por exemplo, o TCU possui um ambiente utilizado pelo Controle Externo como

laboratório de exploração e cruzamento de dados externos e internos.

A área responsável por esse laboratório adota, quando necessário, o “Princípio da Oportunidade”, que justifica a inclusão *ad hoc* de novas bases, mesmo antes de haver tratamentos de qualidade e documentação mais detalhados. O motivo é não deixar, no contexto de alguma auditoria, a janela da oportunidade de análise passar. A partir daí, porém, tal base torna-se elegível a um processo de tratamento mais completo.

Além disso, existe a questão **contextual**. Bases externas de baixa qualidade podem ser úteis para se descobrir achados de auditoria em sistemas de informação, enquanto bases derivadas de boa qualidade, com documentação completa de metadados, facilita a análise relativa ao uso efetivo dos repasses de recursos financeiros, por exemplo.

Quando o modelo apresentado tem baixa qualidade e organização, surgem problemas em três momentos (DYKES, 2016):

- as pessoas não confiam nos dados *a priori*;
- há tantas fontes que não se sabe o que usar; e
- há paralisia de análise, quando é difícil saber quando a análise está boa o suficiente.

Isso dito, podemos nos voltar para gestão e governança dos dados.

4. GESTÃO DE DADOS

Organizações costumeiramente já gerem seus dados, em especial os estruturados e semiestruturados. Quanto à forma, é comum que a gestão siga boas práticas em seus nichos, normalmente pela força da necessidade e pelo conhecimento compartilhado pelos pares internos e parceiros, como os da comunidade TI-Controle, com cerca de 20 órgãos de controle que discutem melhores práticas em TI. A sistematização dos nichos de atuação aparece em *frameworks* como o preconizado no DMBok (MOSLEY *et al.*, 2012), da DAMA International.

Na nova versão desse *framework*, até técnicas mais modernas, como integração por serviços web e *publish-subscribe* são contemplados como formas de dados circularem entre sistemas ou módulos de sistemas, localmente ou em nuvem (BRADLEY, 2016).

Apesar disso, o tema Governança de dados (que é uma espécie de “cola” a todos os processos ligados ao as-

sumo) só agora está ganhando energia, visto que as empresas e organizações que dependem de dados querem extrair o máximo valor desse ativo, cada vez mais abundante e mais complexo. Cabe lembrar que já podem existir áreas que governam bem seus dados. Nesses casos, segundo o político inglês Lucius Cary, “quando não é necessário mudar, é necessário não mudar” (In: SEINER, 2014).

A matéria-prima para o trabalho de qualquer órgão de controle é informação. E seus produtos finais (determinações, recomendações) são fontes primárias de informação a todos os seus auditados.

5. GOVERNANÇA DE DADOS

Finalmente chegamos a uma definição formal, segundo John Ladley (2012):

Governança de dados é a organização e implementação de políticas, procedimentos, estrutura, papéis e responsabilidades que delineiam e reforçam regras de comprometimento, direitos decisórios e prestação de contas para garantir o gerenciamento apropriado dos ativos de dados.

Da definição, discutiremos brevemente a necessidade de princípios, políticas, curadores de dados e clas-

Figura 5:

Visão sintética do *framework* para dados, segundo DMBok



sificação dos dados, glossários e metadados, lideranças e abordagens de implantação.

O importante é saber dosar. Governança de dados muito burocrática é um convite à desobediência; já o excesso de flexibilidade pode levar à desgovernança, a uma gestão de dados menos eficiente. Por isso, deve-se começar pela definição dos **princípios**.

5.1 PRINCÍPIOS

Princípios bem específicos, lógicos e viáveis facilitam todos os passos seguintes. Por exemplo, se o princípio “dados são ativos corporativos” for adotado pela organização, devemos tratar a entidade “dados” como tratamos outros ativos: melhoria contínua, definição de responsabilidades etc (LADLEY, 2012).

Cada organização tem seus princípios. A tabela 1 abaixo é exemplificativa e não-exaustiva.

5.2 POLÍTICAS

Como diz John Ladley (2012), a política é o que dá “dentes aos princípios”. É o documento formal que faz a organização adotar os princípios discutidos. É comum também a política definir responsabilidades pelos dados (curadoria), assim como a estrutura organizacional que irá conduzir e monitorar os esforços de Governança de Dados.

Como um bom exemplo, deixamos a sugestão de leitura da Política de Governança da Informação do Banco Central do Brasil, publicada pela Portaria nº47, de 20 de fevereiro de 2013, atualizada em 2016.

5.3 CURADORIA E CLASSIFICAÇÃO

Os curadores (*data stewards*, em inglês) são as pessoas ou grupos de pessoas que têm responsabilidades de **cuidar dos dados sob sua alçada de negócio**. Essa é uma mudança fundamental, pois compartilha com a TI a missão de cuidar dos dados corporativos.

Suponha que o princípio da qualidade tenha sido adotado. Isso significa que as áreas de negócio têm que se preocupar com a qualidade do dado que geram para si e para as outras, e devem reportar problemas de qualidade em relatórios e outras fontes que venham de outras áreas de negócio. O Escritório de **Governança de Dados**, ou equivalente, é o setor que auxiliará os curadores nessa tarefa.

A partir da Lei nº 12.527/2011, conhecida como a Lei de Acesso à Informação, formalizou-se como diretriz para a Administração Pública a “publicidade [das informações] como preceito geral e do sigilo como exceção”. Assim, os curadores também devem ativamente classificar as informações que produzem ou custodiam, de modo a acelerar a disponibilização dos dados públicos ou de restringi-los nos termos da lei.

5.4 GLOSSÁRIOS E METADADOS

Glossários, ou vocabulários, definem e eliminam ambiguidades em termos de negócio, como o termo “processo”, citado na seção 2. O TCU mantém o Vocabulário de Controle Externo (VCE), criado para, segundo o Ministro-Presidente Aroldo Cedraz, “padronizar o tratamento de informações especializadas e conferir maior

Tabela 1:

Lista de alguns princípios para implantação de Governança de Dados

Princípios	Descrição
Regra de ouro	Todos os dados são tratados como ativo corporativo.
Federação	Há padrões definidos para as estruturas de dados.
Eficiência	Dados relevantes devem estar disponíveis no momento certo, no lugar certo e no formato certo para usuários autorizados.
Qualidade	Dados corporativos são medidos e geridos para terem qualidade.
Gestão de risco	Manter conformidade com a legislação, políticas e normativos internos relativos a dados.
Colaboração	Dados corporativos são recursos compartilhados e publicizados.
Contextualização	O contexto de uso do dado muda sua forma de armazenamento, tratamento e utilização.
Inovação	Novas técnicas são incentivadas, seguindo-se os demais princípios.

Fonte: adaptado de LADLEY (2012)

agilidade e precisão na recuperação dos conteúdos presentes nos sistemas de informação do TCU” (VCE, p.5).

A definição clássica de metadados é “dado sobre o dado”. Mas podemos ampliar esse conceito em duas vertentes: de negócio e técnicos.

Os metadados de negócio são aqueles que contextualizam os dados armazenados ou em movimento. Por exemplo, considere a tabela WZV da base de dados TTNN (hipotéticos). Como a área de negócio e os especialistas de domínio conhecem esse assunto, eles podem **enriquecer os metadados** existentes com informações específicas. Agora, os cientistas de dados saberão que a tabela WZV tem os dados cadastrais de fornecedores e que a base TTNN traz organizações auditadas que, dentre outros, compraram daqueles fornecedores.

Já os metadados técnicos são aqueles que trazem informações de TI sobre os dados, como nomes de campos e tabelas, datas de carga, versão, tamanhos, tipos etc.

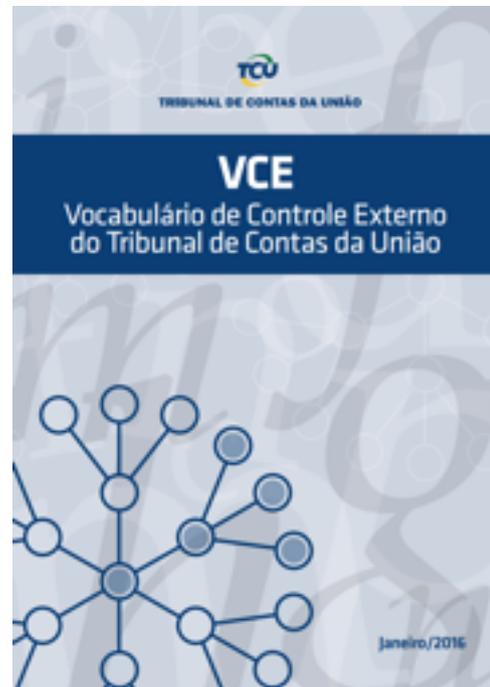
A “magia” acontece quando os metadados técnicos e de negócio são **integrados para descoberta** e uso das áreas de análise e de gestão de mudanças. No TCU, há algumas iniciativas nesse caminho. Os dados do VCE estão aos poucos sendo integrados a outras soluções de tecnologia da informação, de forma a otimizar e uniformizar o uso dos conceitos. Outra é uma ferramenta especializada em metadados que mostra, de forma integrada, informações oriundas de modelagens de dados, integrações por ETL (*Extract Transform Load*) e bases de dados corporativas (figura 7).

5.5 ESTRUTURAS

Cada corporação tem sua cultura e necessidades, e não faria sentido haver apenas uma forma de estrutura

Figura 6:

Capa da edição de janeiro/2016 do VCE do TCU

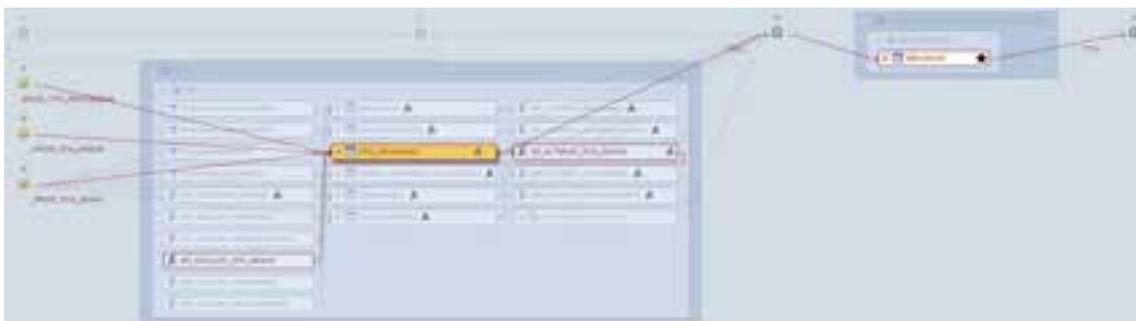


organizacional para governar dados. A figura do CDO (*Chief Data Officer*) é cada vez mais comum. Essas são algumas das formas utilizadas, segundo Bergson Rêgo (2013):

- a. CIO liderando equipe centralizada na TI, que interage com o negócio;
- b. CDO interagindo com TI e uma área de gestão de dados central do negócio;

Figura 7:

Exemplo editado de metadados técnicos integrados



- c. CDO interagindo com TI e várias áreas de gestão de dados do negócio; ou
- d. um CDO para cada grande linha de negócio, interagindo com TI (híbrido).

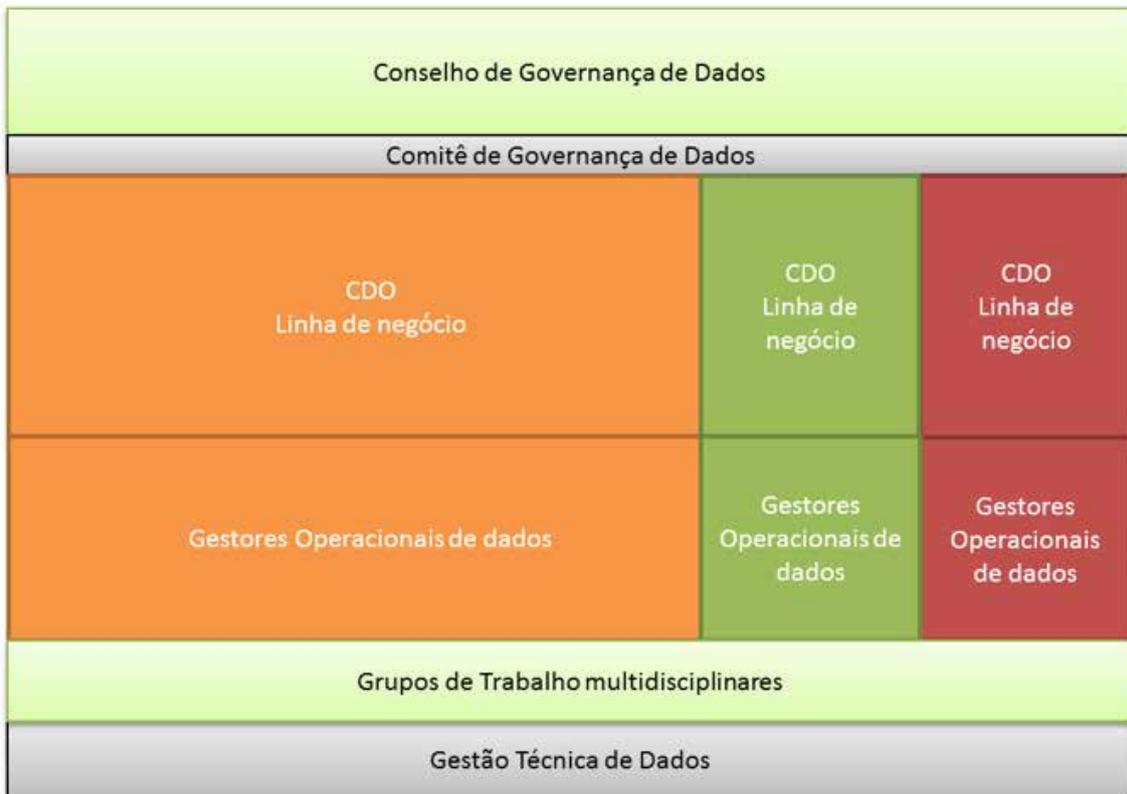
Figura 8:

Estrutura simples hierárquica (RÉGO, 2013)



Figura 9:

Estrutura híbrida, adaptado (IPPOLITI, 2016)



A estrutura híbrida é particularmente interessante, e utilizada na prática em um grande banco multinacional (IPPOLITI, 2016). Dada a grande variedade interna de dados disponíveis, eles resolveram dividir a gestão desses em grandes linhas de negócio (figura 9).

5.6 ABORDAGENS DE IMPLANTAÇÃO

Há três formas estruturadas mais comuns de se implantar Governança de Dados (SEINER, 2016):

- a. Comando e controle;
- b. Tradicional; e
- c. Não-invasivo.

Na abordagem comando e controle, o comitê de mais alto nível define as regras, determina os curadores, compra ferramentas e cobra resultados.

A tradicional identifica os curadores e os orienta com processos mais genéricos e ferramentas existentes, e mede resultados analisando os dados envolvidos.

Por fim, a não-invasiva formaliza a curadoria tácita, foca em aplicar os processos existentes, estimula construção de ferramentas e uso das atuais e mede resultado pelo aumento percebido em eficiência e efetividade de capacidades de análise.

Como já frisamos, nenhuma é necessariamente melhor que a outra. Depende da cultura e necessidades de cada organização.

4. CONCLUSÃO

Apesar da importância atual do tema Governança de Dados, esta será ainda maior no futuro, no qual haverá muita automatização, mesmo de atividades intelectuais, desde que repetíveis. E os dados, sejam oriundos de sistemas informatizados ou de sensores em objetos do dia a dia, serão um instrumento básico para essa automatização.

Isso significa que as organizações deverão capturar dados, armazená-los de forma eficiente e automatizar análises com algoritmos cognitivos (HENDERYKCX, 2016). Segundo Ian Rowlands (2016), metadados de negócio terão sua semântica, e não apenas sua sintaxe, inferida por algoritmos e *crowdsourcing* (grandes grupos pensando e colaborando juntos).

Também é desafio crescente identificar e reduzir o *dark data*, que representa perda de oportunidades, desperdício de recursos e risco (ROWLANDS, 2016). Rowlands reitera trecho de declaração do Au-

ditor-Geral do Canadá, em que esse é um problema a ser combatido:

Um dos temas que unem muitas de nossas auditorias é que os dados coletados de muitas organizações governamentais ou são inutilizáveis, ou não são usáveis ou não são utilizados (Auditor-Geral do Canadá em ROWLANDS, 2016).

Para que tal redução seja possível, impera a necessidade de governar tais dados – na dose certa para cada caso e contexto, sempre com o **olhar na missão e visão da instituição**.

REFERÊNCIAS

BRADLEY, C. The new DMBOK 2 discipline of Data Integration. Disponível em: <<https://www.brighttalk.com/webcast/12405/186095/the-new-dmbok-2-discipline-of-data-integration>>. Acesso em: 12 abr. 2016.

BRASIL, BANCO CENTRAL DO BRASIL – BCB. Portaria nº 47, de 20 de fevereiro de 2013. Diário Oficial da União – DOU, 21 de fevereiro de 2013. Seção 1, p. 24.

BRASIL, Comunidade de Tecnologia da Informação Aplicada ao Controle – TI Controle. Sobre a Comunidade. Disponível em: <<http://www.ticontrôle.gov.br/ticontrôle/sobre-a-comunidade/sobre.htm>>. Acesso em: 20 out. 2016.



BRASIL. LEI nº 12.527, de 18 de novembro de 2011. Lei de Acesso à Informação. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>. Acesso em: 20 out. 2016. 2011.

BRASIL, TCU – TRIBUNAL DE CONTAS DA UNIÃO. Referencial básico de governança aplicável a órgãos e entidades da administração pública. Versão 2 – Brasília: Tribunal de Contas da União. Secretaria de Planejamento, Governança e Gestão. 80p. Disponível em: <<http://www.tcu.gov.br/governanca>>. Acesso em: 20 out. 2016. 2014.

_____. Relatório de Atividades: 4º trimestre de 2014 – Brasília: Tribunal de Contas da União. Secretaria-Geral da Presidência, p.101. Disponível em: <<http://portal.tcu.gov.br/publicacoes-institucionais/relatorios/relatorios-de-atividades>>. Acesso em: 20 out. 2016. 2015.

_____. Vocabulário de Controle Externo do Tribunal de Contas da União – Brasília: Tribunal de Contas da União. Instituto Serzedêllo Corrêa. Disponível em: <<http://portal.tcu.gov.br/vocabulario-de-controle-externo/>>. Acesso em: 20 out. 2016. 2015.

CANADÁ, Office of the Auditor General of Canada – OAG. Spring Reports of the Auditor General of Canada. Auditor General’s Opening Statement. Disponível em: <http://www.oag-bvg.gc.ca/internet/English/osm_20160503_e_41358.html>. Acesso em: 20 out. 2016. 2016.

DMBOK. MOSLEY, M.; BRACKETT, M.; EARLEY, S. HENDERSON, D. DAMA Guia para o corpo de conhecimento em gerenciamento de dados. Technics Publications, versão brasileira 2012.

DYKES, B. Big Data Paralyzing Your Business? Avoid These 3 Common Traps. Forbes, Entrepreneurs. Disponível em: <<http://www.forbes.com/sites/brentdykes/2016/09/28/big-data-paralyzing-your-business-avoid-these-3-common-traps/>>. Acesso em: 20 out. 2016.

HENDERYCKX, J. Sustaining Data Driven Innovation. In: DATA MANAGEMENT CONFERENCE LATIN AMERICA, 2016, São Paulo. Anais eletrônicos... São Paulo: DAMA Brasil, 2016. 42 slides. Disponível em: <<http://www.dmc-latam.com/palestrantes/jan-henderyckx/>>. Acesso em: 19 set. 2016.

IPPOLITI, J. Implementing Data Governance Strategies. In: ENTERPRISE DATA GOVERNANCE ONLINE, 2016. Anais eletrônicos... 20 slides. Disponível em: <<http://video.dataversity.net/video/keynote-implementing-data-governance-strategies/>>. Acesso em: 12 fev. 2016.

LADLEY J. Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program. The Morgan Kaufmann Series on Business Intelligence. Morgan Kaufmann. 2012.

RÊGO, B. L. Gestão e Governança de Dados: Promovendo Dados Como Ativo de Valor Nas Empresas. Rio de Janeiro: Brasport, p. 60-74. 2013.

ROWLANDS, I. Data Management in Motion. In: DATA MANAGEMENT CONFERENCE LATIN AMERICA, 2016, São Paulo. Anais eletrônicos. São Paulo: DAMA Brasil, 2016. 25 slides. Disponível em: <<http://www.dmc-latam.com/palestrantes/ian-rowlands/>>. Acesso em: 19 set. 2016.

SEINER, R. Non-Invasive Data Governance: The Path of Least Resistance and Greatest Success. Technics Publications. 2014.

_____. Comparing Approaches to Data Governance. Disponível em: <<http://tdan.com/comparing-approaches-to-data-governance/20386>>. Acesso em: 5 out. 2016.

SETZER, V. W. Dado, informação, conhecimento e competência. Journal Data & Information 1(1), DAMA Brasil: p. 38-55. 2015.

VANIAN, J. Why Data Is The New Oil. Fortune, Brainstorm Tech. Disponível em: <<http://fortune.com/2016/07/11/data-oil-brainstorm-tech/>>. Acesso em: 20 out. 2016. Julho de 2016.

WODZINSKI, M. et al. Building an Impactful Data Governance Program One Step at a Time. Disponível em: <<https://www.brighttalk.com/webcast/10477/142503/building-an-impactful-data-governance-program-one-step-at-a-time>>. Acesso em: 26 fev. 2015.

O georreferenciamento dos bens imóveis públicos no sistema geodésico brasileiro para fins de incorporação no cadastro técnico multifinalitário: construção da regularização imobiliária dos municípios



Davi Lopes Silva

é graduado em Geografia pela Universidade Estadual do Ceará – UECE, graduando em Computação pela Universidade Estadual do Ceará – UECE, especialista em Geoprocessamento e Georreferenciamento pela Universidade Cândido Mendes – UCAM e atualmente ocupa o cargo de Gerente da Célula de Bens Imóveis de Fortaleza.

RESUMO

Os bens públicos são por natureza e origem patrimônio das cidades, contudo foram deixados de lado nos cadastros técnicos do Brasil, uma vez que o objetivo era a arrecadação tributária, o que resultou em uma diminuta atualização das informações dos perímetros e limites dos bens públicos ao longo da história. Este trabalho destaca a importância de utilizar o georreferenciamento de imóveis no sistema geodésico brasileiro – Sirgas 2000 – para o levantamento das dimensões dos imóveis, visando à integração das informações no Cadastro Técnico Multifinalitário (CTM) que as cidades brasileiras estão implantando. Apresenta-se também a relevância de registrar as atualizações efetuadas nas áreas públicas nos Cartórios de Registro de Imóveis (CRI). Utilizou-se como metodologia a bibliografia brasileira existente acerca dos CTM e dos CRI e os estudos de caso de diversos autores sobre o assunto. Enfatiza-se, ao final, a necessidade de incluir nos cadastros técnicos do CTM as informações dos bens patrimoniais, visando a centralização, controle e transparência dos gastos públicos.

Palavras-chave: Georreferenciamento. Bens públicos. Sistema geodésico brasileiro. Cadastro técnico multifinalitário.



1. INTRODUÇÃO

A partir do aperfeiçoamento das técnicas geodésicas, das novas tecnologias e da importância da preservação do patrimônio público, surgiu a necessidade de localizar e delimitar os bens públicos das cidades de uma forma mais precisa, utilizando-se do georreferenciamento de imóveis.

Aplicando os conhecimentos topográficos e as técnicas de levantamento planimétrico, apresenta-se a relevância de representar objetos espacialmente por meio da criação de polígonos georreferenciados. De forma simples e concisa, busca-se demonstrar a importância de conhecer, delimitar e localizar com precisão os bens imóveis públicos, visando a fomentar a incorporação dessas informações no Cadastro Técnico Multifinalitário (CTM) das cidades.

Nas últimas décadas, os agentes públicos investiram em equipamentos e bens imóveis voltados para o atendimento das necessidades da população e prestação de serviços, o que ampliou consideravelmente o patrimônio imobiliário das cidades, gerando assim novos bens que nem sempre possuíam registros imobiliários assertivos ou atualizados em relação a seus limites e localização.

Os Cartórios de Registro de Imóveis (CRI) receberam e armazenaram, em alguns casos, informações imprecisas, como a medição de distâncias em “passos ou palmas”, o que ao longo da história brasileira foi subs-

tituído pela unidade de medida em metros do Sistema Internacional de Unidades. Ocorre, ainda, que muitos registros foram atualizados sem uma aferição precisa dos órgãos atuantes municipais, o que pode levar, no futuro, à sobreposição de registros imobiliários de particulares sobre terrenos públicos, podendo ocasionar discussões e litígios.

Recomenda-se a adoção de um CTM que envolva um cadastro único dos bens públicos e particulares para que se possa verificar se existe sobreposição de registros imobiliários ou ocupações irregulares, com a possibilidade de recursos administrativos para impugnação nos CRI ou intervenções judiciais mais rápidas.

Este trabalho está amparado na Legislação Brasileira, que trata sobre o georreferenciamento de imóveis na resolução nº 01/2005 do Instituto Brasileiro de Geografia e Estatística (IBGE), que alterou e definiu o novo sistema geodésico brasileiro, além do Código Civil Brasileiro, que detalha os bens considerados públicos. Contudo, não existe uma norma ou regra que determine também a aplicação do georreferenciamento nos imóveis e terrenos estatais. Considera-se também para essa temática a Portaria nº 511/2009 do Ministério das Cidades, que trata sobre as diretrizes para criação, instituição e atualização do CTM nos municípios brasileiros, além de autores acadêmicos que trabalham as questões em pauta, como Carneiro (2003), Pimentel (2012), Erba (2005) e outros.

2. DESENVOLVIMENTO

Para fomento da gestão pública, o município deve conhecer seu território além dos limites geográficos, inclusive deve considerar para otimização de recursos o levantamento do patrimônio público, como as praças, os rios, os logradouros públicos, os hospitais, as sedes do governo, as creches, os postos de saúde ou qualquer bem registrado ou em sua posse.

No passado, para localizar uma propriedade era necessário ler seu registro imobiliário ou memorial descritivo para conhecer suas dimensões e áreas, o que resultaria, ao final, em um levantamento de campo para aferição das medidas que nem sempre correspondem ao esperado. Além disso, nesses documentos cartorários, também reside a descrição dos confinantes, que se for originado de um registro muito antigo, pode não descrever o imóvel ou o terreno limítrofe, mas mencionar o nome dos confinantes ou fazer alusão à geografia do local, como citação de rios, lagos, mares, encostas e montanhas.

Acrescente-se também que utilizar endereços, descrição do relevo e nomes de confinantes não colabora com o passar do tempo para a correta e precisa localização caso haja remoção de edificações, novas construções ou mudança de proprietários, sem deixar de destacar também que os bens públicos podem ser alvo de ocupações irregulares e descaracterização de propriedade por terceiros.

Para sanar esse problema, reduzir litígios e aumentar a precisão de localização de imóveis, surgiu o georreferenciamento, que pode ser definido como o ato de “descrever e definir a localização de um imóvel, por levantamento topográfico, com equipamentos de precisão, utilizados por profissional especializado” (GEORREFERENCIAR, 2015).

Essa técnica consiste em representar, por meio de polígonos, num dado sistema de referência padrão, o perímetro de um imóvel a partir de suas coordenadas de localização, com o objetivo de incorporação no CTM. Posteriormente, essas informações podem ser objeto de sobreposição e comparação com os registros públicos ou servir para retificação de documentos imobiliários nos CRI.

De acordo com Zocolotti (2005), uma poligonal representa uma série de linhas consecutivas com comprimentos e direções conhecidas, obtidas mediante levantamento de campo. Esse pesquisador considera também que “O levantamento de uma poligonal é realizado através do método de caminhamento, percorrendo-se o contorno de um itinerário definido por uma

série de pontos, medindo-se todos os ângulos e lados e uma orientação inicial” (p. 10). Recomenda-se o uso de poligonais fechadas, uma vez que os pontos de partida e de chegada deverão estar na mesma coordenada, visando ao cálculo da área dos bens públicos.

Entre o período de fevereiro de 2005 e fevereiro de 2015, era permitido utilizar em concomitância com o Sistema de Referência Geocêntrico para as Américas (Sirgas 2000), o South American Datum 1969 (SAD 69) e o Córrego Alegre, havendo respaldo na legislação deste país para trabalhos técnicos utilizando o georreferenciamento de imóveis.

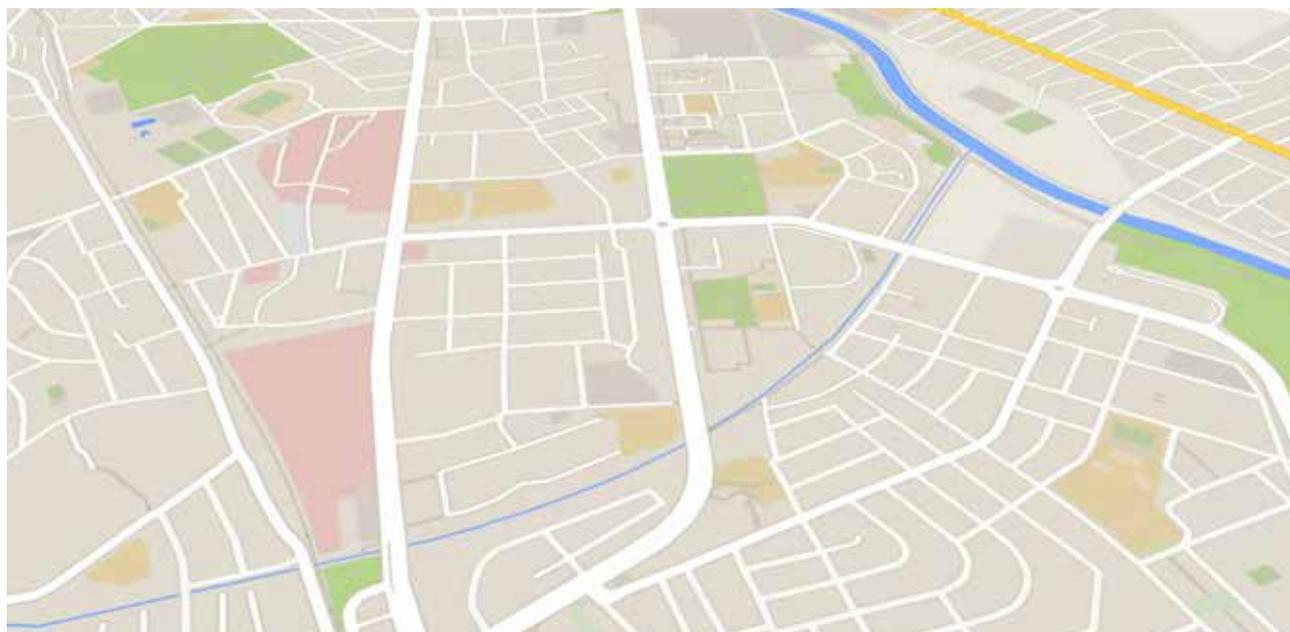
Com o crescimento do uso dessa ferramenta no Brasil, diversos levantamentos foram realizados seguindo esses sistemas de referência, sendo considerados, inclusive, para registro nos CRI, o que poderá levar no futuro a litígios, caso os registros imobiliários não tenham agregado a informação do sistema de referência adotado nos títulos antigos, uma vez que desconhecer o sistema de referência adotado implica a incorreta localização ou deslocamento de poligonais.

Desde 25 de fevereiro de 2015 o Sirgas 2000 é reconhecido e adotado oficialmente para fins de georreferenciamento, definido pela Resolução da Presidência do IBGE, que esclarece ainda:

Para o desenvolvimento das atividades geodésicas, é necessário o estabelecimento de um sistema geodésico que sirva de referência ao posicionamento no território nacional. A materialização deste sistema de referência, através de estações geodésicas distribuídas adequadamente pelo país, constitui-se na infraestrutura de referência a partir da qual os novos posicionamentos são efetuados. (BRASIL, 2005).

Para Erba (2005, p. 25), “Os problemas derivados da localização relativa desaparecem ao adotar o posicionamento absoluto dos imóveis. Nesse sistema, cada detalhe levantado recebe uma coordenada correspondente a um sistema de referência único, podendo ser este municipal ou nacional”. Com isso, o levantamento cadastral mensurado por coordenadas e sistema de referência único contribui para a localização de imóveis, além de possibilitar a comparação dos registros imobiliários existentes e a realidade encontrada em campo.

Conforme Carneiro (2003), o levantamento de informações e consequente cadastro tem no Brasil enfoque de tributação, que desconsidera na maioria das vezes a parte sem interesse fiscal, como nesse estudo de caso, as áreas públicas.



Uma vez que essa autora considera o cadastro municipal como uma ferramenta que deve ir além do aspecto tributário, devendo participar no planejamento, controle urbano e ordenamento territorial, destaca-se:

O cadastro multifinalitário é definido por Dale & McLaughlin (1990) como um sistema de informações territoriais projetado para servir tanto a organizações públicas como privadas, além de servir aos cidadãos. Difere dos demais sistemas de informações territoriais por ser baseado em parcelas. Serve de base para os demais tipos de cadastro (legal, fiscal etc...) (CARNEIRO, 2003, p. 24).

Em consonância com o pensamento de um cadastro territorial nas cidades, está Erba (2005, p. 21), que demonstra uma declaração da Federação Internacional de Geômetras em suas palavras a seguir:

Em concordância com esta última afirmação está a Declaração sobre o Cadastro redigida pela FIG, em 1995, que afirma que o Cadastro é um sistema de informação territorial, normalmente baseado em parcelas, que registra interesses sobre a terra, como direitos, restrições e responsabilidades. Ainda acrescenta que o Cadastro pode ser estabelecido para arrecadação, legal e, ou, de apoio ao planejamento, buscando sempre o desenvolvimento social e econômico, destacando, porém, que não existe a necessidade de pensar em um Cadastro uniforme para todos os países ou jurisdições.

O cadastro de áreas públicas deve seguir o padrão dos demais cadastros técnicos brasileiros, devendo possuir, no mínimo, considerando diretiva do Infrastructure for Spatial Information in Europe (Inspire), as seguintes informações: “geometria da parcela, identificador único, referência geodésica e índice das parcelas para impressão/publicação” (PIMENTEL; CARNEIRO, 2012, p. 210). Nesse aspecto, ao criar um cadastro de áreas públicas, estas devem estar aptas a se integrar aos demais cadastros municipais contendo também esses dados.

Para fins didáticos e de entendimento, adota-se o conceito de parcela cadastral definida pela Portaria nº 511/2009, que trata toda e qualquer propriedade como unidade a ser incorporada no CTM, inclusive os bens públicos, o que se transcreve a seguir:

Art. 2º A parcela cadastral é a menor unidade do cadastro, definida como uma parte contígua da superfície terrestre com regime jurídico único.

§ 1º É considerada parcela cadastral toda e qualquer porção da superfície no município a ser cadastrada.

§ 2º As demais unidades, como, lotes, glebas, vias públicas, praças, lagos, rios e outras, são modeladas por uma ou mais parcelas de que trata o caput deste artigo, identificadas por seus respectivos códigos.

§ 3º Deverá ser atribuído a toda parcela um código único e estável.

Em seu art. 3º, a referida Portaria do Ministério das Cidades explica que “toda e qualquer porção da superfície territorial no município deve ser cadastrada em parcelas”, o que nos remete a partir da legislação para composição de um CTM que, dependendo da situação, pode exigir o cadastro das parcelas das áreas públicas por meio de uma ou mais parcelas ou o que possibilitar maior transparência e preferência dos cadastradores, sendo estas parcelas identificadas por códigos únicos.

Segundo Pimentel e Carneiro (2012, p. 205), a definição da Federação Internacional de Geômetras retrata claramente a questão da composição das informações no cadastro:

Segundo definição da Federação Internacional de Geômetras (FIG, 1995), a parcela é a unidade territorial do cadastro e pode ser definida de muitas maneiras, dependendo da sua finalidade para o cadastro. Por exemplo, uma área com um tipo específico de uso de terra; ou uma área de domínio único ou propriedade individual ou de um grupo. Os limites podem ser formais ou informais e para identificação dos polígonos utiliza-se um código único.

Esse autor recomenda, ainda, que qualquer cadastro de parcelas territoriais deve ser multifuncional e adotar padronização, em suas palavras a seguir:

Indica-se que o levantamento das parcelas seja georreferenciado ao Sistema Geodésico Brasileiro, para identificação inequívoca de seus limites. A projeção UTM é recomendada até que se defina uma projeção específica para cartografia em escala grande. (loc. cit.).

Na concepção de Fonseca (2010, p. 14), um dos objetivos de montar um CTM em âmbito municipal é o de desenvolver também um cadastro do patrimônio público, descrevendo-o como “uma relação de bens pertencentes ao patrimônio público”.

Esse mesmo autor destaca que um cadastro para ser multifuncional deve cumprir uma função social, pois “Hoje já se tem um significado mais amplo e diverso sobre o cadastro, visualizando-se então sua importante função social, devido às diversas informações que ele pode conter, deixando de ter somente a função de taxaçaõ, podendo ser então considerado multifuncional” (loc. cit.).

Em sua tese, Galdino (2006, p. 56) destaca ainda uma preocupação sobre o cadastro na forma de parcelas, uma vez que esse formato deve preservar outras

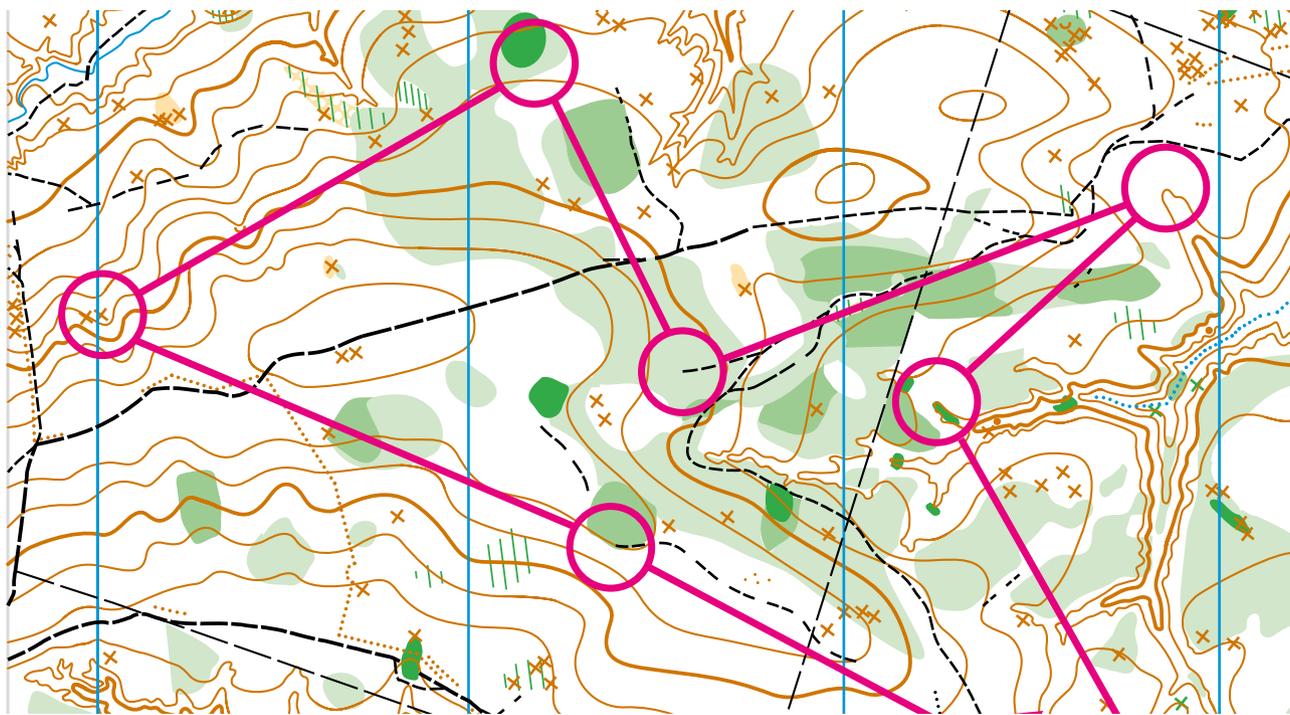
características próprias do regime jurídico e legislativo brasileiro, o que implica a necessidade de adequação do levantamento dos bens públicos também nesse sentido. Esse autor também aborda a questão da responsabilidade dos ordenadores territoriais no tocante à observação dos direitos e restrições da propriedade, seja no meio urbano ou rural, o que se transcreve a seguir:

Um cadastro territorial baseado em parcelas observando os direitos, restrições e responsabilidades da propriedade urbana ou rural, nos aspectos geométricos, jurídicos, econômicos e de gestão, aqui no nosso país, só na atualidade está sendo ventilado e discutido interativamente pelo poder público, instituições e comunidade científica, particularmente nos moldes da FIG e adequado à realidade cultural do país. Carecendo, no entanto, que as instituições e estudiosos perseverem na sua inclusão em legislação específica explicitamente, com a designação da instituição responsável, regras e regulamentações. (loc. cit.).

Ao delimitar e cadastrar os bens imóveis de uma cidade, é possível atuar também na regularização fundiária, como apresenta Carneiro (2003, p. 25):

O trabalho de regularização fundiária consiste em uma série de procedimentos técnicos, jurídicos e administrativos (cadastro e levantamento topográficos, análise da origem dominial do imóveis, ações discriminatórias judiciais, demarcações, planos de legitimação de posses etc.), que





visam acabar com a incerteza domínial, separando as áreas devolutas das particulares e legitimando a posse e o uso de terras públicas.

No caso das áreas públicas, deve-se acrescentar outras informações importantes a serem coletadas para compor o referido cadastro: tipo de bem público, legislação de uso e ocupação do bem, dimensão do terreno e da edificação, identificador de correspondência imobiliária dos limites físicos, valor de aquisição e valor de avaliação, caso exista.

Os bens públicos foram classificados pelo Código Civil Brasileiro como dominicais, de uso comum do povo e de uso especial. Ainda de acordo com a referida Lei, no art. 99:

Art. 99. São bens públicos:

I – os de uso comum do povo, tais como rios, mares, estradas, ruas e praças;

II – os de uso especial, tais como edifícios ou terrenos destinados a serviço ou estabelecimento da administração federal, estadual, territorial ou municipal, inclusive os de suas autarquias;

III – os dominicais, que constituem o patrimônio das pessoas jurídicas de direito público, como objeto de direito

pessoal, ou real, de cada uma dessas entidades. (BRASIL, 2002).

Os bens públicos possuem finalidade desde sua aquisição ou recebimento e devem ser classificados a partir de sua afetação (destinação). Conforme Di Pietro (2000 apud CARNEIRO, 2003, p. 99), necessita-se, ainda, de uma separação no tocante aos bens imóveis próprios, definidos na legislação como especiais e dominicais, e os bens de uso comum do povo, haja vista que os primeiros servem de fomento para instalação de equipamentos e prestação de serviços e o segundo integra o patrimônio com a finalidade de preservação e conservação pelo poder público.

Conforme Carneiro (2003), em virtude da não obrigatoriedade e desinteresse dos proprietários de imóveis em atualizarem os registros imobiliários, criou-se uma dicotomia entre o que existe registrado nos CRI em relação à realidade dos imóveis brasileiros, sejam eles de propriedade pública ou privada. No caso dos bens públicos, estes frequentemente são passíveis de ampliação, redução e incorporação, hajam intervenções efetuadas pelos municípios que muitas vezes não são informadas nos CRI, implicando o abismo entre a realidade e o limite legal da propriedade.

Para Haar (1992 apud ERBA, 2005, p. 24), a diferença entre o limite real e o limite de posse frequentemente resulta em litígio, custos administrativos e



judiciais para a defesa e reintegração da propriedade, o que transcreve-se a seguir:

Com relação ao cadastro, existem dois limites para as parcelas: o limite legal, definido por Haar (1992) como uma linha imaginária que não se pode localizar no terreno sem um sinal que a materialize, exigindo para sua determinação o estudo dos títulos da parcela em questão, mais os títulos das propriedades vizinhas; e o limite da posse, que é determinado pelo uso do imóvel, materializado por entes naturais ou antropológicos.

Em relação ao tópico anterior, é prudente relacionar e informar no ato de incorporação ao cadastro a diferença entre o limite legal e o limite da posse, uma vez que esses dados nem sempre são acrescidos no CRI de imediato, pois existem regras, custos e anuência para o lançamento dessas atualizações e correções.

A partir da diferença entre os limites legais e de posse, como ressaltam Carneiro (2003) e Erba (2005), implica-se também dizer que não existe a boa prática de atualização das informações nos CRI, esses últimos que exigem para fins de atualização e averbação o pagamento de taxas e emolumentos, além da responsabilidade técnica.

Em parte, essa dicotomia entre o cadastro municipal, os registros imobiliários e a realidade dos

imóveis gerou inúmeras consequências para a trajetória imobiliária histórica do Brasil, conforme explica Erba (2005, p. 25) em parte de sua pesquisa a seguir:

No Brasil, grande parte dos trabalhos de medição efetuados pelos profissionais de mensuração objetiva exclusivamente levantar os fatos existentes, determinando assim somente os limites de posse das propriedades, desconhecendo as causas legais correspondentes ao domínio efetivo. Este fato acaba provocando a generalizada e conhecida situação de confusão de limites e sobreposição de títulos de propriedade. Este sistema tem como ponto fraco a falta de precisão causada pela subjetividade que existe no momento em que se define o citado ponto de partida quando a parcela é amarrada à malha urbana. O fato de usar este tipo de referência tem causado grandes problemas nos sistemas de publicidade territorial de muitos países, gerando superposições de títulos e conflitos de limites.

Para facilitar o balanço contábil das cidades, recomenda-se incluir no ato do cadastro o valor de aquisição do bem, visando a praticidade e agilidade na prestação de contas, como também é possível associar os investimentos a cada patrimônio, facilitando a transparência dos gastos públicos.

Contribuiu-se para essa necessidade as Normas Brasileiras de Contabilidade Aplicada ao Setor Público (NBCASP), instituídas pelo Conselho Federal

de Contabilidade (CFC), que requerem das esferas públicas uma prestação de contas dos investimentos, visando a representar o valor real de cada bem público, o que também deve ser associado ao CTM dos municípios.

3. CONCLUSÃO

O Brasil está desenvolvendo sua cultura cadastral baseado nos Cadastros Técnicos Multifinalitários (CTM), mas ainda persiste em focar o aspecto tributário. Nesse aspecto, inserir o georreferenciamento dos bens públicos imóveis nos CTM deve considerar não apenas seus limites e confrontações, mas também agregar informações pertinentes ao tipo de bem a ser cadastrado, visando não apenas a sua delimitação, mas servindo para transparência dos gastos públicos.

Neste trabalho apresentou-se a oportunidade de modernizar e adequar os levantamentos e descrições dos imóveis públicos no sistema geodésico brasileiro para evitar perdas de patrimônio com sobreposição de registros imobiliários de terceiros e para uma maior integração com as plataformas brasileiras em desenvolvimento.

Cabe ressaltar que para proteger as áreas públicas, deve-se considerar cadastrá-los no mesmo ambiente dos particulares para evitar sobreposição ou incorporação no patrimônio particular. Adotando essa metodologia, os municípios poderão agregar outras informações, como os investimentos e uso atual, evitando, assim, desperdício de recurso estatal.

A história brasileira demonstra a dicotomia entre o limite real e o limite legal dos CRI. Contudo, o georreferenciamento de imóveis deve contribuir para atestar a realidade dos imóveis e servir para atualizar essas informações nos CRI, que possuem papel importantíssimo na defesa da propriedade.

REFERÊNCIAS

BRASIL. Casa Civil. Lei nº 10.406, de 10 de janeiro de 2002. Institui o Código Civil. Subchefia para Assuntos Jurídicos. Disponível em: <<http://bit.ly/1hBawae>>. Acesso em: 21 nov. 2016.

_____. Instituto Brasileiro de Geografia e Estatística. Resolução do Presidente nº 1, de 25 de julho de 2005. Altera a caracterização do Sistema Geodésico Brasileiro. Disponível em: <<http://bit.ly/2glKejh>>. Acesso em: 22 nov. 2016.

_____. Ministério das Cidades. Portaria nº 511, de 7 de dezembro de 2009. Diretrizes para a criação, instituição e atualização do Cadastro Territorial Multifinalitário (CTM) nos municípios brasileiros. Diário Oficial da República Federativa do Brasil, Brasília, DF, 8 dez. 2009, seção 1, p. 75.

CARNEIRO, Andrea Flávia Tenório. Cadastro imobiliário e registro de imóveis. Porto Alegre: Instituto de Registro Imobiliário do Brasil, 2003.

CONSELHO FEDERAL DE CONTABILIDADE (CFC). Contabilidade aplicada ao setor público. 2008. Disponível em <<http://bit.ly/1xdNMmH>>. Acesso em: 25 set. 2015.

ERBA, Diego Afonso, O cadastro territorial: passado, presente e futuro. In: ERBA, Diego Afonso; OLIVEIRA, Fabrício Leal de; LIMA JUNIOR, Pedro de Novais (Orgs.). Cadastro multifinalitário como instrumento de política fiscal e urbana. Rio de Janeiro: Studium, 2005 p. 13-40.

FONSECA, Cláudio Eduardo. A importância do cadastro tributário na arrecadação municipal e na auditoria de tributos: estudo de caso do município de Belo Horizonte. 2010. 32 p. Trabalho de Conclusão de Curso (Especialização em Auditoria em Tributos Municipais) – Faculdade de Direito da Universidade Gama Filho, Belo Horizonte, 2010. Disponível em: <<http://bit.ly/2gkA92a>>. Acesso em: 22 nov. 2016.

GALDINO, Carlos Alberto Pessoa Mello. Cadastro de parcelas territoriais vinculado ao sistema de referência geocêntrico: Sirgas 2000. 2006. 255 f. Tese (Doutorado em Engenharia Civil) – Universidade Federal de Santa Catarina, Florianópolis, 2006.

GEORREFERENCIAR. In: MICHAELIS Dicionário Brasileiro da Língua Portuguesa. 2015. Disponível em: <<http://bit.ly/2fJMzRj>>. Acesso em: 20 nov. 2016.

PIMENTEL, José; CARNEIRO, Andréa Flávia. Cadastro territorial multifinalitário em município de pequeno porte de acordo com os conceitos da portaria n. 511 do ministério das cidades. Revista Brasileira de Cartografia, Rio de Janeiro, v. 64, n. 1, p. 202-212, 2012. Edição especial.

ZOCOLOTTI FILHO, Carlos Alberto. Utilização de técnicas de poligonização de precisão para o monitoramento de pontos localizados em galerias de inspeção: estudo de caso da U. H. de Salto Caxias. 2005. 112 f. Dissertação (Mestrado em Ciências Geodésicas) – Universidade Federal do Paraná, Curitiba, 2005.

Uso de técnicas de inteligência artificial para subsidiar ações de controle



Luís André Dutra e Silva

é servidor do Tribunal de Contas da União, Bacharel em Ciência da Computação pelo UniCeub, com certificação em Engenharia de Software pelo IEEE e em Gestão de Projetos pela Universidade de Stanford.

RESUMO

Os serviços cognitivos são uma alternativa baseada em Inteligência artificial para a obtenção de soluções que são capazes de detectar padrões de qualquer tipo em textos, imagens ou qualquer outra fonte de dados. A seguir, é feita a descrição dos experimentos de utilização de técnicas de inteligência artificial em bases não estruturadas, realizados pelo TCU. As etapas do projeto são detalhadas e relacionadas às aplicações e possibilidades futuras. Os resultados desses experimentos foram muito promissores e pretende-se adotar as técnicas utilizadas nesse período em novas iniciativas tecnológicas para o TCU e para a Administração Pública em geral.

Palavras-chave: Serviços cognitivos. NLP. Inteligência artificial. *Machine Learning*. Mineração de textos.

1. INTRODUÇÃO

Grande parte das informações consultadas e produzidas pelo TCU são recebidas de órgãos jurisdicionais, registradas em relatórios e instruções processuais, votos, acórdãos, despachos e outros documentos. Esses registros são textuais e complexos e exigem recursos sofisticados de interpretação para se obter o conhecimento linguisticamente representado, especialmente em razão de os dados não estarem estruturados. Essa característica requer inúmeras análises e combinações



para exploração e agregação de valor às informações e ao processo de tomada de decisão.

Esse fato implica significativo esforço por parte dos servidores do Tribunal para estruturação desses dados e sistematização de conhecimentos para uso e tomada de decisão. Exemplo disso é o esforço requerido para a realização de determinadas atividades de gestão e controle, como o monitoramento de deliberações do TCU e as iniciativas de classificação e triagem de processos de Tomada de Contas Especial (TCE).

Nesse contexto, o uso de ferramentas e algoritmos amparados em modelos de *machine learning* para automatização da interpretação de documentos revela-se essencial e estratégico para classificação e extração automática de informações contidas em fontes de dados não estruturados.

As técnicas de inteligência artificial exploradas e sistematizadas nesse contexto permitem o aprendizado de máquina das características mais complexas de conceitos presentes em diferentes documentos. Com isso, pretende-se estruturar e tornar disponíveis e úteis informações inicialmente dispersas em diferentes documentos e formatos.

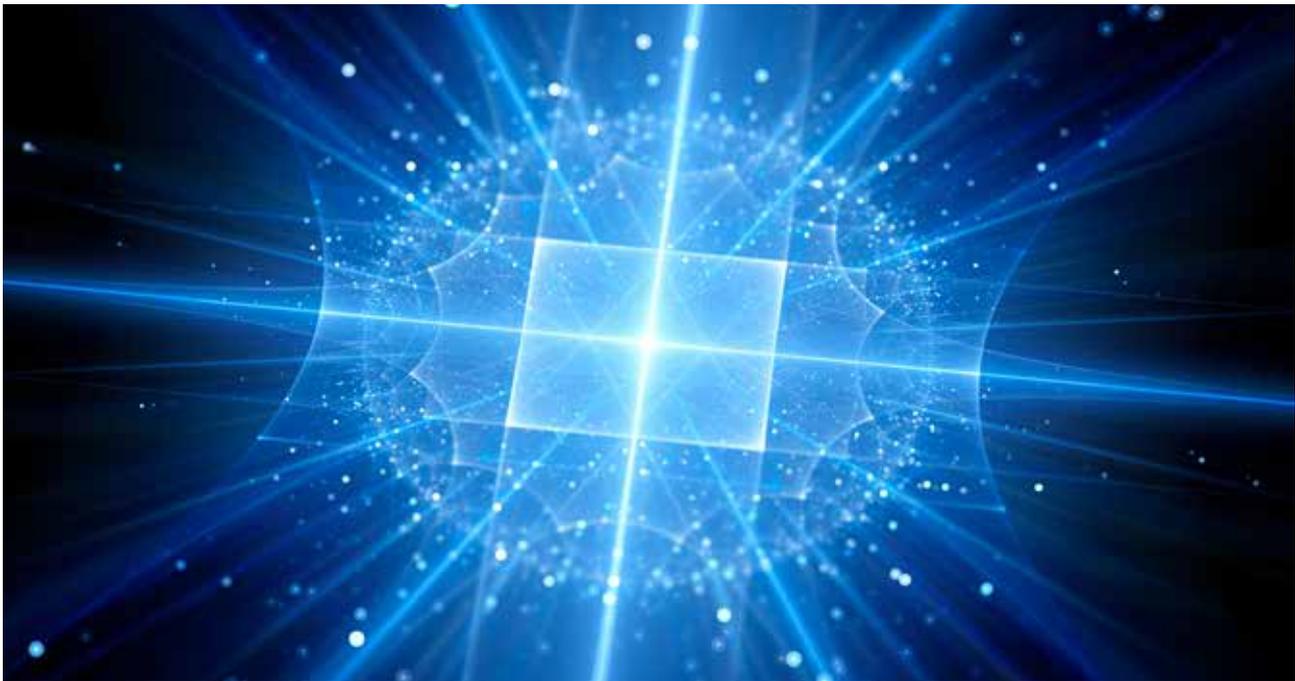
A título de informação, o desenvolvimento de experimentos iniciais com o uso de técnicas de *Deep Learning* revelou-se bastante promissor. O protótipo foi realizado no período de 1º de julho a 31 de dezembro de 2015, e a prova de conceito realizada utilizou como base de treinamento cerca de 257 mil acórdãos de 1993 a 2013 e classificou deliberações contidas no texto de

5.300 acórdãos proferidos entre 2014 e 2015. O resultado obtido revelou precisão média de mais de 96%.

2. DESENVOLVIMENTO DOS TRABALHOS

Dado o caráter exploratório e o pioneirismo de uso de técnicas de inteligência artificial no âmbito do TCU, o desenvolvimento dos trabalhos foi dividido em etapas de experimentação com foco inicial para processos de tomada de contas especial. Dado o caráter experimental, muitas das análises realizadas foram descartadas para que a melhor solução surgisse. Em que pese essa orientação, o trabalho sempre teve presente a ideia de replicação para outros objetos de controle e desafios de negócio.

A primeira etapa consistiu na obtenção dos documentos de processos de Tomada de Contas Especial diretamente do repositório do GED (Gerenciamento Eletrônico de Documentos), com o propósito de estabelecer uma base local para futura mineração das peças de TCE. O volume de documentos utilizados nessa etapa foi de cerca de 680 mil peças, e, para a extração deles foi utilizado um algoritmo em linguagem Python para acesso ao banco de dados de desenvolvimento e ao GED em produção com a finalidade de realizar download de todas as peças em pastas relativas a cada processo de TCE. Essa etapa foi realizada no período de 20 a 24 de fevereiro de 2016, e o principal desafio enfrentado foi o fato de que o download das peças foi frequentemente interrompido devido a problemas de conexão, e não foi possível, no



Word; foram realizadas extrações de informações nesses relatórios que permitiram preencher com boa precisão uma minuta de instrução para o tipo de irregularidade “Omissão no dever de prestar contas”; foram realizados cruzamentos das informações extraídas com sistemas estruturados, tais como o Sisobi, Siconv, Siafi; as mesmas extrações realizadas em documentos em formato Word passou a ser realizada também para os relatórios digitalizados em PDF, que são de menor qualidade devido ao processo de OCR; foi disponibilizada uma interface de validação do extrator para os especialistas em TCE. As extrações realizadas por esse experimento foram consideradas úteis na gestão do estoque de TCE. Segundo os especialistas, o trabalho que é realizado de forma manual para este fim pode ser parcialmente automatizado, reduzindo, assim, o tempo necessário para a distribuição/triagem de TCE. No entanto, a geração automática de minutas foi considerada problemática pelos especialistas devido ao fato de apresentarem dados que não preenchem totalmente as necessidades dos auditores que instruem TCE.

Na oitava etapa, realizada entre 27 de abril e 25 de maio de 2016, foram feitas as primeiras entregas com o propósito de disponibilizar as primeiras versões dos serviços previstos no cronograma do projeto. Conforme cronograma do projeto de criação de serviços cognitivos, estão disponíveis para validação e acesso os seguintes produtos em sua primeira versão: serviço REST de extração de deliberações de acórdãos; interface de validação

do Extrator de Deliberações; serviço REST de extração de entidades de relatórios de TCE; interface de validação do Extrator de Conteúdo de TCE.

A nona etapa, realizada entre 8 e 15 de julho de 2016, teve como propósito a melhoria de documentos digitalizados por meio da utilização de serviços terceirizados. Foi utilizada uma amostra de dez relatórios com grandes problemas de digitalização, e a API Google Vision foi utilizada para essa tentativa. O desafio principal era melhorar a qualidade do texto extraído dos documentos digitalizados de TCE sem recorrer a novo processo de digitalização. No entanto, o resultado alcançado não foi melhor, em termos de quantidade de erros, do que o software de OCR adotado no âmbito do TCU (Adobe).

A décima etapa, entre 16 e 25 de julho de 2016, teve como propósito o desenvolvimento da primeira versão do serviço de reconhecimento de entidades mencionadas, NER (*Named Entity Recognition*), capaz de extrair de qualquer texto os nomes de pessoas físicas, pessoas jurídicas, CPF, CNPJ e referências normativas. O volume de documentos utilizados para treinar os modelos neurais de reconhecimento de entidades foi de cerca de 58 mil acórdãos e a base Amazônia da Universidade de Lisboa, contendo as anotações manuais de cada tipo de entidade mencionada. Para esse fim, foi utilizado o framework de *machine learning* Apache OpenNLP em Python e Java, com anotações manuais e automáticas em textos de acórdãos. Com a utilização de algoritmos de *machine learning* no lugar de regras pré-definidas, a precisão e as demais

métricas associadas podem não ser as ideais enquanto não houver uma grande quantidade de textos anotados manualmente. A primeira versão do extrator de entidades genérico foi exposta como serviço web na infraestrutura de produção. Como foi separada uma base de validação não treinada de 10 mil sentenças aproximadamente para avaliar a qualidade do serviço web NER, foi obtido um score F1 de cerca de 81% para pessoas físicas, o que corresponde ao estado da arte em termos de extração de entidades mencionadas. A métrica F1 é a mais adequada para demonstrar o equilíbrio entre precisão e recall, que deve ser o objetivo a ser buscado.

Na décima primeira etapa, entre 1º de agosto e 14 de setembro de 2016, ocorreu o desenvolvimento da segunda versão do extrator de entidades genérico em Java puro, utilizando nativamente a biblioteca OpenNLP e os modelos treinados anteriormente para a primeira versão do NER. O principal desafio foi o fato de que a portabilidade de código Python para Java nem sempre é possível em muitas customizações. Portanto, essa iniciativa poderia não ser bem-sucedida caso um ou mais frameworks utilizados em Python não existissem de forma similar em Java. Não obstante, o serviço NER, de acordo com a arquitetura padrão de serviços do TCU, Arquitetura de Referência 8, foi desenvolvido e encontra-se em produção no ambiente JBoss 6 EAP.

A décima segunda etapa, ocorrida entre 16 de setembro de 2016 e 11 de outubro de 2016, teve como objetivo a construção de um serviço capaz de detectar possíveis erros materiais em acórdãos proferidos pelo TCU antes de serem oficializados. Foram utilizados cerca de 600 documentos para o teste da ferramenta dentre cerca de 1 mil acórdãos apontados por uma inspeção realizada pela Corregedoria do TCU como possuidores de erros materiais em sua elaboração. O resultado alcançado pelo serviço cognitivo foi uma captura de cerca de 40% dos erros materiais presentes nos acórdãos, após a verificação da grafia correta dos nomes dos responsáveis, CPF e CNPJ inexistentes ou pertencentes a pessoas físicas já falecidas ou a pessoas jurídicas inativas.

A décima terceira etapa, em andamento desde 18 de outubro de 2016, tem como objetivo a disponibilização de um serviço de extração automática de ontologias em formato OWL, de documentos de texto, utilizando algoritmos de *machine learning*. São utilizados cerca de 280 mil documentos de jurisprudência do TCU. É utilizada a técnica *Latent Semantic Indexing* realizando a decomposição da matriz W de documentos \times conceitos, obtida pela técnica *Bag of Words* e normalização TF-IDF, usando *Singular Value Decomposition*. *Latent Semantic*

Indexing é um método estatístico que liga termos em uma estrutura semântica útil, sem análise sintática ou semântica e sem intervenção manual. Usando esse método, cada documento é representado não por termos, mas por conceitos que são verdadeira e estatisticamente independentes de uma forma que os termos não são.

3. CONCLUSÃO

Os serviços cognitivos desenvolvidos podem ser utilizados de forma bem-sucedida por outros sistemas, com a finalidade de aprimorar os processos de trabalho do TCU e da Administração Pública, pois é inerente ao trabalho do TCU e outros órgãos a necessidade de estruturar textos produzidos continuamente e que dependem da classificação e extração das informações contidas nessas bases não estruturadas.

Portanto, a disponibilização desses serviços pode servir para conferir maior exatidão ao trabalho de instrução de processos, bem como para colocar à disposição, para os profissionais de Controle Externo, contextos pertinentes a todo trabalho desenvolvido de forma textual.

Ao elaborar determinado relatório, por exemplo, toda a jurisprudência existente sobre o tema desenvolvido poderia ser automaticamente relacionada ao texto nascente de forma a simplificar o processo de busca por informações relevantes ao conteúdo a ser elaborado. Outro exemplo de uso seria a elaboração de resumos de textos recebidos de órgãos jurisdicionados de forma automática, para agilizar o processo de análise desses documentos externos.

Além disso, podem-se elaborar, por meio de serviços cognitivos, clippings personalizados a respeito de cada assunto tratado em determinado trabalho. Assim, os auditores terão sempre as informações mais atualizadas que subsidiem a elaboração dos relatórios de fiscalização e todos os demais documentos necessários.

REFERÊNCIAS

- NOVELLI, Andreia, OLIVEIRA, José. Simple Method for Ontology Automatic Extraction from Documents. *International Journal of Advanced Computer Science and Applications*. Vol 5. No. 12. 2012
- MADDI, Govind, VELVADAPU, Chakravarthi. Ontology Extraction from text documents by Singular Value Decomposition. *ADMI*. 2001
- BERRY, Michael, DUMAIS, Susan, O'BRIEN, Gavin. Using Linear Algebra for Intelligent Information Retrieval. *SIAM Review*. Vol 37. No.4, pp-573-595, December 1995.