

Tecnologia big data em bibliotecas: estudo contributivo a partir de pontuações literárias

Big data technology in libraries: a contributory study based on literary scores

Larissa Bacelar Pontes Lopes

Graduada em Biblioteconomia pela Universidade de Brasília (UnB).

larissabponteslopes@gmail.com

Márcio Bezerra da Silva

Doutor em Ciência da Informação pela Universidade Federal da Bahia (UFBA). Docente na Faculdade de Ciência da Informação da Universidade de Brasília (UnB).

marciobdsilva@unb.br

RESUMO

Estudo que apresenta um cenário literário de uso da tecnologia *big data* com destaque às bibliotecas. Aborda no referencial teórico: pirâmide informacional, baseada nos elementos dado, informação, conhecimento e sabedoria; Internet, abordando as *webs* 1.0, 2.0 e 3.0; e *big data*, ponderando suas características básicas, representadas por volume, velocidade e variedade (3V), e exemplos de aplicação da tecnologia nos segmentos corporativo, governamental, científico e cotidiano social. Caracteriza-se como um estudo exploratório, bibliográfico e qualitativo pela coleta de dados. Apresenta, como resultados, o elenco de exemplos de uso da *big data* nas bibliotecas da *Washington University*, *Harvard University* e *Seattle Public*; e a elaboração de um quadro sugestivo de aplicações da tecnologia em setores de biblioteca segundo as características “3V”, como volume-velocidade-variedade na administração e na circulação, volume-variedade no setor de desenvolvimento de coleções e valor-volume no setor de referências. Conclui-se que as características “3V”, encaradas como básicas da tecnologia *big data* e complementadas pelas propriedades complexidade, veracidade e valor, podem auxiliar na tomada de decisão em setores das bibliotecas, como na administração, na circulação, no desenvolvimento de coleções e na referência, em um cenário de produção e compartilhamento de informações promovidos pelas tecnologias de informação e comunicação, como a Internet e em sentido exponencial.

Palavras-chave: Pirâmide informacional. Internet. *Web*. *Big data*. *Big data* em bibliotecas.

ABSTRACT

Study that presents a literary scenario of use of the big data technology, with emphasis on libraries. Discusses in the theoretical referential: informational pyramid, based on the given elements, information, knowledge and wisdom; Internet, addressing the 1.0, 2.0 and 3.0 webs; and big data, considering its basic characteristics, represented by volume, speed and variety (3V), and examples of application of technology in the corporate, governmental, scientific and daily social segments. Characterizes as an exploratory, bibliographic and qualitative study by data collection. Presents, as results, the list of examples of the use of big data in the libraries of Washington University, Harvard University and Seattle Public; and the elaboration of a table suggestive of technology applications in library sectors according to the “3V” characteristics, such as volume-speed-variety in administration and circulation, volume-variety in the collection development sector and value-volume in the sector of references. Concludes that that the “3V” characteristics, seen as basic to big data technology and complemented by the complex, veracity and value properties, can assist in decision making in library sectors, such as administration, circulation, development of co-lectures and reference, in a scenario of information production and sharing promoted by information and communication technologies, such as the Internet and in an exponential sense.

Keywords: Informational pyramid. Internet. *Web*. Big Data. Big Data in libraries.

1 INTRODUÇÃO

A dinâmica social é um processo que se apresenta em constante transformação, influenciada, principalmente em tempos contemporâneos, por tecnologias digitais, alterando nas formas como as pessoas se comunicam entre si. Nesta contextualização espaço-temporal citam-se marcos responsáveis por mudanças sociais, como a escrita, a qual possibilita a preservação da memória humana. Em seguida, elenca-se a máquina (de imprensa) de Gutemberg, a qual impulsionou a produção de livros e, conseqüentemente, acarretou na circulação de informações de maneira mais veloz, via suporte físico. Também se destaca o telefone de Graham Bell que, ao romper as fronteiras territoriais, impulsionou a comunicação entre países. Porém, conforme Briggs e Burke (2004), foram as tecnologias de comunicação e informação (TIC), representadas pelo computador e pela Internet, as responsáveis por transformarem, significativamente, a dinâmica social.

A partir da Internet, enraizada na Segunda Guerra Mundial, e evoluindo conforme as gerações da camada *World Wide Web* (WWW), enquanto “[...] um sistema de documentos de hipermídia que são interligados e executados na Internet” (NONATO, 2015, p. 4), as dinâmicas sociais foram se alterando, em uma constante metamorfose relacional entre pessoas e aplicações digitais, assim como entre máquinas, com parcial e/ou nenhuma participação humana.

Compreender a evolução da WWW, iniciando na 1.0, focada na publicização de pessoas, produtos e serviços, de forma estática, permitindo apenas pesquisa e leitura, passando pela 2.0, fomentando a integração, participação e cooperação entre usuários, até chegar na 3.0 (*web* semântica), desejando a estruturação de grandes quantidades de dados, os quais são relacionados semanticamente, de forma a subsidiar a tomada de decisões sob a perspectiva informacional dos indivíduos, pois, se por um lado a Internet disponibilizou/disponibiliza uma massiva quantidade de dados aos seus usuários, por outro, a qualidade desses dados tornou-se uma problemática.

Diante da supracitada perspectiva, evoca-se a pirâmide informacional (figura 1), tendo o dado como um elemento de caráter quantitativo, localizado na base, e a inteligência, também nomeada como sabedoria, fixada, opostamente, no pico da representação, de caráter qualitativo e de maior abstração. Neste sentido, a evolução das

gerações da WWW caminha em paralelo à “[...] problemática da geração, da organização, da transferência e do aproveitamento da informação [...]” (URDANETA, 1992, p. 99, tradução nossa).

Figura 1 - Pirâmide informacional.



Fonte: Urdaneta (1992, p. 100).

De acordo com a pirâmide informacional (figura 1), quanto mais simples ou nulo for o tratamento realizado em certo conteúdo, mais próximo esse conteúdo estará da base piramidal. Segundo Urdaneta (1992), a partir da hierarquia por níveis, como uma relação entre qualidade vs. quantidade, desenha-se um sentido iniciado pelo dado, que se transforma em informação, que por sua vez se torna conhecimento e, finalmente, consolida-se como sabedoria, nível em que ocorre o alcance máximo da cognição humana, gerando novas descobertas e fomentando a tomada de decisões.

Como um meio de subsidiar a tomada de decisões no tempo contemporâneo da WWW, de forma veloz e, por vezes, de maneira indicativa (recomendada), menciona-se a *big data*, caracterizada por Taurion (2015) como um conjunto de tecnologias, processos e práticas que permitem a análise de um grande volume e variedade de dados que, em sua maioria, não eram acessíveis, em um espaço de tempo curto e passível de gerar valor de decisão.

Produzir, tratar e disseminar quantidades cada vez maiores de dados ocorrem em diversos segmentos da sociedade, vislumbrando cenários de aplicação da *big data* como um caminho para a tomada de decisões. Entre esses espaços estão as bibliotecas, especialmente quando rompem o paradigma de ser um ambiente exclusivo de

organização da informação, como um local de armazenamento, para um local que oferta serviços de recuperação dessa informação organizada, pensando nas necessidades de seus usuários.

Ao focar nas intenções dos usuários, profissionais da informação, como os bibliotecários, podem coletar, tratar e analisar dados com fins de tomada de decisão, pressuposto que fomentou a investigação de como se daria a adoção da *big data* nos serviços realizados e ofertados por bibliotecas.

Na eminência da necessidade de atualização constante dos profissionais bibliotecários frente às novas tecnologias, da multiplicidade de setores que compõem uma biblioteca, dentre a oferta de produtos e serviços na revolução digital, surge a indispensabilidade de análise e acompanhamento dos novos recursos e dos já existentes, com fins de futuras decisões quanto à continuidade, atualização ou encerramento dos produtos e serviços, bem como pela satisfação dos usuários. A necessidade de avaliação da qualidade, de acordo com Rozados (2004), vai ao encontro de indicadores enquanto insumos de avaliações e que subsidiam a adequação, o rendimento e as falhas dos serviços de informação. Diante da possibilidade de coletar dados quantitativos e qualitativos dos setores da biblioteca e reuni-los periodicamente, de forma a resultar em volume informacional que pode ser analisado e conseqüentemente, gerar indicadores por meio de recursos tecnológicos como a mineração de dados e *big data* - possibilita análise de dados em grandes quantidades e em alta velocidade - nasceu o interesse em verificar a realidade de produção e análise no contexto da biblioteca, neste caso, a partir de uma investigação literária sobre características e práticas da *big data*.

De forma ampla, a pesquisa objetivou apresentar um cenário literário do paradigma contemporâneo da *big data* em bibliotecas. Especificamente, objetivou-se: elencar exemplos de uso de *big data* em bibliotecas; e, por fim, exibir e descrever um quadro sugestivo de aplicações da *big data* em bibliotecas. Destaca-se que se trata de um estudo propositivo, a partir do resgate de conceitos e exemplos já consolidados, de iniciativas para implementação em bibliotecas com enfoque na *big data* e sem a intenção de esgotar o tema em questão.

O percurso metodológico caracterizou-se como um estudo descritivo, de natureza básica, de método indutivo, e que utiliza a técnica de pesquisa bibliográfica, com abordagem de coleta de dados qualitativa. A partir dos ambientes *Google* buscador,

Biblioteca Digital de Dissertações e Teses (BDTD) e Base de Dados Referencial de Artigos e Periódicos em Ciência da Informação (BRAPCI), para consultas e coleta de dados, somados a livros, anais de eventos e *blogs* especializados em Tecnologia e Informação (TI), utilizou-se a combinação das palavras-chave “*big data*” e/ou “biblioteca” e “*library*”, como forma de filtrar os resultados obtidos durante o período de agosto de 2018 a março de 2019 e, assim, embasar a descrição das sugestões a partir da confecção de um quadro propositivo de uso da *big data* em bibliotecas.

O processo de busca e seleção das fontes de informações, em plataformas nacionais e internacionais, culminou no fato metodológico de, a partir de análise sistemática dos materiais selecionados, eger exemplos, coincidentemente, pertencentes à América do Norte, mais especificamente, provenientes dos Estados Unidos da América.

2 DA INTERNET ESTÁTICA À WEB SEMÂNTICA

Dentre as inovações tecnológicas que influenciam na evolução da forma de comunicação das sociedades destaca-se a Internet, cunhada em tempos de guerra, tendo as suas primeiras versões criadas durante a Segunda Guerra Mundial. A partir da combinação de esforços de centros governamentais e centros universitários a Internet “[...] se desenvolveu em um ambiente seguro, propiciado por recursos públicos e pesquisa orientada para a missão [...]” (CASTELLS, 2003, p. 26). Se por um lado, o governo propiciou capital e liberdade, por outro, os centros universitários serviram de difusores da iniciativa (CASTELLS, 2003).

O que permitiu a Internet alcançar um nível global foi o desenvolvimento da WWW, em que Berners-Lee, em colaboração com Robert Cailliau: “[...] construiu um programa navegador/editor em dezembro de 1990 [...]” (CASTELLS, 2003, p. 19), que possibilitava máquinas a “[...] obter e acrescentar informação de e para qualquer computador conectado através da Internet: HTTP, HTML e URI (mais tarde chamado URL).” (CASTELLS, 2003, p. 19).

A evolução da Internet resultou nas diversas fases da *web*, a iniciar pela *web* 1.0, a qual permite, de forma restrita, a recuperação e a leitura de informações pelos usuários (AGHAEI; NEMATBAKHS; FARSANI, 2012). Chamada de “*read only web*”, essa primeira

versão começou como “um lugar de informação para que empresas difundissem as informações delas para a população” (AGHAEI; NEMATBAKSH; FARSANI, 2012, p. 1, tradução nossa), similar a um catálogo, em que apresentavam seus serviços e produtos. O objetivo dessa *web* era estabelecer presença/publicidade *online* apresentando “[...] páginas que só podem ser entendidas por humanos [...]” (CHOUDHURY, 2014, p. 8096, tradução nossa) e uma “ausência de representação dinâmica [...]” (CHOUDHURY, 2014, p. 8097, tradução nossa), caracterizada pela inexistência de interações com o usuário e elementos prioritariamente textual. Contudo, salienta-se que, ainda que caracterizada como a fase inicial da WWW, a *web* 1.0 se faz presente em ambientes atuais, como *websites* de órgãos públicos, empresas, instituições de ensino etc., entretanto, complementadas por recursos das fases posteriores, como o compartilhamento de informações em formatos diversos e a participação ativa dos usuários, marcos predominantes da segunda fase da *web*.

O acréscimo de funcionalidades na *web* 2.0 em relação à primeira fase resultou na ampliação e rapidez na forma de comunicação e compartilhamento de informações entre os seres humanos. Conceituada por Dale Dougherty, em 2004, como “*read-write-web*”, a *web* 2.0 apresentou como atrativo a possibilidade de interação, ou seja, os usuários tornam-se colaboradores, pois são capazes de relacionarem-se com o ambiente digital e por meio dele (AGHAENI; NEMATBAKSH; FARSANI, 2012). Se por um lado, o aumento da colaboração influenciou o surgimento de novas aplicações *web*, como as redes sociais; por outro, significa que há “[...] mais interação com menos controle [...]” (AGHAENI; NEMATBAKSH; FARSANI, 2012, p. 3, tradução nossa), uma vez que a possibilidade de criação e atualização de conteúdo na *web* social, aumenta os casos de “[...] vandalismo, confusões e erros de informação ou de uso de ferramentas (como apagamento incidental [*sic*] de dados) [...]” (PRIMO, 2007, p. 6) por se tratar de um ambiente flexível e dinâmico (ALVIM, 2007).

Por fim, o grande fluxo de informação propiciado pela *web* 2.0, causado pelo alto volume de criação e transformação de informação, coloca em uso o discernimento humano, que nem sempre obtém êxito para filtrar o conteúdo exposto no ambiente digital para seu consumo, bem como realizar as escolhas mais adequadas diante da variedade de informações/opções. Neste sentido, em um cenário com múltiplas possibilidade e opções de escolha, a *web* 3.0 mostra-se com uma tentativa, de por meio da tecnologia semântica,

direcionar o usuário dentre a ampla gama de resultados, considerando os perfis dos usuários, especificamente sobre seus interesses, gostos e preferências.

A “*read-write-execute*” (*web* para leitura, escrita e execuções), ou *web* 3.0, intenciona “[...] definir dados estruturados e conectá-los de forma a possibilitar descobertas mais efetivas, automação, integração, e reuso por meio de aplicações diversas” (CHOUDHURY, 2014, p. 8097, tradução nossa). Trata-se de uma *web* legível por máquinas, através da adoção de tecnologias disponíveis e complementares, que obtém informações por meio da análise e de “*links*” de conjuntos de dados diversos (AGHAEI; NEMATBAKHS; FARSANI, 2012).

A terceira fase oferece maior autonomia ao usuário por meio da interconexão entre documentos e dados relacionados, da ampliação do alcance das informações publicadas *online*, do processamento automático das informações e da maior integração das informações (GARCÍA-MARCO, 2010).

A *web* 3.0 suprime, tanto os tradicionais conceitos de *website*, como a dinâmica de posse de dados para compartilhamento, por meio de serviços especializados (CHOUDHURY, 2014). Os dados, nessa fase da *web*, são compartilhados e os serviços possuem perspectivas variadas, ou seja, diversas aplicações em dispositivos e múltiplos formatos, com “[...] foco no contexto e na personalização [...]” (TORRES, 2007, p. 1, tradução nossa). Dessa forma, a *web* de dados, ainda que construída por documentos assim como os *websites* tradicionais de hipertexto, os *links* são feitos “[...] entre objetos arbitrários descritos por RDF” (BERNERS-LEE, 2009, p. 1, tradução nossa), utilizando *Uniform Resource Identifier* (URI) que são capazes de identificar “[...] qualquer tipo de objeto ou conceito” (BERNERS-LEE, 2009, p. 1, tradução nossa), ou seja, tornar os documentos recuperáveis com maior facilidade.

Em suma, a partir do *Resource Description Framework* (RDF) é possível organizar a *web* por *links* entre elementos digitais, possibilitando que a intenção da *web* 3.0, de integrar e conectar formatos diversos para permitir os “[...] usuários de achar, compartilhar e combinar informações mais facilmente” (CHOUDHURY, 2014, p. 8098, tradução nossa) seja viabilizada.

3 O PARADIGMA DA BIG DATA: PONTUAÇÕES

Ao passo que a *web* 3.0 foi se desenvolvendo, áreas do conhecimento foram se apresentando úteis ao contexto, como é o caso da Inteligência Artificial (AI) e da Ciência de Dados (DS), com ela, novas aplicações, como o *machine learning* e a *big data*. Neste sentido, serviços como *WolframAlpha*, que responde a perguntas dos usuários para recuperar informações acerca de termos buscados; e a *Siri*, fornecida pela *Apple*, que, ao utilizar a voz dos usuários em seus dispositivos móveis, responde a requisições feitas, como uma pergunta sobre clima de certa região.

Enquanto a *web* semântica demonstra o potencial de recuperar e reunir dados úteis em um vasto conjunto de informações disponíveis, ferramentas como a *big data* possuem um objetivo mais ambicioso, ou seja, gerar conhecimento a partir de uma extensa e variada gama de informações, disponibilizadas em alta velocidade e, assim, contribuindo diretamente na geração de sabedorias pelo usuário, o que lhe embasará na tomada de decisões.

A contemporaneidade, permeada por tecnologias aplicadas em processos de diferentes complexidades, da comunicação via aparelhos móveis até o uso de sensores em estradas, presencia uma demasiada geração de dados, de diferentes formatos, em grandes quantidades e em curto espaço de tempo. Resulta-se em um cenário fomentador à aplicação da *big data*, a qual configura-se como um recurso emergido no prisma da *web* 3.0, “[...] aplicado à informação que não pode ser processada ou analisada por métodos usando processos ou ferramentas tradicionais” (ZIKOPOULOS *et al.*, 2012, p. 3, tradução nossa).

Considerando as possibilidades de adoção, a *big data* caracteriza-se por três elementos básicos “[...] volume, variedade e velocidade” (ZIKOPOULOS *et al.*, 2012, p. 5, tradução nossa), também conhecidos como “3V”. O **volume** refere-se à quantidade de dados gerada diariamente, de forma exorbitante, construindo um cenário de produção de dados exponencialmente quantitativo desde os anos 2000, assim como defendem os autores ao afirmarem que:

No ano 2000, 800,000 petabytes (PB) de dados estavam armazenados no mundo. [...] Nós esperamos que esse número alcance 35 zettabytes (ZB)

até 2020. Sozinho, o Twitter gera mais de 7 terabytes (TB) de dados todos os dias, Facebook 10 TB e algumas empresas geram terabytes em dados a cada hora, todos os dias do ano. (ZIKOPOULOS *et al.*, 2012, p. 5-6, tradução nossa)

Além do quantitativo, a *big data* apresenta-se como uma solução de análise de grandes volumes de dados a partir de sua **variedade**, a qual alude a todos os tipos de formatos, uma vez que, nos dias atuais, a explosão das tecnologias gera dados complexos e em multiplicidade. Desta forma, a solução *big data* atinge a seguinte cobertura de dados:

[...] não inclui apenas dados relacionais tradicionais, mas também dados brutos, semiestruturados, e não estruturados provindos das páginas web, arquivos de login da web [...], buscas indexadas, fóruns de mídia social, e-mail, documentos, [...] e assim por diante. (ZIKOPOULOS *et al.*, 2012, p. 7, tradução nossa).

A existência de sistemas que analisam dados de formatos variados é aparentemente uma realidade impraticável, de difícil exercício. Entretanto, trabalhar com dados estruturados indica a possibilidade de recuperação devido a organização que possuem, neste caso, “[...] grupos possuindo as mesmas descrições” (NASCIMENTO *et al.*, 2015, p. 5). Por outro lado, ao tratar de dados desestruturados, as questões de guarda e recuperação tornam-se complexas, uma vez que envolvem informações aleatórias e estruturas não definidas. É nesse contexto que a *big data* se torna uma solução viável, a partir de potenciais análises baseadas em dados estruturados e não estruturados para apoiar a tomada de decisão (ZIKOPOULOS *et al.*, 2012).

Por último, o predicado **velocidade** considera a rapidez com que são gerados os dados, segundo certo volume e dimensão de variedade. A característica em questão “[...] refere-se a aspectos da dinâmica de crescimento e processamento dos dados, em que o processo de tomada de decisão está muito interligado ao momento e a rapidez que tal decisão é tomada” (CONEGLIAN *et al.*, 2018, p. 139). Sendo assim, a rapidez nas ações de gerenciamento e análise de dados é posicionada como um componente de vantagem competitiva, como a extração de uma informação relevante para a tomada de decisão em tempo real. Trata-se da chamada *streams computing*, enquanto um processamento de “[...] dados que são gerados continuamente ou ao vivo à medida que estes dados se tornem disponíveis [...]” (CHEDE, 2009, p. 1).

Além das características supracitadas, elencam-se três propriedades: complexidade, veracidade e valor. A **complexidade** destina-se ao estabelecimento de relações e *links* entre dados que provém de fontes variadas, uma vez que é de grande facilidade perder o controle sobre esses dados. A propriedade **veracidade** refere-se à validade e inconsistência dos conjuntos de dados. O **valor** da informação está relacionado com o conhecimento que ela pode gerar, principalmente em organizações, em que a geração de conhecimento reflete diretamente nas estratégias adotadas (KATAL; WAZID; GOUDAR, 2013). Há ainda, autores que atribuam características adicionais: validade, variabilidade, visualização, viralidade e viscosidade, alcançando os 10Vs (CHEN; MAO; LIU, 2014).

A partir das características em “V” (volume, variedade e velocidade) infere-se que a solução *big data* torna-se pertinente “[...] quando todos, ou a maioria, dos dados precisam ser analisados versus uma amostra de dados [...]” (ZIKOPOULOS *et al*, 2012, p. 16, tradução nossa). Assim, a sua adoção será motivada por “[...] análises interativas e exploratórias quando as medidas da análise dos dados não foram pré-determinadas pelas organizações” (ZIKOPOULOS *et al*, 2012, p. 16, tradução nossa). A partir dos resultados obtidos por essas análises é possível tomar decisões, tanto para traçar metas, objetivos, ações, como para atingi-los, conseqüentemente (ARAÚJO JÚNIOR; SOUZA, 2018).

Diante de seus predicados, a *big data* tem a capacidade de atuar em segmentos variados, incentivando o aumento da produtividade e competitividade das organizações. Por exemplo, o uso de aplicações *web* viabilizam predições de mercado (corporativo) com base em informações provindas de fontes variadas, como a análise de redes sociais e o histórico de buscas dos usuários (CHEN; ZANG, 2014). Neste contexto é possível citar a empresa *WalMart*¹, ao implementar a *big data* como fins de “[...] rastrear toda venda gravada em seus terminais de venda” (CHEN; ZANG, 2014, p. 3, tradução nossa) e, a partir da análise do (grande) volume de dados produzidos, cerca de 267 milhões de transações, torna-se capaz de “[...] aumentar a eficiência deles nas estratégias dos preços e nas campanhas de publicidade” (CHEN; ZANG, 2014, p. 3, tradução nossa).

Exemplos também são identificados não apenas quanto ao volume (em ascensão) de informação produzida, pois, por vezes, exige-se que as organizações também lidem

¹ Site: <https://www.walmart.com.br/>.

com a velocidade na qual essa produção é gerada. Para tanto, cita-se a plataforma *FICO Falcon*², a qual trata-se de um sistema que detecta fraudes em cartões de crédito, gerenciando “[...] mais de 2.1 bilhões de contas válidas ao redor do mundo” (CHEN; ZANG, 2014, p. 3, tradução nossa). A velocidade torna-se parâmetro essencial para a *FICO Falcon*, uma vez que necessita receber informações em tempo real para identificar eventuais anomalias.

Não somente empresas do setor privado utilizam dos benefícios oferecidos pelo recurso *big data*. No contexto dos órgãos da administração pública, ao lidar com centenas de pessoas, de idades diferentes e com necessidades variadas, a população anseia que a “máquina” governamental “[...] seja mais efetiva na administração pública” (CHEN; ZANG, 2014, p. 3, tradução nossa). Considerando o valor da informação, a *Gendarmerie Nationale*³, grupo policial francês comprometido com a segurança pública do Estado, utiliza da velocidade em tempo real, veracidade e volume de informações dos sistemas aos quais possuem acesso, o que, conseqüentemente, aumentou a economia de tempo e de recursos nas tomadas de decisões (SAP SE, 2017).

O emprego da *big data* mostra potencial de uso também no segmento de pesquisas científicas, como “[...] o grande telescópio de levantamento sinóptico (*the Large Synoptic Survey Telescope – LSST*⁴) [que] irá gravar 30 trilhões de *bytes* de dados em imagens em um único dia” (CHEN; ZANG, 2014, p. 4, tradução nossa). A análise desses dados resultará em informações acerca da origem do universo, ao validar o “amontoador” de informações capturadas pelo telescópio.

Presente em variados segmentos, entre sociais, corporativos e/ou governamentais, a *big data* encontra-se também em aplicações de uso cotidiano. A partir dos dados coletados de usuários, aplicações *web*, como o *Google Now*, o *Netflix* e a *Amazon* são capazes de realizar recomendações personalizadas e aprimorá-las a partir do aprendizado do comportamento deles.

Em suma, a solução *big data* constitui a combinação de características, entre os “3V” e complexidade, veracidade e valor, as quais subsidiam o estabelecimento de diretrizes acerca do uso, aplicação e resultados (possíveis) de dados coletados e

² Site: <https://www.fico.com>.

³ Site: <https://www.gendarmerie.interieur.gouv.fr/>.

⁴ Site: <https://www.lsst.org/lsst/>.

informações geradas. A ferramenta agrega valor aos dados, oferta informações tratadas e compreensíveis, contribui na possibilidade de constituir conhecimento e potencializa a tomada de decisão a partir da sabedoria construída, para, assim, estipular e avaliar estratégias e metas nas organizações.

4 RESULTADOS DA PESQUISA

Considerando os exemplos elencados em setores diversos, entre pessoal, corporativo, governamental e científico, ingere-se a *big data* como um recurso de ampla implementação, em variados segmentos. Neste sentido, por meio do percurso metodológico adotado, identificou-se a aplicação da *big data* em bibliotecas, gerando a detecção de três ambientes, especificamente a *Harvard University Library*⁵, a *Washington University Library*⁶ e a *Seattle Public Library*⁷.

Somadas às características básicas e exemplos de aplicação da *big data* no contexto geral, a sua possibilidade de adoção em nichos diversos torna o campo da Educação um segmento passível de aplicação, especificamente nas bibliotecas, enquanto unidades onde o fluxo de informação é intenso, tanto na produção, quanto na disseminação, fazendo com que o tradicional conceito de biblioteca seja repensado. Em outras palavras, com o progresso tecnológico, o conceito de biblioteca foi ampliado, fomentando o surgimento de espaços extensivos ao ambiente físico, como as bibliotecas digitais, por exemplo. A biblioteca comumente relacionada a espaços físicos, e que objetiva guardar e organizar informações gravadas em suportes físicos para disponibilizar a sociedade (SANTOS; ASSUNÇÃO, 2012), passa a romper paradigmas a partir da sua versão digital.

Por mais que rupturas abruptas tenham obrigado as bibliotecas a adaptarem-se, (re)surgindo em novos formatos, a sua missão não foi alterada, ou seja, de “[...] oferecer um conjunto organizado de informações afins, armazenadas com o objetivo de consulta ou utilização posterior” (LUCAS, 2004, p. 16). Para tanto, as bibliotecas necessitam de gerenciamento e, por conseguinte, de tomadas de decisão, especialmente pela

⁵ Site: <https://library.harvard.edu/>.

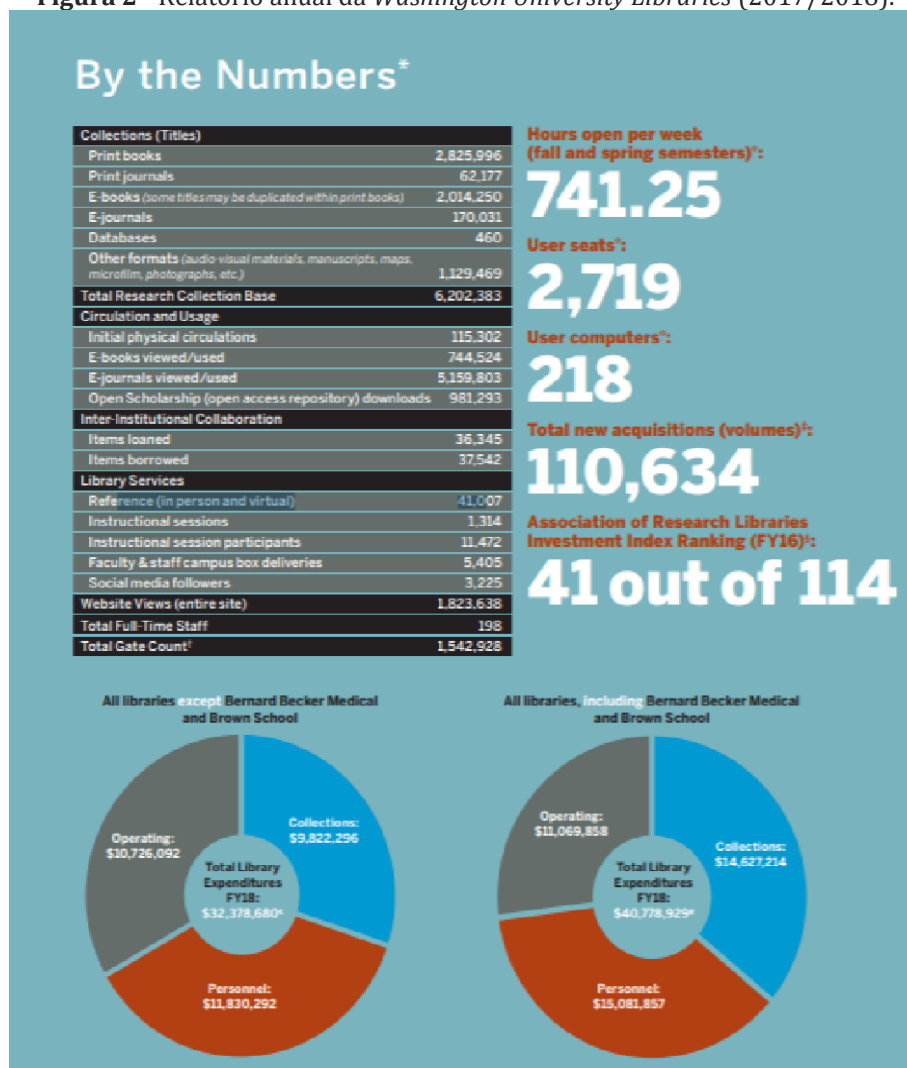
⁶ Site: <https://library.wustl.edu/>.

⁷ Site: <https://www.spl.org/>.

contemporânea explosão de informação, tornando-se, assim, apta para a implantação da *big data* em processos específicos.

O paradigma contemporâneo da *big data* contribui para a implementação de soluções em bibliotecas voltadas, principalmente, para a visualização de dados e, a partir disso, gerar conhecimentos para embasar as tomadas de decisão. Na *Washington University*, por exemplo, a partir do conjunto de dados coletados e reunidos pelas bibliotecas setoriais da Universidade, nasceu o projeto *Making the numbers speak*, que consiste em representar os dados graficamente e interpretá-los (CHEN *et al.*, 2015), conforme ilustra a figura 2. Os resultados guiaram a direção estratégica das bibliotecas, preocupadas em “[...] tomar melhores decisões em relação às bibliotecas” (CHEN *et al.*, 2015, p. 2, tradução nossa).

Figura 2 - Relatório anual da *Washington University Libraries* (2017/2018).



Fonte: *Washington University Libraries* (2018, p. 13).

A partir da reunião de relatórios semestrais e anuais (figura 2) geram-se informações, e combinadas com o método de *Balanced Scorecard*, as bibliotecas da *Washington University* embasam os seus planejamentos estratégicos. A partir do alinhamento com a característica volume presente na *big data*, os resultados, de visualizar graficamente os volumes de dados coletados, refletem nos comportamentos organizacionais. A partir dos dados dos projetos desenvolvidos em 2015-2016, por exemplo, foi lançado o projeto *Geospatial Information Systems* em 2017 (CHEN *et al.*, 2016), ilustrado na figura 3.

Figura 3 - Relatório anual da *Washington University Libraries* (2017/2018).



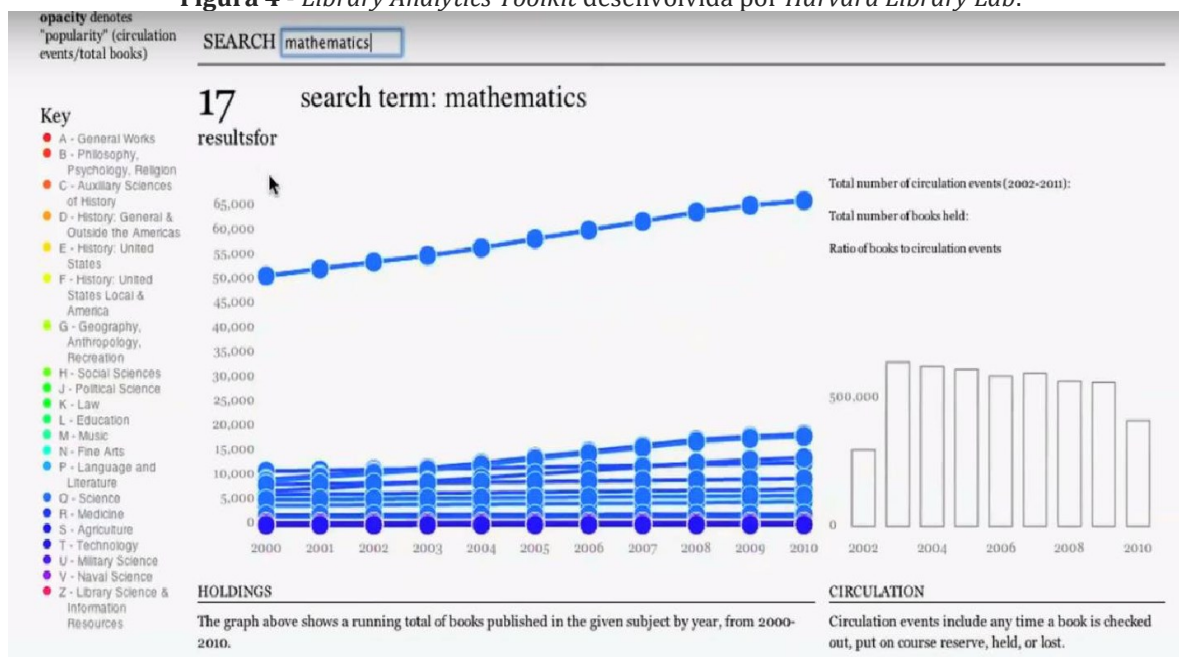
Fonte: *Washington University Libraries* (2017, p. 16).

O projeto *Geospatial Information Systems*, de colaboração entre as bibliotecas da *Washington University* e o departamento administrativo de *Facilities Planning and Management*, resultou em um mapa interativo do campus, no qual, visitantes, alunos e profissionais podem ter uma visão geral, via *desktop* e dispositivos *mobiles*, para localizar estacionamentos, bibliotecas setoriais, prédios acadêmicos, lanchonetes e ainda obter informações sobre os estabelecimentos (WASHINGTON UNIVERSITY LIBRARIES, 2017).

Outro exemplo se refere aos laboratórios de inovação da biblioteca da *Harvard University*, ao proporem expansão das funcionalidades da página *Check out the checkouts*, pois o *site* apresenta as obras devolvidas em formato de *ranking*. Essa funcionalidade, chamada de *Library Analytics Toolkit* (figura 4), permite, além de ser configurada por cada biblioteca setorial da Universidade individualmente, que os dados sejam estudados por

meio de um painel visual (HARVARD LIBRARY LAB, 2011), ou seja, conforme um “[...] *dashboard* de análise interativo que providenciaria informação detalhada e em tempo real sobre os eventos “[...] da biblioteca no sistema de Harvard” (HARVARD LIBRARY LAB, 2011, p. 2, tradução nossa), assim aproximando-se da velocidade presente nos “3V”.

Figura 4 - Library Analytics Toolkit desenvolvida por Harvard Library Lab.



Fonte: Harvard Library Lab (2011).

A figura 4 exemplifica o *dashboard* interativo desenvolvido pela *Harvard Library Lab*, permitindo visualizar, em *timeline*, informações de circulação referentes às obras relacionadas ao termo pesquisado pelo usuário. Por meio dessas informações, verificam-se comportamentos e interesses dos usuários e, assim, é possível decidir sobre a constituição do acervo da *Harvard University*. O gráfico de elementos azuis (*holdings*) apresenta o total de obras sobre certo assunto, no caso matemática, publicadas por ano, no intervalo de 2002 a 2010. Quanto ao gráfico em barras (*circulation*), corresponde a quantitativos por ano, de eventos de circulação, ou seja, inclui quando o livro foi emprestado, reservado, tirado da estante ou perdido.

Em 2004, o projeto *Making visible the invisible* propôs reunir os dados coletados pela *Seattle Public Library* e apresentá-los aos seus usuários por meio de painéis (LEGRADY, 2005). A proposta centra-se na visualização de mudanças no fluxo de informação ao longo do tempo, através do processamento de dados. Consequentemente, esses dados são apresentados eletronicamente, refletindo “[...] em tempo real a natureza

dinâmica da biblioteca como recurso de informação [...]” (LEGRADY, 2005, p. 5, tradução nossa), demonstrando com clareza alinhamento com a propriedade de velocidade presente na *big data* e, de forma menos aparente, a aproximação com a característica variedade, ao relacionar dados de tipos diversos.

Nos painéis, de hora em hora, são apresentadas informações sobre os livros e as mídias devolvidos (figura 5), em números e em ordem cronológica, bem como a frequência de palavras-chave indexadas, conectando-as em relacionamentos de igualdade e hierarquia com outros itens do acervo (LEGRADY, 2005).

Figura 5 - Painel da *Seattle Public Library* sobre empréstimos de obras.



Fonte: *Blog Future of the Book* (2005).

A *Seattle Public Library* explorou o uso da *big data*, não apenas para o tratamento das informações geradas, como também ser um meio de interagir com os usuários da Instituição, como uma espécie de campanha institucional ao proporcionar a divulgação dos serviços e produtos proporcionados, exemplificados pela circulação das obras do acervo e pela transparência dos resultados alcançados, neste caso, a partir de relatórios em tempo real. Desta forma, fazendo uma analogia aos aeroportos, as informações obtidas pela *Seattle Public Library* são apresentadas em painéis semelhantes aos quadros de informações dos voos, entre partidas e decolagens, atualizados de “tempos em tempos” e veiculando mensagens aos transeuntes.

Por fim, a partir das características da *big data* e possibilidades de sua aplicação, foi possível sugerir, mediante descrição, um conjunto de ações para o seu uso em bibliotecas (quadro 1). De forma ampla, as sugestões vão ao encontro da iniciativa da *Seattle Public Library*, por viabilizar (recomendar), tanto a aproximação do usuário com os serviços disponibilizados pela biblioteca, quanto a divulgação da oferta de produtos e serviços, além de métricas alcançadas pela unidade de informação, neste caso, por meio de painéis que possibilitam a visualização de informações, como quantidade de obras

emprestadas em um certo momento, materiais mais solicitados etc., retratando-se como um exemplo prático das características velocidade, veracidade, volume e valor. Em consequente, a *Harvard University Library* e a *Washington University Library* inspiram o setor da administração, especificamente no planejamento estratégico das bibliotecas e consequentes tomadas de decisão. A *Harvard University Library*, por meio da identificação de comportamentos organizacionais, do gerenciamento e da elaboração de relatórios dos eventos da biblioteca, combina as características da variedade, valor e complexidade. Quanto à *Washington University Library*, deslumbram-se possibilidades de colaborações entre setores diversos com o intuito de analisar e apresentar o alto volume de informações de maneira personalizada, mais precisa, além de considerar a devida complexidade para criar relações/conexões entre informações e, consequentemente, apoiar as cooperações setoriais e respectivas tomadas de decisão.

Deve-se destacar que *Harvard Library* e a *Washington Library* compartilham do estímulo de uso do atributo volume com fins de realizar estudos dos usuários. A identificação dos interesses, necessidades e comportamentos dos usuários, pelo setor de referência, por exemplo, serviria de informações ao setor de desenvolvimento de coleção, para direcionar a tomada de decisão na constituição do acervo, utilizando-se das características de complexidade, que conectaria informações diversas sobre as obras, os formatos variados em que estão disponíveis e o volume, sobre a quantidade de informações acerca do acervo, dos usuários e das possibilidades de busca e recuperação.

Quadro 1 – Sugestões de aplicação da *big data* em bibliotecas.

Característica da <i>big data</i>	Setor da biblioteca	Ação
Variedade; Valor; Complexidade	Administração	Identificar comportamentos organizacionais para embasar e direcionar o planejamento estratégico da biblioteca, a partir dos resultados alcançados.
Velocidade; Valor	Administração	Gerenciar e relatar eventos da biblioteca em tempo real, possibilitando tomadas de decisão com maior rapidez.
Volume; Complexidade	Administração	Permitir visão geral da situação da biblioteca, resultando no direcionamento às tomadas de decisão e visualização de novas oportunidades, como, por exemplo, possibilidades de novos projetos e colaborações entre setores.
Velocidade; Volume; Veracidade	Circulação	Aproximar e interagir com os usuários, via exibição de informações em tempo real, no momento do empréstimo e devolução de

		materiais, por exemplo. E apresentar, de maneira transparente, o fluxo (dos processos) de funcionamento da biblioteca e os resultados alcançados, tanto pela organização, quanto pelos setores específicos.
Complexidade; Volume; Variedade	Desenvolvimento de Coleção	Conectar informações acerca das obras, da relevância dos assuntos e do interesse sobre o comportamento dos usuários, aos quais subsidiam as decisões sobre a constituição do acervo.
Valor	Referência	Divulgar produtos e serviços ofertados pela biblioteca, além de apresentar os resultados alcançados, pela organização como um todo e/ou por seus setores, aos seus clientes, ou seja, aos usuários da biblioteca.
Volume	Referência	Realizar estudos de usuários para identificar interesses, necessidades e comportamentos, como forma de orientar, ajustar e desenvolver melhorias para os processos e setores da biblioteca, o que, conseqüentemente, trará melhorias aos usuários.

Fonte: Dados da pesquisa (2019).

Em suma, a partir dos elencados exemplos ao longo do artigo, em contextos geral e específicos nas bibliotecas, que utilizam a *big data*, acentuados por suas características básicas e pelo quadro sugestivo de aplicação do recurso em bibliotecas, infere-se que a *big data* é uma solução de considerável utilidade e versatilidade, em que agrega qualidade no atendimento aos usuários e agilidade na tomada de decisão pelas unidades de informação.

5 CONSIDERAÇÕES FINAIS

O presente artigo considerou a pirâmide informacional (dado, informação, conhecimento e sabedoria) como o fundamento de que processos cognitivos geram conhecimentos, que por sua vez produzem informações, que assumirão valor, ou não, e possibilitarão, ou não, a tomada de decisões estratégicas.

A partir do desenvolvimento da *web 3.0* surgem recursos tecnológicos como a *big data*, a qual constitui elemento interdisciplinar em áreas do conhecimento como Computação, Ciência de Dados, Marketing e Negócios, com o objetivo de facilitar o processo de integração entre dados e geração de novas informações. Conseqüentemente,

infere-se no avanço da pirâmide informacional até o topo, ou seja, a sabedoria, contribuindo no processo de tomadas de decisões.

Diante do cenário construído, de aplicações do paradigma da *big data*, identificaram-se exemplos nos segmentos de mercado (corporativo), governamental, científico e cotidiano social, os quais fundamentaram nas possibilidades de uso da *big data* em bibliotecas, especificamente: nos setores de gestão e de tomada de decisão das bibliotecas, ao possibilitar a identificação de “fragilidades” informacionais da organização, alinhar metas com recursos disponíveis e, tanto embasar, quanto traçar estratégias e objetivos para a biblioteca, como a promoção de cooperação entre setores; no mapeamento dos perfis e interesses dos usuários, o que permitiria captar tendências ao monitorar e acompanhar o uso do acervo, constituindo uma forma de estudo do usuário, e, a partir dos resultados, embasar decisões de novas aquisições ao acervo, bem como possibilitar um meio de integrar a biblioteca ao cotidiano do contexto que integra e dos usuários; por fim, na apresentação dos serviços e produtos ofertados pela biblioteca, como forma de engajar os usuários, assim como divulgar os resultados alcançados, tanto pela unidade de informação como um todo, quanto pelos setores, o que pode, conseqüentemente, estabelecer uma espécie de comunicação transparente entre usuários e organização.

Concluiu-se que as características “3V”, encaradas como básicas da solução *big data* e complementadas pelas propriedades complexidade, veracidade e valor, podem auxiliar na tomada de decisão em setores das bibliotecas, como na administração, na circulação, no desenvolvimento de coleções e na referência, em um cenário de produção e compartilhamento de informações promovidos pelas TIC, como a Internet e em sentido exponencial.

REFERÊNCIAS

AGHAEI, Sareh; NEMATBAKHS, Mohammad Ali; FARSANI, Hadi Khosravi. Evolution of the World Wide Web: from web 1.0 to web 4.0. **International Journal of Web & Semantic Technology**, v. 3, n. 1, jan. 2012. Disponível em: <https://pdfs.semanticscholar.org/8cb3/93c3229e8f288febfa4dac12a0f6298efb93.pdf>. Acesso em: 18 maio 2020.

ALVIM, Luísa. Blogues e bibliotecas: construir redes na Web 2.0. **Cadernos Bad**, n. 1, p. 38-74, 2007. Disponível em: <https://www.bad.pt/publicacoes/index.php/cadernos/article/view/781/780>. Acesso em: 19 maio 2020.

ARAÚJO JÚNIOR, Rogerio Henrique de; SOUZA, Renato Tarciso Barbosa de. Estudo do ecossistema de Big Data para conciliação das demandas de acesso, por meio da representação e organização da informação. **Ciência da Informação**, Brasília, v. 45, n. 3, p. 187-198, set./dez. 2016. Disponível em: <https://www.brapci.inf.br/index.php/res/v/18140>. Acesso em: 14 maio 2020.

BERNERS-LEE, Timothy John. **Linked data**. jun. 2009. Disponível em: <https://www.w3.org/DesignIssues/LinkedData>. Acesso em: 14 maio 2020.

BRIGGS, Asa; BURKE, Peter. **Uma história social da mídia: de Gutenberg à Internet**. 3. ed. Rio de Janeiro: Zahar, 2004.

CASTELLS, Manuel. **A galáxia da internet: reflexões sobre a internet, os negócios e a sociedade**. Rio de Janeiro: Zahar, 2003.

CHEDE, Cezar. **Um novo paradigma: stream computing e o InfoSphere Streams da IBM**. 2009. Disponível em: <https://www.ibm.com/developerworks/community/blogs/ctaurion/date/200908?lang=en>. Acesso em: 14 maio 2020.

CHEN, Chun-I Philip; ZANG, Chun-Yang. Data-intensive applications, challenges, techniques and technologies: a survey on big data. **Information Sciences**, v. 275, n. 10, p. 314-347, aug. 2014. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0020025514000346>. Acesso em: 21 maio 2020.

CHEN, Hsin-liang *et al.* Library assessment and data analytics in the big data era: practice and policies. **Proc. Assoc. Info. Sci. Tech.**, v. 52, n. 1, p. 1-4, 2016. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/pra2.2015.14505201002>. Acesso em: 21 maio 2020.

CHEN, Min; MAO, Shiwen; LIU, Yunhao. Big data: a survey. **Mobile Networks and Applications**, v. 19, n. 2, p. 171-209, apr. 2014. Disponível em: <https://doi.org/10.1007/s11036-013-0489-0>. Acesso em: 29 dez. 2020.

CHOUDHURY, Nupur. World Wide Web and its journey from Web 1.0 to Web 4.0. **International Journal of Computer Science and Information Technologies**, v. 5, n. 6, 2014. Disponível em: <http://ijcsit.com/docs/Volume%205/vol5issue06/ijcsit20140506265.pdf>. Acesso em: 26 maio 2020.

CONEGLIAN, Caio Saraiva *et al.* O papel da web semântica nos processos do big data. **Encontros Bibli**, v. 23, n. 53, p. 137-146, set./dez. 2018. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p137>. Acesso em: 15 maio 2020.

FICO FALCON. **FICO Falcon fraud manager** - customer success stories. 2016. Disponível em: <https://www.slideshare.net/FICO/fico-falcon-fraud-manager-customer-success-stories>. Acesso em: 1 maio 2020.

GARCÍA-MARCO, Francisco-Javier. El reto de la transparencia y la web 3.0: um nuevo contexto para las publicaciones oficiales. *In: ROVER, Aires José; GALINDO, Fernando (Eds.). O governo eletrônico e suas múltiplas facetas*. Zaragoza: Prensas Universitarias de Zaragoza, 2010. p. 241-264. Disponível em: http://www.egov.ufsc.br/portal/sites/default/files/lefis_10.pdf. Acesso em: 26 maio 2020.

HARVARD LIBRARY LAB. **Library Analytics Toolkit**: Library Lab grant proposal. 2011. Disponível em: <https://osc.hul.harvard.edu/liblab/projects/library-analytics-toolkit>. Acesso em: 20 maio 2020.

KATAL, Avita; WAZID, Mohammad; GOUDAR, R. H. Big data: issues, challenges, tools and good practices. *In: INTERNATIONAL CONFERENCE ON CONTEMPORARY COMPUTING*, 6., 2013, Noida, India. **Anais eletrônicos** [...]. IEEE, 2013, p. 404-409. Disponível em: https://www.researchgate.net/publication/261454906_Big_data_Issues_challenges_tools_and_Good_practices. Acesso em: 26 maio 2020.

LEGRADY, George. "Making visible the invisible": Seattle Library data flow visualization. *In: INTERNATIONAL CONFERENCE ON HYPERMEDIA AND INTERACTIVITY IN MUSEUMS*, 5., 2005, Paris. **Proceedings** [...]. Paris: Archives & Museum Informatics Europe, 2005. Disponível em: <http://www.archimuse.com/publishing/ichim05/Legrady.pdf>. Acesso em: 10 maio 2020.

NASCIMENTO, Diego de Lima *et al.* Big Data: a nova formula para a inovação, a concorrência e produtividade. *In: Congresso Acadêmico de Tecnologia e Informática*, 7., 2015, Mato Grosse. **Anais** [...]. Mato Grosso: UNEMAT, 2015. Disponível em: https://www.researchgate.net/publication/313468003_Big_Data_A_nova_formula_para_a_inovacao_a_concorrncia_e_produtividade. Acesso em: 25 maio 2020.

NONATO, Marcos Daniel de Arruda. **World Wide Web**: www. Tefé, AM: DanG3, 2015. Disponível em: <https://www.livrosdigitais.org.br/livro/1447A68L1BQ3T>. Acesso em: 24 maio 2020.

PRIMO, Alex. O aspecto relacional das interações na Web 2.0. **E-Compós**, Brasília, v. 9, ago. 2007. Disponível em: <https://www.e-compos.org.br/e-compos/article/view/153/154>. Acesso em: 20 maio 2020.

ROZADOS, Helen Beatriz Frota. **Indicadores como ferramenta para gestão de serviços de informação tecnológica**. 2004. 239 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004. Disponível em: <https://lume.ufrgs.br/handle/10183/5668>. Acesso em: 24 maio 2020.

SANTOS, Carla Marques do; ASSUNÇÃO, Suelene Santana. Biblioteca digital: uma evolução da biblioteca convencional. *In: ENCONTRO REGIONAL DE ESTUDANTES DE BIBLIOTECONOMIA, DOCUMENTAÇÃO, CIÊNCIA E GESTÃO DA INFORMAÇÃO*, 15., 2012, Juazeiro do Norte. **Anais** [...]. Juazeiro do Norte: Universidade Federal do Cariri, 2012. Disponível em: <https://www.brapci.inf.br/index.php/res/v/81194>. Acesso em: 24 maio 2020.

SAP SE. **GN**: processing data quickly for effective action with SAP HANA. 2017. Disponível em: <https://pt.slideshare.net/sap/gn-processing-data-quickly-for-effective-action-with-sap-hana>. Acesso em: 22 maio 2020.

TAURION, Cezar. **Big data**. Rio de Janeiro: Brasport, 2015.

THE INSTITUTE FOR THE FUTURE OF THE BOOK. **Making visible the invisible:** George Legrady installation at Seattle Central Library. New York, 16 sep. 2005. Blog: If:Book: a project of The Institute for the Future of the Book. Disponível em: <http://futureofthebook.org/blog/2005/09/16/making-visible-the-invisible-g/>. Acesso em: 13 maio 2020.

TORRES, Marcos Daniel Marado. **What is web 3.0?** ago. 2007. Disponível em: <https://mindboosternoori.blogspot.com/2007/08/what-is-web-30.html>. Acesso em: 13 maio 2020.

URDANETA, Iraset Páez. **Gestión de la inteligencia. Aprendizaje tecnológico y modernización del trabajo informacional:** retos y oportunidades. Caracas, Venezuela: Universidad Simon Bolivar, 1992.

WASHINGTON UNIVERSITY LIBRARIES. **Annual report 2017/2018.** Washington, 2018. 14 p. Disponível em: https://library.wustl.edu/wp-content/uploads/2018/12/Libraries_2017_18_AR_FINAL.pdf. Acesso em: 15 maio 2020.

WASHINGTON UNIVERSITY LIBRARIES. **Off the shelf.** Washington, fall 2017. 26 p. Disponível em: https://library.wustl.edu/wp-content/uploads/2017/12/120517_OTSF17.pdf. Acesso em: 15 maio 2020.

ZIKOPOULOS, Paul *et al.* **Understanding big data:** analytics for enterprise class hadoop and streaming data. New York: McGraw Hill, 2015.

Recebido em: 01 de março de 2021
Aprovado em: 02 de novembro de 2021
Publicado em: 08 de dezembro de 2021