

Estudo do ecossistema de Big Data para conciliação das demandas de acesso, por meio da representação e organização da informação

Rogério Henrique de Araújo Júnior

Coordenador Adjunto do Programa de Pós-Graduação em Ciência da Informação da Universidade de Brasília, Doutor em Ciência da Informação pela mesma instituição, <http://lattes.cnpq.br/4752236489901895>, araujojr@unb.br.

Renato Tarciso Barbosa de Souza

Coordenador do Programa de Pós-Graduação em Ciência da Informação da Universidade de Brasília, Doutor em História Social pela Universidade de São Paulo, <http://lattes.cnpq.br/9941441906608746>, renasou@unb.br.

Submetido em: 22/08/2017. Aprovado em: 24/10/2017. Publicado em: 22/02/2018.

RESUMO

Trata do ecossistema de *Big Data* como subsídio no processo de tomada de decisão. Demonstra que todas as etapas do ciclo documentário integram o *Big Data*, em especial a dedicada à organização da informação, que concilia as demandas de acesso por meio da representação da informação. Apresenta a aplicação do estudo no ambiente da Secretaria de Inspeção do Trabalho (SIT) do Ministério do Trabalho (MTB), com o uso de técnicas e métodos como o diagnóstico organizacional, o Business Intelligence (BI) e o Grupo Cooperativo de Trabalho. A conclusão do estudo aponta formas de melhoria da gestão dos processos no âmbito da administração pública brasileira.

Palavras-chave: Ecossistema de Big Data. Organização da informação e do conhecimento. Representação da informação. Gestão da informação. Sistemas de informação.

Study of the Big Data ecosystem to conciliate access demands, through the representation and organization of information

ABSTRACT

It deals with the Big Data ecosystem as a subsidy in the decision-making process. It demonstrates that all stages of the documentary cycle integrate Big Data, especially that dedicated to the organization of information, which reconciles the demands of access through the representation of information. It presents the application of the study in the environment of the Labor Inspection Secretariat (SIT) of the Ministry of Labor (MTB), using techniques and methods such as organizational diagnosis, Business Intelligence (BI) and Cooperative Work Group. The conclusion of the study points to ways of improving the management of processes within the Brazilian public administration.

Keywords: *Big Data ecosystem. Organization of information and knowledge. Representation of information. Information management. Information systems.*

Estudio del ecosistema Big Data para conciliar las demandas de acceso, a través de la representación y organización de la información

RESUMEN

Se trata del ecosistema de Big Data como subsidio en el proceso de toma de decisión. Demuestra que todas las etapas del ciclo documental integran el Big Data, en especial la dedicada a la organización de la información, que concilia las demandas de acceso por medio de la representación de la información. Se presenta la aplicación del estudio en el ambiente de la Secretaría de Inspección del Trabajo (SIT) del Ministerio de Trabajo (MTB), con el uso de técnicas y métodos como el diagnóstico organizacional, el Business Intelligence (BI) y el Grupo Cooperativo de Trabajo. La conclusión del estudio apunta formas de mejora de la gestión de los procesos en el ámbito de la administración pública brasileña.

Palabras clave: Ecosistema de Big Data. Organización de la información y del conocimiento. Representación de la información. Gestión de la información. Sistemas de información.

INTRODUÇÃO

O estudo do ambiente de facilidades técnicas para armazenamento, processamento e entrega de dados é pressuposto para a gestão da informação e do conhecimento nas organizações. Este ambiente é chamado de ecossistema de *Big Data*. O armazenamento de grandes conjuntos de dados nas organizações em uma plataforma de compartilhamento, como o modelo de armazém de dados ou *datawarehouse*¹, tem trazido novos desafios para a coleta, guarda e disponibilização de dados e informações a serem utilizados no processo de tomada de decisão nas organizações, sobretudo no que diz respeito à exploração dos dados. Segundo Inmon, Terdeman e Imhoff (2001), o processo de exploração consiste na formulação de hipóteses que são aplicadas ao ambiente de dados e tem como fator crítico de sucesso o agrupamento de dados detalhados relevantes e integrados.

Outro aspecto relacionado ao tema diz respeito à organização e representação da informação, o qual enseja uma série de questionamentos quanto aos modelos de armazenamento de dados e de acesso. Há que se considerar, também no âmbito do ecossistema de *Big Data*, o processo de descrição dos dados disponibilizados com aplicação de regras de negócio para grandes transferências.

Considerando que o *Big Data* aporta grande volume de dados estruturados e desestruturados para o processo decisório, a gestão deste aporte deve permitir às organizações condições ideais para análise e julgamento precisos, que subsidiarão decisões que contribuam com o desenvolvimento e aprimoramento corporativo.

Há um fato que se impõe aos sistemas de informação que é a exigência crescente do emprego de imagens que ilustrem o uso de dados por meio de gráficos, esquemas ou tabelas em que possam incidir análise. Esta situação é o que podemos denominar análise visual de dados. No ecossistema de *Big Data*, a possibilidade de visualização dos dados provenientes das diversas fontes disponíveis tem como consequência a valoração das informações a serem usadas no processo decisório nos ambientes corporativos.

O reúso de dados sob a égide de novos pressupostos em torno da solução de problemas traz nova perspectiva para o processo de tomada de decisão, daí o papel da visualização analítica proposta por Aguilar, Pinto Semeler e Soares (2017), segundo a qual o valor que pode ser extraído dos armazéns de dados exige novas formas de visualização, que trazem para a atividade de análise o suporte de tecnologias e soluções que apontam para um campo de trabalho denominado Ciência dos Dados (*Data Science*). Esse campo configura-se como uma oportunidade para a exploração de

¹ Datawarehouse: coleção de dados orientada ao assunto, integrada, não volátil e variante em tempo (INMON, TERDEMAN e IMHOFF, 2001).

grandes volumes de dados por reunir profissionais e desenvolvedores focados na construção de algoritmos para sua visualização.

Diante desse cenário, os aspectos caros ao *Big Data*, ou seja, I. volume que se refere a aplicação de regras de negócio na transferência de dados; II. velocidade que diz respeito a rapidez na fluência dos dados que permite o tratamento em tempo hábil de acordo com as exigências dos usuários e III. variedades dos dados gerados em diferentes tipos de formatos, estruturados, numéricos em bancos de dados tradicionais, textos, entre outros, devem ser tratados como um processo típico de gestão estratégica da informação.

Os pressupostos de pesquisa que apoiam a compreensão do tema exposto neste artigo podem ser assim formulados:

- construção de um modelo de estruturação do ecossistema de *Big Data* deve estar calcada no mapeamento e aplicação de regras de negócio para as transferências de dados;
- plena garantia de acesso às facilidades de armazenamento, processamento e entrega de dados é parte da conciliação das demandas de acesso, por meio da representação e organização da informação; e
- definição de um modelo de organização da informação e do conhecimento é determinante para a construção de uma solução de *Big Data*, o que permitirá o gerenciamento de grande volume de dados estruturados e desestruturados úteis ao processo corporativo de tomada de decisão.

A coleta, o tratamento e a análise da grande quantidade de dados disponíveis para uso das organizações são desafios à efetividade da decisão dos gestores públicos. A situação se complica de acordo com a exigência de requisitos tecnológicos existentes, dentre eles:

- advento do documento eletrônico;
- bases de dados públicas, abertas ou restritas, sobrecarregadas de dados;

- grande quantidade de informações não estruturadas disponíveis em linha (on line); e
- sistemas de informação gerencial antiquados (ARAÚJO JÚNIOR, SOUSA e ALBUQUERQUE, 2015).

Diante disso, as possibilidades de integração dos aspectos tecnológicos e de insumos informacionais pertencentes a um ecossistema de *Big Data* é o que se pretende estudar neste artigo na ambiência da Secretaria de Inspeção do Trabalho (SIT) do Ministério do Trabalho (MTB).

ECOSSISTEMA DE BIG DATA

Big Data é um conjunto de tecnologias voltadas para a gestão de grandes volumes de dados em diferentes formatos (RAMOS, 2015). Uma das suas características é possibilitar a gestão dos dados em grande velocidade de processamento, mas não se refere especificamente à sua organização, nem a rastreabilidade dos dados processados. O conceito vem evoluindo nos últimos dez anos e a sua origem está no *Business Intelligence* (BI), que é o resultado do processo de mineração de dados analíticos voltados para a combinação de dados, aporte de comentários, formatação de metadados e de conteúdo a serem usados no processo decisório. Para Barbieri (2001), o BI é o uso de diversas e variadas fontes de informação que as organizações utilizam para definir estratégias e subsidiar processos organizacionais. Nesse aspecto, os conceitos de *Big Data* e BI são convergentes, levando alguns pesquisadores e técnicos a considerar as duas concepções como uma apenas. Para diversos autores, tais como Davenport (2014), Garcia Marquez e Lev (2016) e Inmon e Linstedt (2014), os conceitos são similares, pois se ajustam em modelos que envolvem captura, apresentação e geração de relatórios para a decisão. Todavia, nenhum autor aponta para a organização dos dados armazenados.

O *Big Data* possui cinco pressupostos básicos que permitem a modelagem dimensional de dados e sua posterior recuperação. Os cinco pressupostos ou 5Vs, de acordo com Walker (2015), são:

- volume: permite a modelagem e análise de extensos volumes de dados;
- velocidade: fornece resultados de pesquisa com agilidade, velocidade em um lapso de tempo compatível com a demanda;
- variedade: processa e fornece conjuntos de dados em diferentes tipos e formatos informacionais;
- veracidade: processa e fornece conjuntos de dados fidedignos, ou seja, íntegros e fiéis aos dados processados e armazenados nos bancos de dados; e
- valor: agrega valor informacional como subsídio ao processo decisório, provendo-o com informação de valor claro para as organizações.

O valor reveste a informação de uma importância estratégica como subsídio ao processo de tomada de decisão, ilustrando com exatidão a proposta de pesquisa que visa estudar o ecossistema de *Big Data* como provedor de informações com valor agregado mensurável e de rastreabilidade que acrescentam relevância aos conjuntos de dados. Cabe ressaltar que essa afirmação pode ser entendida como uma condicionante operacional para a construção de sistemas informacionais baseados em grandes conjuntos de dados.

Como complemento ao entendimento dos cinco pressupostos do *Big Data*, sobretudo no que se refere ao item valor, apresentamos a tabela 1 com as diferenças básicas que um processo de BI enxerga entre dados operacionais e dados informacionais, segundo Barbieri (2001), as quais podem ser adaptadas na compreensão do ecossistema de *Big Data*.

Tabela 1 – Comparação entre dados de natureza operacional e informacional

Características	Dados operacionais	Dados informacionais
Conteúdo	Valores correntes	Valores sumarizados, calculados e integrados de várias fontes
Organização dos dados	Por aplicação nos sistemas de informação	Por assuntos/negócios
Natureza dos dados	Dinâmica	Estática até a atualização dos dados
Formatos das estruturas	Relacional, próprio para computação transacional	Dimensional e simplificado, próprio para atividades analíticas
Atualização dos dados	Atualização campo a campo	Acesso, sem atualização
Uso	Altamente estruturado com processamento repetitivo	Desestruturado com processamento analítico e heurístico
Tempo de resposta	Otimizado em tempo de resposta	Análises mais complexas com tempos de respostas maiores

Fonte: Barbieri, 2001

A comparação entre dados operacionais e dados informacionais aponta para os procedimentos a serem desenvolvidos para trazer ao ecossistema de *Big Data* o foco da informação com valor agregado, que consubstancia a importância do estudo do fenômeno da explosão da informação, bem como responde à questão de pesquisa: Como retirar de grandes quantidades de dados informação útil para a tomada de decisão nas organizações?

O ecossistema de *Big Data* também é conceituado como um ambiente composto por um complexo de facilidades técnicas para armazenamento, processamento e recuperação de dados (GLASS e CALLAHAN, 2015). No estudo do ecossistema de *Big Data* para a conciliação das demandas de acesso, por meio da representação e organização da informação, as características a serem consideradas para a delimitação do escopo da pesquisa serão aquelas propostas por Barbieri (2001), para diferenciar dados operacionais de dados informacionais: conteúdo, organização dos dados, natureza dos dados, formatos das estruturas, atualização dos dados, uso e tempo de resposta.

PROCESSO DE TOMADA DE DECISÃO

O processo para a tomada de decisão enseja ampla consideração das variáveis que influenciam um problema. A decisão precisa ser fruto de um amadurecimento na análise de alternativas de solução e, ao mesmo tempo, espelha o estilo gerencial da liderança responsável pelas decisões. Segundo Gomes, Gomes e Almeida (2009), as decisões podem ser classificadas como simples, complexas, específicas ou estratégicas, e podem apresentar consequências imediatas, de curto ou longo prazo.

Considerando que o processo de tomada de decisão envolve o exame das alternativas de solução, o decisor tem que conhecer com precisão o problema e as suas consequências para estruturar uma decisão que influenciará positivamente o desenvolvimento organizacional.

Segundo Araújo Júnior (1998), a decisão funciona como uma espécie de termômetro da efetividade gerencial, pois coloca à prova a capacidade de resposta da empresa ante os desafios e problemas, assim como a capacidade de apresentar soluções viáveis que propiciem à organização uma visão futura do seu desenvolvimento, ou seja, considera o problema como fruto de uma teia de relações, e não como um fato isolado ou até mesmo casuístico.

Para a consecução de um processo de importância estratégica para as organizações, o processo decisório deverá estar apoiado em insumos informacionais que permitam ao gestor conhecer os ambientes organizacionais para melhor compreender o problema e suas consequências. De outro modo, não há como tomar uma decisão racional e consequente sem que o decisor tenha acesso à informação. Este desafio está na base da utilidade de um ecossistema de *Big Data* para as organizações, ou seja, prover a decisão baseada em informação de valor agregado.

O processo decisório conta com poderosa ferramenta na sua otimização: a informação. Ela é gerada a partir dos dados coletados, sem, contudo, limitar-se a eles. Na verdade informação são dados coletados, organizados, ordenados, aos quais são atribuídos significados e contextos (McGEE e PRUSAK, 1994).

Considerando que o melhor subsídio para o processo decisório é a informação, passamos à consideração de um modelo de gestão de dados e informações com três condicionantes básicos: coleta de dados, processamento e disseminação da informação por meio de produtos e serviços informacionais. Entretanto, os gestores ainda enxergam a informação como um insumo complexo e muitas vezes com valor intangível. Para Davenport (2000), os gestores adeptos de uma abordagem ecológica da informação acreditam que a informação não é facilmente arquivada em computadores; quanto mais complexo o modelo de informação (arquitetura de informação), menor será a sua utilidade; a informação pode ter muitos significados em uma organização; e a tecnologia é apenas um dos componentes do ambiente informacional e não se apresenta como meio adequado para operar mudanças na organização.

A partir dessas considerações, fica claro que o papel da informação na otimização da decisão deverá ser promovido por meio de um sistema que é considerado como uma estrutura interativa de dados sobre a organização, cujo objetivo é servir como base de conhecimento para consultas que proporcionem ao gestor maior clareza no processo de tomada de decisão (ARAÚJO JÚNIOR, 1998).

No processo decisório, a informação é um recurso estratégico e, para esse intento, o ecossistema de *Big Data* surge como solução, já que incorpora todas as técnicas e ferramentas de processamento de grandes quantidades de dados e se comporta como um típico sistema de informação gerencial nas respostas aos gestores.

Araújo Júnior (1998) afirma que a relevância da informação pode ser também compreendida em sua influência na estratégia competitiva, onde uma organização configura suas atividades comerciais e apresenta produtos e serviços diferenciados baseados na consideração das necessidades dos clientes, fato que reforça e integra a visão da importância do *Big Data* e dos sistemas de apoio à decisão nas organizações.

McGee e Prusak (1994) apresentam três perspectivas relacionadas à informação, quando disponíveis em um sistema de busca e recuperação:

- informação para a definição da estratégia: considera o ambiente de competição da organização. Presta-se a identificar ameaças e oportunidades para a empresa, além de facilitar a construção de um cenário propício a resposta competitiva mais eficaz;
- informação e execução da estratégia: a oferta de produtos e serviços está baseada na elaboração de processos baseados na tecnologia. A informação para a execução estratégica constitui-se em uma das ferramentas para a diferenciação de produtos e serviços; e
- informação e integração: os indicadores de desempenho que a informação fornece são importantes para o reposicionamento competitivo das organizações.

ORGANIZAÇÃO E RECUPERAÇÃO DA INFORMAÇÃO

A organização e a recuperação da informação são, respectivamente, o processo e a etapa final do ciclo documentário. Para Café e Sales (2010), a organização da informação é o processo de arranjo de acervos tradicionais ou eletrônicos realizado por meio da descrição física e de conteúdo (assunto) de seus objetos informacionais. A recuperação da informação, segundo Araújo Júnior (2007), é o processo de localizar documentos e itens de informação que tenham sido objeto de armazenamento, com a finalidade de permitir o acesso dos usuários aos itens de informação, objetos de uma solicitação.

A relação que se estabelece entre a organização da informação e a recuperação da informação está ancorada na relação direta que há entre o arranjo de acervos por meio da descrição física e de conteúdo dos objetos informacionais e a posterior localização de documentos e itens de informação armazenados em um sistema.

De todo o modo, cabe a observação de Araújo Júnior (2007), que complementa o entendimento da relação entre a organização da informação e sua recuperação, ao afirmar que o monitoramento sistemático das necessidades dos usuários tem relação direta com os mecanismos de busca, que têm como resultado informações em conformidade com os requisitos predeterminados.

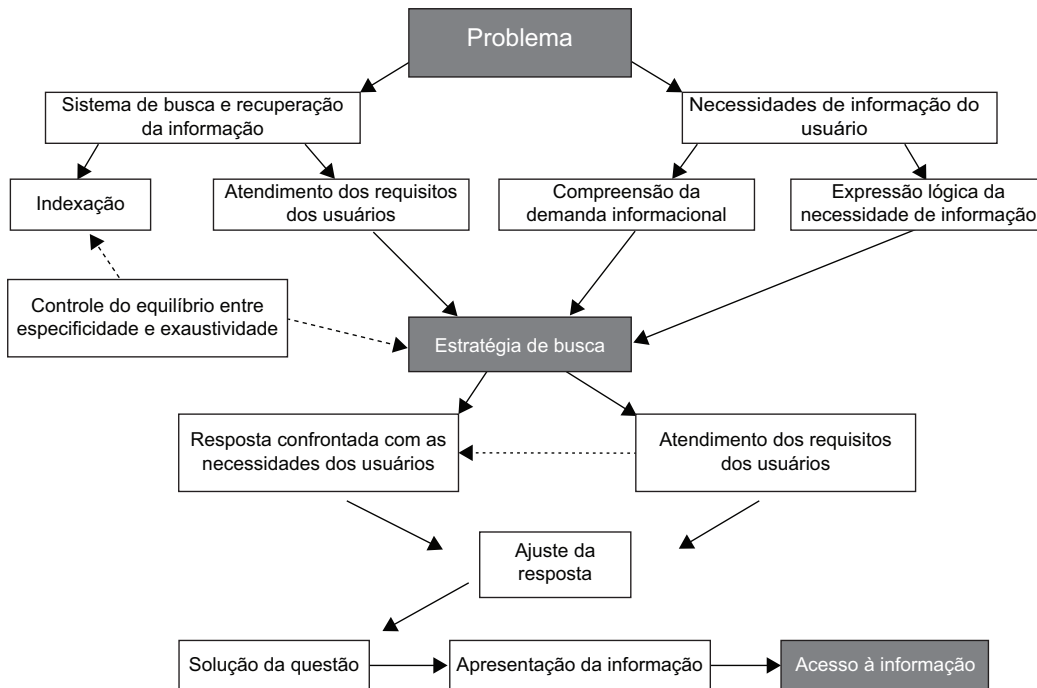
Para Guedes e Araújo Júnior (2014), o processamento da informação será mais eficiente se o armazenamento tiver sido de melhor qualidade, e os produtos resultantes deste processamento deverão ser apresentados na linguagem adequada ao usuário (o que equivale à codificação apropriada para cada destinatário). Os autores acrescentam ainda que a atenção com o ruído, redundância, canal de comunicação e codificação tem reflexos positivos na etapa de interrogação e busca, porque o sistema está em boa forma, organizado e livre de informações desnecessárias ou distorcidas, facilitando a recuperação da informação, que é a própria finalidade do trabalho documentário.

O ciclo documentário está subdividido em 12 operações, conforme enumera Robredo (2005):

- i. seleção: escolha, dentre um conjunto de documentos, daqueles que podem responder às disciplinas ou tópicos cobertos por dado serviço de documentação;
- ii. aquisição: atividade administrativa para adoção de providências de compra, intercâmbio, doação, pagamento e controle de chegada;
- iii. registro ou tombamento: atividade administrativa na qual são atribuídos números de registro para controle da incorporação dos documentos aos acervos das unidades informacionais;
- iv. descrição: atividade de processamento técnico relacionada com o registro das características descritivas dos documentos;
- v. análise ou condensação: etapa de processamento técnico em que são desenvolvidas atividades de sumarização do conteúdo dos documentos, por meio de número limitado de sentenças ou frases que devem expressar a essência temática do documento;
- vi. indexação: atividade de processamento técnico destinada à identificação dos conceitos de que trata o documento, expressando-os na terminologia utilizada pelo profissional da informação em linguagem natural ou com a ajuda de vocábulos ou termos derivados de linguagens documentárias;
- vii. armazenamento dos documentos: etapa em que os documentos, a partir de dados que os descrevam por meio de sistemas de classificação, são armazenados de modo a serem encontrados quando solicitados pelos usuários nas unidades informacionais;
- viii. armazenamento da representação condensada dos documentos: etapa onde os elementos descritivos do conteúdo dos documentos, bem como os números que lhes foram atribuídos na etapa de registro ou tombamento, constituem um conjunto de dados que caracteriza, de forma condensada, cada documento;
- ix. processamento da informação condensada: atividade em que os registros que contêm os elementos representativos dos documentos são submetidos a diversos processamentos manuais ou automatizados;
- x. produtos do processamento: etapa onde são formados os produtos que vão compor o portfólio das unidades informacionais. Vários podem ser os serviços documentários a serem ofertados aos usuários;
- xi. interrogação e busca: etapa na qual a interação entre a gestão do ciclo documentário e a realização de pesquisa nos acervos se dá por uma ação de comunicação e mediação entre os profissionais da informação e os usuários da unidade informacional onde o ciclo documentário se desenvolve; e
- xii. recuperação da informação: esta etapa constitui a finalidade do trabalho documentário. Ela é o resultado das operações realizadas no processo de busca.

É possível entender que a organização da informação deverá ser aplicada como subsídio para a definição do tratamento dos dados do *Big Data*, visando a recuperação da informação demandada. Essa relação subsidia a integração do ecossistema de *Big Data* para a conciliação das demandas de acesso, por meio da representação e organização da informação. A figura 1, que trata do processo de busca e recuperação da informação, traz uma visão geral dos elementos constituintes desse processo, além de colocar o problema como ponto de partida para a organização da informação que deverá apontar para o seu acesso.

Figura 1– Elementos constituintes do processo de recuperação da informação



Fonte: adaptado de Araújo Júnior, 2007

SOBRE O AMBIENTE DE ESTUDO E APLICAÇÃO

O desenvolvimento da pesquisa tem como ambiência a Secretaria de Inspeção do Trabalho (SIT) do Ministério do Trabalho (MTB), especificamente na atividade de inspeção do trabalho, cuja finalidade é a fiscalização do Fundo de Garantia por Tempo de Serviço (FGTS), que é feita por meio de auditorias realizadas pelos auditores-fiscais do trabalho (AFTs) junto aos empregadores em todo o território nacional.

As inspeções são realizadas no local de trabalho e por meio das análises de documentos como folhas de pagamentos, registro de ponto, livros fiscais, sistemas de informação e dados disponíveis sobre as atividades laborais desenvolvidas. São também consultados dados e informações existentes em diversos sistemas institucionais informatizados que compõem o ecossistema de *Big Data*, tais como a Relação Anual de Informações Sociais (Rais), Cadastro Geral de Empregados e Desempregados (Caged), Seguro-Desemprego, Sistema Empresa de Recolhimento do FGTS e Informações à Previdência Social

(Sefip), Base de Dados da Receita Federal, assim como informações disponibilizadas pela Caixa Econômica Federal (CAIXA) referentes a recolhimentos realizados de FGTS, por meio do Sistema do Fundo Garantidor de Crédito (FGC). Também são utilizadas pela fiscalização informações referentes às confissões de débito feitas pelos empregadores em procedimento administrativo próprio e notificações de débito emitidas pelo Ministério do Trabalho. A CAIXA disponibiliza periodicamente para a SIT arquivo com informações referentes aos recolhimentos de FGTS relativos a toda a base de empregadores (Arquivo SADO), para alimentação de sistema institucional utilizado para planejamento fiscal e, também, como fonte de consulta dos auditores-fiscais do trabalho quando da fiscalização.

A integração da informação, quando concretizada, possibilitará o planejamento eficaz das ações de fiscalização e auditoria. Na SIT, o acompanhamento da arrecadação do FGTS e de outros direitos trabalhistas deve ser realizado por meio da integração de várias bases de dados, permitindo o uso e o compartilhamento

das informações disponíveis no Sistema da Receita Federal do Brasil, Receitas estaduais e municipais, além das informações gerenciadas pelas outras Secretarias do próprio Ministério do Trabalho, o que enseja o estudo da viabilidade de implementação de uma visão de ecossistema de bases a ser disponibilizada e efetivamente usada no âmbito do *Big Data*.

É importante ressaltar que o Ministério do Trabalho é ente participante do desenvolvimento do eSocial, sistema de escrituração digital da administração pública, cujo objetivo é estruturar todas as informações relacionadas ao trabalhador, as quais atualmente estão reunidas por meios múltiplos e em diferentes plataformas, muitas vezes protegidas por sigilo. O acesso não autorizado, a disponibilização voluntária ou acidental da senha de acesso ou de informações e a quebra do sigilo constituem infrações ou ilícitos que sujeitam o usuário à responsabilidade administrativa, penal e civil (MINISTÉRIO DO TRABALHO, 2017). Atualmente, o eSocial opera com o módulo do empregador doméstico e armazena todos os eventos trabalhistas, inclusive as folhas de pagamento dos empregados, documento essencial para a fiscalização do FGTS.

Percebe-se que a auditoria e fiscalização do FGTS é tarefa bastante complexa, que exige a análise de um volume de dados considerável, provenientes de diversas fontes. Assim como temos desperdício de material de construção, desperdício de alimentos, desperdício de água, desperdício de energia, é possível verificar empiricamente que há grande desperdício de informação. Isso tem custo muito alto, seja pela duplicidade de esforços, de recursos humanos, materiais e financeiros, o que afeta a eficiência e eficácia da auditoria do FGTS. A pergunta é: como podemos tornar a auditoria eficiente e eficaz, com rapidez, qualidade e a baixo custo com uso de um modelo de ecossistema de *Big Data* para conciliação das demandas de acesso, por meio da representação e organização da informação?

Não obstante, é preciso considerar que os desafios inerentes à gestão dos dados são análise, captura, curadoria, pesquisa, compartilhamento, armazenamento, transferência, visualização e informações sobre privacidade dos dados. Maior precisão nos dados pode levar à tomada de decisões com mais confiança. Além disso, melhores decisões podem significar maior eficiência operacional, redução de risco e redução de custos.

ETAPAS METODOLÓGICAS

O estudo apresentado neste artigo é de caráter descritivo, isto é, os fatos foram diagnosticados, analisados, registrados, classificados e, por fim, interpretados. O ambiente de aplicação permitiu a consideração de um fenômeno inserido em contexto organizacional real, como foi o caso do Centro de Gestão da Informação, Conhecimento e Análise da SIT do MTB, por isso a sua adoção como estratégia para o estudo do ecossistema de *Big Data*.

Segundo Yin (2005), o estudo de caso é uma forma válida e distinta de estudo e investigação, ao considerar situação em que há mais variáveis de interesse do que pontos de dados; presença de várias fontes de evidências; e desenvolvimento prévio de proposições teóricas para conduzir a coleta de dados. Esse foi o caso em questão na SIT.

O estudo foi subdividido em quatro etapas que estão descritas a seguir.

Etapas 1: consistiu na organização e representação da informação que envolveu o diagnóstico das fontes de informação para o trabalho de fiscalização da SIT. Simultaneamente, foi realizado o levantamento preliminar dos modelos: tipos e grandeza de dados, definição do modelo de organização dos dados, construção do modelo da arquitetura para captura e uso dos dados, assim como aplicação física da arquitetura com utilização de tecnologias *Big Data*.

A técnica utilizada foi a modelagem para *Business Intelligence* (BI), que é advinda da engenharia de *software* com aplicabilidade no desenvolvimento corporativo, que permitiu o uso de diversas fontes

de informação como insumo para a formulação de estratégias, possibilitando o diagnóstico previsto nessa etapa e a definição de regras para a formatação de grandes volumes de dados. Segundo Araújo Júnior, Sousa e Albuquerque (2015), o manuseio de repositórios menores e específicos de informação evita a busca e recuperação em “montanhas” de dados, muito mais complexa no seu manejo.

Etapa 2: consistiu na modelagem da informação, que envolveu o levantamento preliminar dos silos de dados de incorporação, determinação das necessidades de informação e construção da gestão de sua distribuição. O método utilizado foi o JAD, sigla para expressão da língua inglesa que pode ser interpretada como “juntos para construir aplicações”. Esse método é largamente utilizado na etapa de levantamento de requisitos para a construção de sistemas informatizados, sendo que mais recentemente passou a ser utilizado também para a construção de outros elementos, tais como os planos estratégicos das organizações (ARAÚJO JR., SOUSA e ALBUQUERQUE, 2015).

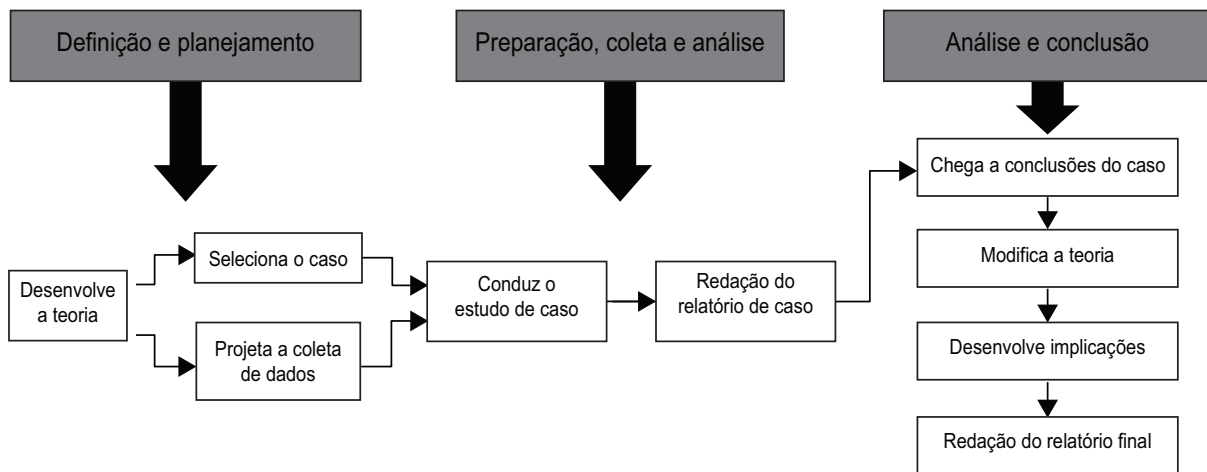
O método prevê a organização de ambientes voltados para a realização de trabalhos cooperativos. Essa organização refere-se à estrutura das reuniões de trabalho, aqui denominadas sessões. Cada sessão possui um objetivo declarado e, com base nesse objetivo, são produzidos os resultados.

As sessões são conduzidas por um facilitador, que domina o método do Grupo Cooperativo de Trabalho. O facilitador é apoiado por um ou mais monitores e são observadas três regras básicas: a quantidade de participantes; o tempo de duração e os intervalos.

Etapa 3: compreendeu a criação do processo de inserção de dados, extração, transformação e carga dos dados (ETL). Na sequência foi iniciada a construção do processo de orquestração, a atualização das bases de dados e a transferência do conhecimento. O método usado foi um dos resultados do desenvolvimento do estudo, e a consequência direta dessa etapa foi a construção de um protótipo que congregou o estudo do ecossistema do *Big Data* do Ministério do Trabalho.

Etapa 4: compreendeu os processos de representação e organização da informação multimodal, que definiu funcionalidades generalizadas, interface de comunicação com sistemas externos, definição das interfaces generalizadas de BI e definição de padrões generalizados de relatórios. O método utilizado foi a técnica da questão gerencial, componente do processo de organização da informação para avaliação e gestão, envolvendo a tomada de decisão (informação gerencial). A figura 2 resume as etapas de investigação.

Figura 2 – Método de estudo de caso



Fonte: adaptado de Yin, 2005

O objetivo do emprego desse método foi o de identificar as necessidades de informação gerencial das comunidades de gestores interessadas no acompanhamento e na avaliação de desempenho dos sistemas que compõem o ecossistema de *Big Data*. As principais vantagens auferidas do seu uso foram garantia do registro preciso das necessidades de informação gerencial dos gestores; obtenção de subsídios para a seleção dos indicadores de desempenho; facilitação da modelagem multidimensional para a estruturação dos mecanismos adequados de armazenamento da informação em bancos de dados; e auxílio na criação e na atualização de indicadores disponíveis.

CONSIDERAÇÕES FINAIS

O tratamento de grandes quantidades de dados que deveriam ser utilizados para gerar informações úteis em processos decisórios constitui desafio para a gestão da informação nas organizações. A complexidade cada vez maior dos mercados e as exigências da regulação governamental em setores estratégicos da economia levam as organizações públicas e privadas a acumular grandes massas documentais em formatos digitais e convencionais. Essa realidade impõe uma busca sistemática por soluções que considerem a transformação de dados em informações.

Nesse contexto, os processos tradicionais de coleta, processamento e disseminação passam a ser considerados além das fronteiras de estudo da ciência da informação e se tornam cada vez mais corriqueiros entre gerentes e mesmo entre os especialistas em gestão estratégica. Antes, o interesse por estudos relativos aos processos de busca e recuperação de informação estava circunscrito a alguns grupos de profissionais, notadamente arquivistas e bibliotecários, mas hoje permeia as preocupações dos tomadores de decisão e os formuladores de políticas públicas e empresariais.

Um bom exemplo para essa nova abrangência na organização da informação está na crescente demanda pelos sistemas de inteligência competitiva, que objetivam a promoção do uso racional de

informações extraídas, por meio de *Big Data* do contexto de atuação das organizações, a fim de melhor posicionar suas ações estratégicas. Um dos elementos centrais desses sistemas é a informação gerada no âmbito organizacional, sendo a coleta, o tratamento, a análise e a disseminação tarefas básicas para o seu desenvolvimento, coincidentes com as etapas do ciclo documentário.

Assim, identificamos três tipos de contribuição da investigação, ora apresentada, para a melhoria da gestão dos processos no âmbito da administração pública:

- a determinação das competências essenciais necessárias à concepção de um modelo de gestão estratégica de dados e de informação passa por estudos sistemáticos dos modelos de gestão dos ecossistemas de *Big Data*, além de conhecimento das necessidades de informação dos usuários que farão uso desses modelos;
- as etapas do ciclo documentário devem integrar a concepção do *Big Data*, sobretudo no que se refere à organização e representação da informação, permitindo estabelecer um processo de busca e recuperação apoiado nas necessidades informacionais dos usuários do ecossistema; e
- o uso de ferramentas de BI propicia o aprimoramento de processos tradicionais da ciência da informação, além de constituir um campo profícuo de pesquisa, de modernização e de adaptação de novas soluções para melhoria do desempenho dos sistemas de informação e dos ecossistemas de *Big Data*.

A proposta de um modelo de ecossistema de *Big Data* para conciliação das demandas de acesso, por meio da representação e organização da informação contribui para a linha de pesquisa de organização e gestão da informação e do conhecimento, pois possibilita o desenvolvimento de soluções de tecnologia da informação aplicada à elaboração de instrumentos de gestão de documentos.

A coleta, o tratamento e a análise de grandes quantidades de dados para uso do governo federal é um desafio claro à efetividade da

decisão dos gestores públicos, que, além disso, têm que lidar com a pouca familiaridade com requisitos tecnológicos e com a cultura limitada da organização da informação, tendo como consequência um processo de análise superficial para subsidiar o processo decisório.

Nesse contexto, a integração dos aspectos tecnológicos e dos insumos humanos qualificados enseja a necessidade de estudo das competências essenciais que devem congregiar ações de gestão dos fluxos informacionais, a fim de fornecer requisitos básicos para a construção de um modelo de ecossistema de *Big Data* que apoie efetivamente a decisão estratégica.

REFERÊNCIAS

AGUILAR, A. G. *Visualização de dados, informação e conhecimento*. Florianópolis: Editora UFSC, 2017.

ARAÚJO JÚNIOR, R. H. de. *Estudo de necessidades de informação dos gerentes do setor editorial e gráfico do Distrito Federal*. 1998.178 p. Dissertação (Mestrado em Ciência da Informação) - Universidade de Brasília.

_____. *Precisão no processo de busca e recuperação da informação*. Brasília: Thesaurus, 2007.

_____.; SOUSA, R. T. B. de; ALBUQUERQUE, S. F. de. Métodos, técnicas e instrumentos de organização e gestão da informação nas organizações. In: BAPTISTA, D. M.; ARAÚJO JÚNIOR, R. H. de. (Org.). *Organização da informação: abordagens e práticas*. Brasília: Thesaurus, 2015. p. 44-68.

BARBIERI, C. *Business intelligence: modelagem e tecnologia*. Rio de Janeiro: Axcel Books, 2001.

BRASIL. Ministério do Trabalho. *E Social padroniza portal para melhorar atendimento à população*. Brasília: Ministério do Trabalho, 2017. Disponível em: <<http://trabalho.gov.br/noticias/4567-esocial-padroniza-portal-para-melhorar-atendimento>>. Acesso em: 10 Ago. 2017.

CAFÉ, L.; SALES, R. Organização da informação: Conceitos básicos e breve fundamentação teórica. In: ROBREDO, J.; BRÁSCHER, M. (Orgs.). *Passeios no Bosque da Informação: Estudos sobre Representação e Organização da Informação e do Conhecimento* - EROIC. Brasília: IBICT, 2010. p.115-129. Disponível em: <<http://www.ibict.br/publicacoes/eroic.pdf>>. Acesso em: 10 Ago. 2017.

DAVENPORT, T. H. *Big Data no trabalho: derrubando mitos e descobrindo oportunidades*. Rio de Janeiro: Campus, 2014.

GARCIA MARQUEZ, F. P.; LEV, B. *Big Data management*. New York: Springer, 2016.

GLASS, R.; CALLAHAN, S. *The Big Data-driven business: how to use Big Data to win customers, beat competitors, and boost profits*. New Jersey: John Wiley & Sons, 2015.

GOMES, L. F. A. M.; GOMES, C. F. S.; ALMEIDA, A. T. de. *Tomada de decisão gerencial: enfoque multicritério*. São Paulo: Atlas, 2009.

GUEDES, W.; ARAÚJO JÚNIOR, R. H. de. Estudo das similaridades entre a teoria matemática da comunicação e o ciclo documentário. *Inf. & Soc.: Est., João Pessoa*, v.24, n.2, p. 71-81, maio/ago. 2014. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/view/16498/11492>>. Acesso em: 08 Ago. 2017.

INMON, W. H.; LINSTEDT, D. *Data architecture: a primer for the data scientist: Big Data, data warehouse and data vault*. Burlington: Elsevier, 2014.

INMON, W. H.; TERDEMAN, R. H.; IMHOFF, C. *Data Warehousing: como transformar informações em oportunidades de negócio*. São Paulo: Berkeley, 2001.

McGEE, J. PRUSAK, L. *Gerenciamento estratégico da informação*. Rio de Janeiro: Campus, 1994.

RAMOS, A. *Infraestrutura Big Data com opensource*. Rio de Janeiro: Ciência Moderna, 2015.

ROBREDO, J. *Documentação de hoje e de amanhã*. Brasília: Edição de autor, 2005.

WALKER, R. *From Big Data to big profits: success with data and analytics*. New York: Oxford University Press, 2015.

YIN, R. K. *Estudo de caso: planejamento e métodos*. Porto Alegre: Bookman, 2005.