

BIG DATA EM SAÚDE: DESAFIOS E PERSPECTIVAS NA VISÃO DA CIÊNCIA DA INFORMAÇÃO

CARLOS ALBERTO FERREIRA*

MARCIO GONÇALVES**

Resumo: *O presente trabalho aborda os principais desafios que a Área da Ciência da Informação enfrenta para a construção de soluções de Sistemas de Big Data, relacionados com a Área da Saúde. Analisa as perspectivas que podem ser trabalhadas e a contextualização de Sistemas de Big Data para a área da Saúde com o principal viés que é o uso e a recuperação da informação. Objeto de estudo e análise da Ciência da Informação.*

Palavras-chave: Big Data; Ciência da Informação; Ciência da saúde; Encontrabilidade da informação; Recuperação da informação.

Abstract: *This work addresses the main challenges that the Information Science Area faces in the construction of Big Data Systems solutions, related to the Healthcare Area. It analyses the perspectives that can be worked on and the contextualisation of Big Data Systems for the Healthcare Area with the main bias being the use and retrieval of information. Object of study and analysis of Information Science.*

Keywords: Big Data; Information Science; Health science; Information findability; Information retrieval.

INTRODUÇÃO

Com a explosão informacional e a partir do desenvolvimento das chamadas tecnologias de informação e comunicação (TIC), muitos pesquisadores têm se dedicado ao estudo da informação para uma melhor maneira de empregá-la com o objetivo de agregar valor a serviços e produtos, aumentando a produtividade, mapeando públicos e diversificando oportunidades.

A geração de informação pela área da saúde gira em torno de 30% de tudo que é produzido, principalmente no que tange ao setor de Assistência Médica. Conseguiu-se ver que tais dados, a partir de um único paciente crônico, geram em torno de 80 Mb de dados ao ano. Conforme apresentado no documento das redes de atenção à saúde do Ministério da Saúde do Brasil.

Em meio à disseminação rápida e até em tempo real da informação e o grande volume de dados propagados a cada instante, surge o *Big Data*, que é a reunião de várias bases de dados em valor, velocidade, veracidade, volume e variedade. Desse modo é importante

* Universidade Federal do Estado do Rio de Janeiro (UNIRIO). Email: carlos.cafe@unirio.br. ORCID: <https://orcid.org/0000-0002-7880-4119>.

** Faculdades Integradas Hélio Alonso (FACHA). Email: marciog.goncalves@gmail.com. ORCID: <https://orcid.org/0000-0001-5773-2139>.

ressaltar a importância da análise e interpretação do *Big Data* nas mais diversas formas de uso, como o da área da Saúde, por exemplo, para que seja empregado com eficiência nas mais diferentes perspectivas, aprimorando assim a qualidade de serviços, produtos e até o bem-estar da vida das pessoas.

As novas TIC, com o seu poder multiplicador e a sua versatilidade de aplicação a todas as tarefas humanas, passando pelo lazer, pela cultura e chegando até a saúde em todos os espectros, contribuíram de forma significativa para o avanço de um tipo de pensamento hegemônico.

O presente trabalho tem por objetivo apresentar um panorama do uso do *Big Data* e seus avanços, sob o contexto da chamada explosão informacional, tendo em vista a rapidez e o volume de dados disseminados atualmente. Nesse sentido, a metodologia escolhida foi a de referencial teórico, baseada nos estudos mais recentes sobre a temática abordada.

1. CONTEXTUALIZANDO BIG DATA

Como todo fenômeno, o *Big Data* tem uma causa que pode ser vista e analisada de forma ampla como, por exemplo, a popularização da Internet e a redução do custo dos equipamentos tecnológicos. Essas questões trouxeram uma maior disponibilidade, a redução substancial do custo de armazenamento de dados — o que favoreceu uma ampla possibilidade de interconexão de dispositivos — e a inclusão de «novas» formas de coleta de dados, como a Internet das Coisas, as funções de geolocalização, a inteligência artificial, entre outras.

O *Big Data* acabou se popularizando devido a um dos termos mais conhecidos da área de informação — «a explosão informacional» —, que possibilitou novas necessidades, não somente de coleta, mas, principalmente, no que tange ao armazenamento e disponibilização da informação.

Torna-se emergencial a criação de formas e recursos de armazenagem e representação da informação em ambientes digitais como os bancos de dados, mas, em especial, as estruturas de *Big Data* em saúde, trazendo, conseqüentemente, formas e maneiras mais efetivas de recuperação da informação.

A explosão informacional trouxe vários desafios para os profissionais da informação, onde dois dos principais são: a forma de coleta e a organização da informação armazenada. Uma das características mais marcantes da era pós-industrial é o aumento das redes de comunicação que usam computadores para se efetivar. Com esse fato, tem-se um aumento significativo de pessoas que puderam alcançar dados disponíveis nas redes. Deste advento surge, então, a sociedade em rede, que é tratada na pulverização das tecnologias da informação e conhecimento.

Diante desta revolução na produção de dados, a sociedade verificou que informações e, por conseguinte, o conhecimento, ora armazenados em sistemas de informação, passaram a ter mais relevância e valor para uso, em escala exponencial. O conceito

Big Data pode ser resumidamente definido como uma coleção de bases de dados tão complexa e volumosa que se torna muito difícil fazer algumas operações simples (remoção, ordenação, sumarização) de forma eficiente utilizando Sistemas Gerenciadores de Bases de Dados tradicionais. Em pleno século XXI, a Informação está intrinsecamente ligada à diversidade de dispositivos de entrada de dados e os inúmeros formatos que são oferecidos. Esse fenômeno, por assim dizer, de coleta em demasia de dados se nomeia em linhas gerais como *Big Data*.

A mineração e tratamento destes dados podem ser considerados um dos maiores desafios da atualidade. Isso devido não só pela heterogeneidade dos dados coletados, mas também, pela diversidade de uso de tais dados. Segundo Sayão e Sales (2012), dados não estruturados têm origem em um novo paradigma da ciência, orientada a dados, ou *e-Science*, paradigma este norteado pela necessidade de armazenamento, compartilhamento e reuso de dados científicos. Diante deste contexto de estruturação de dados para uso e reuso em Banco de Dados e em soluções de *Big Data*, um fato se apresenta como extremamente relevante: a criação e definição de padrões de sistematização de dados. Alves e Souza (2007) apresentam que a utilização de padrões de metadados, o uso de nomes e também de modelos informacionais permitem a integração e a troca via intercâmbio de dados e de informações entre outros sistemas de maneira mais fluida.

Diante da problematização de como seria e funcionaria uma sistematização de dados, pode-se ver Manyika et al. (2011) corroborando com tal afirmativa quando estes apresentam *Big Data* referindo-se a um conjunto de dados cujo tamanho está além da capacidade típica dos *softwares* tradicionais de Banco de Dados em capturar, armazenar, gerenciar e analisar um volume consubstancial de dados.

Em linhas gerais, pode-se dizer que *Big Data* é a capacidade de analisar um volume inédito e descomunal de dados digitais. Afirma-se que o *Big Data* é o termo usado para descrever o crescimento, a disponibilidade e o uso exponencial de informações estruturadas e não estruturadas.

Em uma linguagem técnica, *Big Data* é definido como o conjunto de soluções tecnológicas capazes de lidar com dados digitais em: Volume, Velocidade, Variedade, Veracidade e Valor.

Quando se fala de volume, este refere-se à capacidade de lidar com grande volume de dados, tanto em armazenamento como processamento, além de ser e representar o processo de escalabilidade. Em relação ao segundo «V», a velocidade trabalha o tempo de resposta dentro do necessário de acordo com a regra de negócio, suportando alta demanda.

O terceiro «V» é o da variedade que trabalha a habilidade de se adaptar em termos de infraestrutura elástica: os tipos de dados e sua multiplicidade de origens e as diversas formas de codificação. O quarto «V» trata da veracidade, que trabalha os mecanismos de pré-processamento e padronização: utilização de metodologias científicas e mecanismos de reprodutividade do processo.

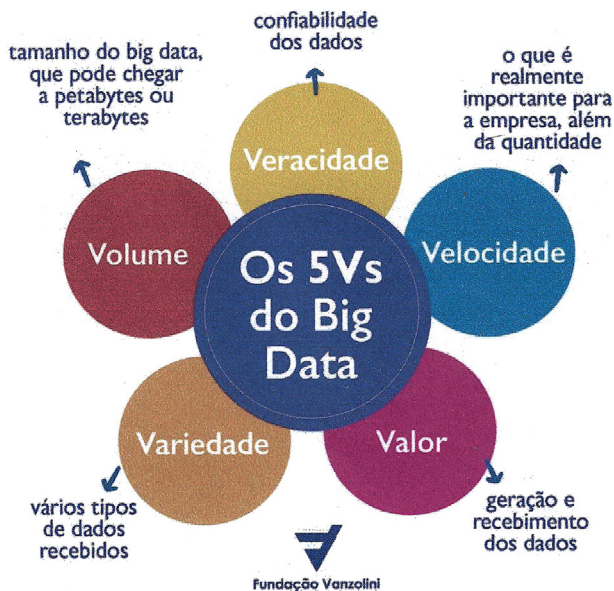


Fig. 1. Dados Digitais no *Big Data*
 Fonte: Fundação Vanzolini 2021

O quinto e último «V» fala do valor, que trabalha a união de todas as características antes mencionadas e tem o foco na entrega de valor às organizações, aos clientes, à sociedade e às pesquisas acadêmicas.

Na prática, a tecnologia de *Big Data* permite analisar qualquer tipo de informação digital em tempo real, sendo fundamental para a tomada de decisões, e tem como base sempre o grande volume de criação e disponibilização de dados correntes.

Já em outro momento tem-se que tratar que não há definições claras sobre o que constitui *Big Data*. Ward e Barker (2013) descobriram que provavelmente existem tantas definições quanto o número de pessoas que você pergunta. Eles pesquisaram as principais empresas do setor de tecnologia da informação e usaram os resultados para desenvolver esta definição na qual afirmam que *Big Data* é um termo que descreve o armazenamento e a análise de conjuntos de dados grandes e ou complexos, usando uma série de técnicas, incluindo, entre outras: NoSQL, MapReduce e aprendizado de máquina.

Dumbill et al. (2012) conceituam *Big Data* como sendo dados que excedem a capacidade de processamento dos sistemas de banco de dados convencionais. Os dados são grandes demais, se movem rápido demais ou não se enquadram nas restrições das arquiteturas de banco de dados. Para obter valor com esses dados, você deve escolher uma maneira alternativa de processá-los.

Para Mayer-Schönberger e Cukier (2014), *Big Data* é definido como coisas que se pode fazer em larga escala, e que não podem ser feitas em menor medida, para extrair novas ideias ou criar novas formas de valor em maneiras que mudam mercados, organizações, relações entre cidadãos e governos e muito mais.

Diante de tais fatores, um dos grandes desafios trabalhados no *Big Data* e que vão diretamente influenciar no seu apoio ao negócio, trata-se da coleta de dados.

2. **BIG DATA: MUITO MAIS QUE COLETA DE DADOS**

Em um mundo permeado pela inteligência artificial e conectado por redes, e possuindo uma maior quantidade de objetos industrializados, os indivíduos e os grupos, mais do que desenvolverem e utilizarem as tecnologias para adaptarem-se às suas necessidades, eles têm que desenvolver ou adquirir capacidades e habilidades complexas, para entender, minimamente, seu próprio ambiente.

É notório pensar na evolução da humanidade com relação à comunicação, mas, conforme Lovink apud Baldi e Oliveira (2014), «não é suficiente pensar como a Internet mudou o mundo, mas como o mundo transformou a Internet». Assim, tem-se quais as influências da sociedade no desenvolvimento de soluções *web*.

Neste contexto, a denominação *Big Data* é recente e trata-se de uma estratégia ou campo de estudo. Para alguns pesquisadores, trata-se de um termo difícil de entender. De acordo com Pedron, Goularte e Zilber (2015), não se trata apenas de uma ferramenta, mas sim, na verdade, de uma geração de novas tecnologias e arquiteturas projetadas para extrair valor econômico de grande volume de dados. Ou, pode-se dizer, ainda, que se refere mais comumente a um conjunto de dados, cujo volume está acima da habilidade de ferramentas tradicionais de bancos de capturar, gerenciar e analisar.

As técnicas usadas no *Big Data* podem ser tratadas nos mais variados campos obtendo resultados que não se limitam apenas pela criatividade do cientista de dados. Um dos usos mais comuns para o *Big Data* seria o *Analytics* em empresas que estão vinculadas para entender seus clientes e prever seus comportamentos. Usando desta forma dados de redes sociais, histórico de pesquisa e análise de texto, as empresas conseguem prever as necessidades dos seus clientes. Como foi no caso da empresa Target, que baseado no histórico de compras de uma cliente adolescente, concluiu que a mesma estava grávida antes mesmo de seu próprio pai saber.

Não são apenas as empresas que podem se beneficiar. Os dispositivos que colhem dados dos usuários e retornam informações para os mesmos usando o *Big Data Analytics* estão há anos no mercado, como, por exemplo das «pulseiras fitness»: as Jawbone UPs e Fibtis, que coletam dados como consumo de calorias, níveis de atividade e padrões de sono, dando em troca para o usuário *insights* de como ter uma vida mais saudável com gráficos informativos e dicas de como se exercitar e dormir melhor. Da mesma forma, as plataformas de *streaming* de vídeos usam dados dos seus usuários não só para o seu algoritmo de recomendação personalizada, mas, também, para desenvolver séries e filmes que seus usuários vão gostar.

As organizações estão criando um sistema de compartilhamento contínuo, cujo propósito é que todos os membros aprendam a trabalhar com novos processos e novas

soluções para compreender a importância da aprendizagem permanente vinculada às metas e estratégias organizacionais. Eboli (2004) corrobora com tal pensamento quando afirma que o objetivo principal de um sistema de compartilhamento é o desenvolvimento e a instalação de competências organizacionais e humanas, consideradas críticas para a viabilização das estratégias de negócio.

Quando se trata de *Big Data*, é importante ressaltar que na visão da Ciência da Informação, a representação da informação pode ser construída e realizada por elementos descritivos e temáticos. Conforme Maimone, Silveira e Tálamo (2011), os dois tipos de representação da informação ocorrem de forma fluida e contínua para facilitar a recuperação da informação pelo usuário. Neste contexto, é possível tratar uma representação com um cunho maior no âmbito semântico sobre a informação, objetivando assim aumentar a precisão na descrição de recursos informacionais, e, conseqüentemente, reduzir a revocação de informações trazidas para os usuários.

Algo que deveria ser um consenso, tanto para os fornecedores de soluções quanto para as organizações de usuários, é que uma solução que aborda a coleta, tratamento e análise de dados precisa ser uma decisão de negócio, e não uma decisão unilateral da área de Tecnologia. Mesmo sabendo-se que o valor do *Big Data* para as organizações é de imensa importância, não existe ainda uma abordagem padronizada sobre como iniciar e, principalmente, concluir tal projeto.

Para que um projeto de *Big Data* possa acontecer de forma satisfatória, necessita-se atender a pelo menos quatro etapas principais para que assim possa ser implementado.

1. Definição do caso de Negócio (*Business Case*);
2. Planejamento do projeto;
3. Definição dos requisitos técnicos;
4. Criação de um *Total Business Value Assessment*.

Estas etapas, quando cumpridas, ganham profissionalismo no tratamento técnico dos dados para sua posterior análise crítica dos resultados.

3. BIG DATA NA SAÚDE: ENTENDENDO ESSA LIGAÇÃO

No estudo de Bloom, Standing e Lloyd (2008), eles reconhecem que assimetrias são vistas para penetrar os mercados de saúde, que são caracterizados pelos altos níveis de incerteza. Por exemplo, pacientes podem conseguir descrever seus sintomas, mas eles têm informações inadequadas para relacionar sua condição a um determinado tipo de tratamento e curso da medicação. Isto cria uma relação desigual de poder entre especialistas e pacientes, onde o primeiro pode explorar em seu próprio interesse.

Uma aplicação considerada relevante para a área da saúde está direcionada com o que é possível no *Big Data Analytics*: decodificar toda a fita de DNA em minutos e permitir encontrar novas curas, ter um entendimento melhor e prever padrões de doenças.

Projetos como o Artemis, criado pelo Instituto de Tecnologia da Universidade de Ontário, no Canadá, juntamente com a IBM, que está sendo aplicado no Hospital para Crianças em Toronto, utilizam-se os dados coletados pelo *Big Data Analytics* no monitoramento de bebês doentes e prematuros. Gravando e analisando cada batimento e padrão respiratório de cada um dos pacientes, desta forma a unidade conseguiu desenvolver um algoritmo que consegue prever infecções com 24 horas de antecedência comparado com qualquer sintoma físico que se manifestasse.

Netto, Berton e Takahata (2021) trazem a definição de *Big Data* na área da Saúde apresentando que este refere-se a um conjunto de dados eletrônicos de saúde tão grandes e complexos que são difíceis ou quase que impossíveis de gerenciar com soluções de *softwares* e *hardwares* considerados tradicionais. Da mesma forma, não podem ser manipulados de forma fácil com ferramentas e métodos tradicionais ou comuns de gerenciamento de dados. Os autores continuam afirmando que entre os conteúdos que mais se utilizam do chamado *Big Data Analytics* estão dados oriundos de processamento de imagens, de sinais, genômicas e todos outros tipos de alimentadores como, por exemplo, prontuários de pacientes.

Para que se possa entender e desta forma exemplificar o uso de soluções de *Big Data* na área da Saúde, faz-se necessário apresentar modelos de suas arquiteturas, pois desta forma consegue-se ter uma maior amplitude de abordagem conceitual para o entendimento de como soluções de *Big Data* trazem formas de auxílio na tomada de decisões na área da Saúde.

Na Figura 2, tem-se um exemplo de Arquitetura de Banco de Dados relacional, que pode ser considerada tradicional, onde os *Data Warehouse* são abastecidos não só pelos *Data Marts*, mas também pelos usuários. Esta arquitetura trabalha com dados estruturados e semiestruturados, mas não contempla múltiplas fontes de informação, o que acaba impossibilitando um efetivo uso para dados não estruturados como fonte de entrada.

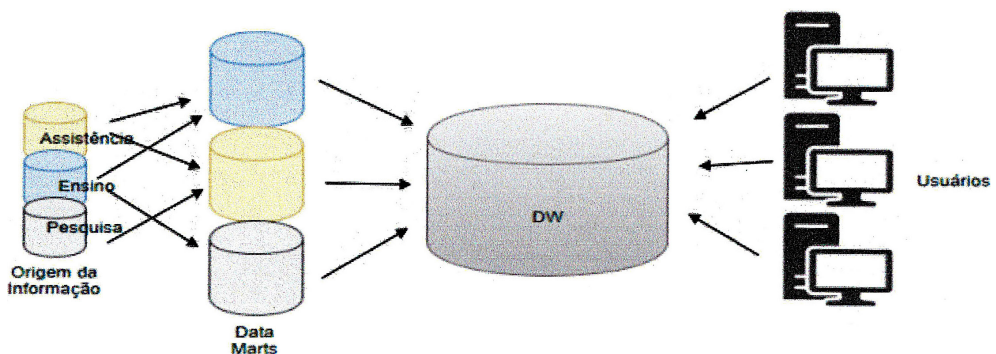


Fig. 2. Arquitetura de Banco de Dados Relacional
Fonte: Malossi 2018

É importante ressaltar que a estrutura geral de dados e, também, os recursos relacionados dos dados são partes integrantes das arquiteturas corporativas.

Tem-se, na Figura 3, a arquitetura de um Sistema de *Big Data* usando a solução Azure da Microsoft, onde o fluxo de dados segue não somente com a alimentação de dados brutos estruturados, mas pode ocorrer com dados semiestruturados, e também não estruturados, isto é, texto livre. Tais inserções podem ocorrer com qualquer tipo de dados, que podem ser dados comerciais, atividades dos usuários e outros tipos de fontes de informações.

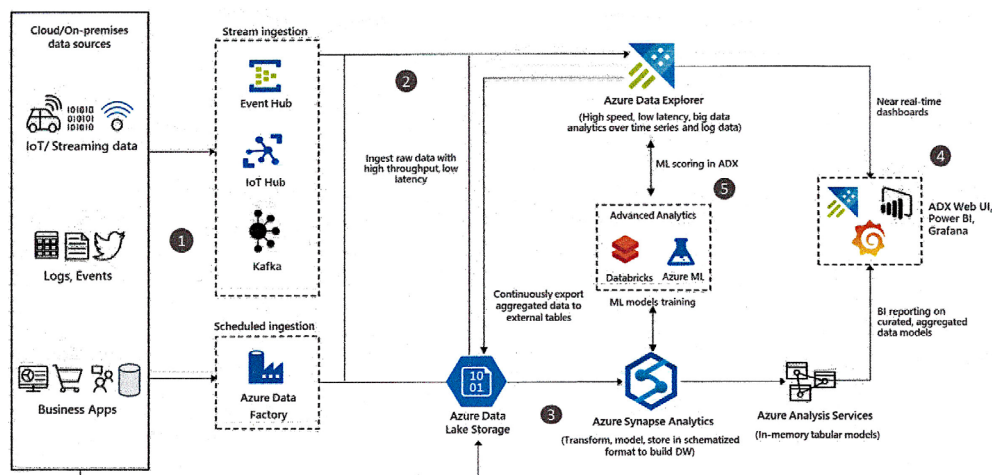


Fig. 3. Arquitetura de um *Big Data* usando uma ferramenta como Azure
 Fonte: Microsoft 2022

A inserção destes dados pode seguir não só pela sua latência de uso, isto é, alto ou baixo, mas também em suas interconexões com os dados armazenados já agregados à solução de *Big Data*.

Ao final da Arquitetura ora mencionada os resultados apresentados são além do reconhecimento de padrões: fornece análises de séries temporais e assim auxilia na detecção e previsão de anomalias e um aprendizado de máquina (*Machine Learning*).

Na Figura 4, tem-se um exemplo de uma arquitetura de *Big Data* específico para a área da Saúde, no qual procura-se respeitar possíveis estratégias de uso de dados com a finalidade de tomada de decisão. Tal arquitetura respeita a entrada de dados de modelos multidimensionais oriundos de diversas redes e com características de formatos analíticos e de grandes volumes, o que em aspectos técnicos são somente processados e analisados por Banco de Dados e sendo muito difícil para humanos de fazê-los.

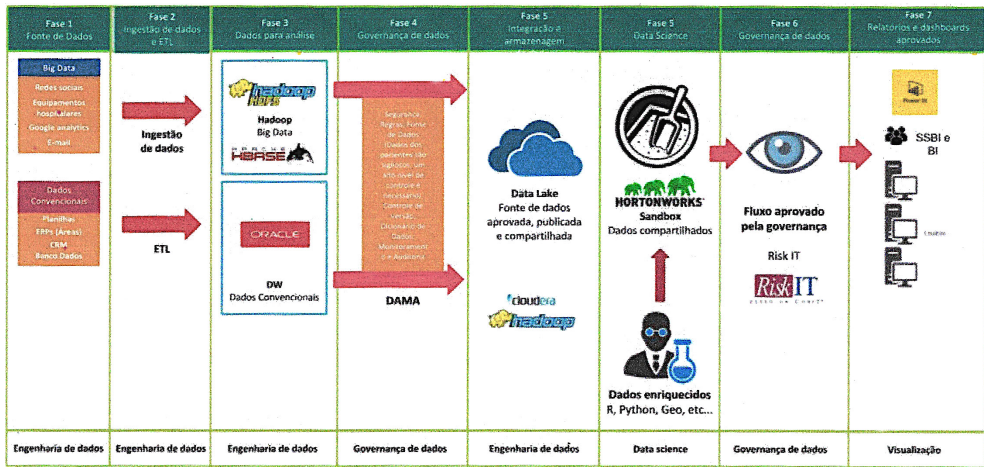


Fig. 4. Arquitetura de Big Data para a área da Saúde
 Fonte: Malossi 2018

Entradas diversas, como as de redes sociais, que são basicamente dados textuais, precisam ser armazenadas em estruturas de *Data Warehouse* ou de Banco de Dados conforme já feito por *Data Lake*. Ou, quem sabe, ainda, estocadas em uma outra estrutura de guarda de dados em rede.

Na abordagem dos três tipos de arquiteturas há exemplos de entrada de dados, assim como a forma de tratamento dos mesmos com o intuito de uma melhoria dos processos e assim alcançar um diferencial competitivo sustentável no qual tem-se mais opções para que um sistema baseado em redes possa tratar de forma híbrida algo que traga benefícios diretos para o sistema de recomendação desejada. Ou seja, apoiar as necessidades do usuário.

4. PERSPECTIVAS: O SISTEMA DE RECOMENDAÇÃO BASEADO EM BIG DATA

A expressão Recuperação da Informação (RI) é geralmente empregada para designar atividades relacionadas com localização de documentos que tratam de um determinado assunto. Um Sistema de Recuperação da Informação (SRI), por sua vez, fornece a ideia de componentes destinados a facilitar o acesso ao conteúdo dos documentos. De qualquer modo, a recuperação da informação está sempre voltada para o mesmo propósito: a necessidade do usuário. De acordo com Chowdhury (2015), a interface de um sistema é projetada com a finalidade de que os usuários possam interagir com o sistema facilmente, cujo propósito é o de buscar e recuperar a informação requerida.

O tipo de usuário, como se pode perceber, depende da natureza da informação. A informação a ser solicitada muitas vezes é oriunda de um problema ou necessidade

não solucionado. A isso, chama-se de necessidade informacional. O conceito de necessidade informacional não é exato, segundo Chowdhury (2015), pois é um conceito vago e que não permanece constante. O que se sabe é que ela pode ocorrer quando um indivíduo reconhece que seu estado atual de conhecimento é insuficiente para lidar com a tarefa em mãos, ou de forma a solucionar os conflitos numa certa área, ou, até mesmo, na hora de preencher um vazio em uma área do conhecimento. Chowdhury (2015) afirma que a necessidade de informação varia de pessoa para pessoa, mudando com o passar do tempo e mudando conforme novas informações chegam aos usuários.

O conceito de Sistemas de Recomendação (SR), assim como suas usabilidades, são há tempos conhecidos na área da Ciência da Informação. E tem como característica, como o próprio nome/termo já diz, o objetivo de oferecer alguma recomendação de um produto, serviço ou informação a um possível cliente ou a um usuário que busca, de maneira personalizada, suas necessidades de pesquisa. Desta maneira, os Sistemas de Recomendação são considerados como ferramentas que analisam, a partir de algoritmos, os dados disponíveis em formato *on-line* para indicar algo.

D'Arc (2021) afirma que nos anos 90 os Sistemas de Recomendação surgiram com o intuito de suprir a dificuldade de encontrar o que se procura (em alguns casos até a possibilidade de «induzir» na escolha, pois, muitas vezes, o usuário não sabe exatamente o que busca), por conta da vasta quantidade de informações disponíveis.

Historicamente, os Sistemas de Recomendação eram conhecidos como filtragem colaborativa, justamente por serem feitos a partir da colaboração de diversas pessoas. Entretanto, por não ser apenas realizado por meio da colaboração, o termo foi generalizado para Sistemas de Recomendação conforme afirma D'Arc (2021).

Tendo por intermédio a mineração de dados, os Sistemas de Recomendação conseguem localizar usuários e seus interesses em comum. A importância da mineração de dados se dá por esta ser identificadora dos padrões de dados que têm potencial para utilização sendo, conforme Garcia e Frozza (2013) afirmam, um sinônimo de KDD (*Knowledge Discovery in Database*) ou Descoberto (ou Extensão) de conhecimento em base de dados.

O uso de Sistemas de Recomendação para auxiliar no uso dos dados coletados e armazenados em bancos de *Big Data* são atualmente fontes de agregação de valor incalculável, pois tais dados e a consequente informação podem servir de fontes de informação potenciais para a área da saúde, trazendo desta forma mais agilidade aos diagnósticos e facilitando no processo de tomada de decisão no fluxo de trabalho médico.

Em qualquer sistema, mas em especial nos Sistemas de Recomendação, é importante a disponibilização de dados já estruturados que consigam abordar padrões predefinidos, possibilitando desta forma uma ampla recuperação e interpretação, favorecendo possíveis análises mais confiáveis. Para que se possa compreender melhor tem-se o pensamento de Sousa (2012) que trata de dizer que uma recuperação de forma satisfatória

exige uma representação adequada dos dados que ora são disponibilizados em Sistemas de Informação.

Um novo paradigma está sendo trabalhado quando se trata de Sistemas de Recomendação, pois a governança dos algoritmos já é algo extremamente presente nos dias atuais. Gray (2009) postulou alguns paradigmas não somente científicos, mas amplamente tecnológicos no qual dois merecem ser destacados. O primeiro diz respeito à Internet quando esta mídia está se tornando um sistema orientado a objeto, cada vez mais dependente de metadados e ontologias, vocabulários controlados e esquemas conceituais que possibilitam dar mais confiabilidade ao significado das coisas e à interação entre diferentes sistemas. Outro paradigma levantado por Gray (2009) diz respeito às bibliotecas de dados digitais que podem ser fomentadas, para além de bibliotecas de publicações acabadas.

5. O OLHAR DA CIÊNCIA DA INFORMAÇÃO PARA O *BIG DATA*

Na primeira década do século XXI, Humby (2018) cunhou a afirmação que os dados são o novo petróleo e tal fato está cada vez mais nítido, visto que, em vários nichos de mercado ou populacional, os dados possuem mais valor e agregam riquezas às pessoas e instituições que as utilizam.

O conceito e a relevância dos dados, para a Ciência da Informação, são notórios desde a primeira metade da década de 90 do século passado, quando Aldo Barreto publicou o texto *A Questão da Informação*. Barreto, nesta produção, mostrou que deveria ser necessário ir além das principais questões da informação. Este autor dedicou-se à evolução do conceito de dados e pretendeu mostrar como o seu entendimento influencia muito a estrutura do receptor.

A Biblioteconomia e a Ciência da Informação vêm ao longo do tempo estudando a explosão documental. Depois, passou-se a verificar o fenômeno da explosão da informação. Atualmente, com o advento do *Big Data*, pode-se dizer que há um latente interesse pelo estudo da explosão da produção de dados.

Baldi e Oliveira (2014) afirmam que o ciberespaço permite a todos introduzir documentos na rede, porém, nem sempre é possível recuperá-los. Tal posicionamento corrobora com um dos questionamentos da Ciência da Informação: a produção de dados de forma cada vez mais exponencial não permite com facilidade a devida recuperação, com isso faz-se necessário cada vez mais os estudos e análises sobre representação para posterior recuperação da informação.

E a relação *Boom* Informacional X *Boom* Documental X *Boom* de Dados Relacionais? Esta pergunta está baseada no fato pelo qual a sociedade atual tem uma produção vertiginosa de informação e conseqüentemente de conhecimento documentado. A cada momento se agiganta mais a produção de dados e em conseqüência a dificuldade que é de trazer o armazenamento e recuperação desta informação que se apresenta a cada instante de forma mais anárquica e com uma dificuldade de processamento e uso nunca vistos antes.

CONCLUSÕES PARCIAIS OU REFLEXÕES TEMPORÁRIAS

Um ponto a ser ressaltado é a relevância do estudo do *Big Data* na saúde, pois, assim, pode-se acompanhar a adequação de práticas de uso dos dados de acordo com o avanço das tecnologias digitais mais modernas. Faz-se necessário, portanto, garantir práticas de proteção dos dados dos usuários de forma a não deixá-los vulneráveis, mas que garantam e protejam a cidadania.

À medida em que nos movimentamos no dia a dia, deixamos pegadas e rastros digitais. Se esses dados forem computados para o controle da saúde e do bem-estar dos indivíduos, a sociedade avança para a chegada de um estágio de maior controle das doenças. Para isso, políticas públicas que pensam em usar tecnologias modernas a favor das pessoas são bem-vindas.

É relevante ressaltar que, atualmente, 60% dos projetos de *Big Data* não são concluídos, o que nos leva a refletir sobre se o problema estava no projeto de construção da solução e no qual não houve uma documentação específica analisando os usos e usuários de forma adequada.

Quando se refere a construção de soluções em especial na área da Saúde, onde os dados são mais sensíveis e necessitam de duas coisas que projetos de *Big Data* primam. A primeira é a quantidade e a segunda é a velocidade de processamento, visto que os mesmos não por conta do estrondoso volume de dados necessitam de um processo de processamento cada vez mais apurado e preciso a fim de apoiar a tomada de decisão dos gestores em saúde.

Faz-se necessário discutir, principalmente, como será feito o registro dos dados clínicos da população, em especial a coleta de dados. Afinal, eles são o ponto de partida porque os especialistas e os computadores acompanham a evolução do paciente e entendem os padrões epidemiológicos ou outra situação que a área da saúde necessite.

Fala-se muito em quantitativo de dados gerados e armazenados, onde acredita-se que anualmente 44 trilhões de *gigabytes* são produzidos pela humanidade. Tal dado leva a entender que somente a Inteligência Artificial e a Computação de Alto Desempenho podem conseguir tratar tantas alternativas, em especial para a Medicina de Precisão.

Pode ser visto que três fatores são fundamentais para que o projeto de *Big Data* consiga melhorar e chegar próximo ao seu objetivo. O primeiro deles é a forma como foi estruturada a informação. Em especial a forma de coleta dos dados disponíveis. Em um segundo momento, tem-se a localização e disponibilização da informação que faz toda a diferença para o tomador de decisão, em especial a questão da precisão para o uso imediato. No terceiro momento, tem-se a conexão das fontes de informação, o que para muitos está como o maior e mais difícil obstáculo a ser derrubado.

É notória a necessidade de discutir, tais fatos, mas quando se leva em conta a disparidade de tipos de entradas de dados, assim como os diversos tipos de fontes, nota-se que a barreira que deva ser derrubada ainda parece um tanto quanto intransponível visto

que os dados que se sabem para construção de um *Big Data* na área da saúde acabam sendo somente a ponta de um *iceberg* de tamanho proporcional à quantidade de dados gerados pela humanidade em apenas um ano.

Deve-se, sim, ter mais alternativas para não somente padronizar os tipos de dados de entrada nos sistemas de *Big Data*, mas também formas de agregação de dados e correlações a fim de auxiliar o usuário final que seria não somente o profissional da saúde, mas também o maior interessado: o paciente dos cuidados médicos.

Desta forma, é possível afirmar que nos próximos anos deve-se acompanhar grandes saltos na evolução da informação na área da saúde, muito baseados na oferta de tecnologias que procuram garantir a transformação do volume exponencial de dados gerados por máquinas e pessoas em informações preciosas para a prática de uma medicina mais personalizada e eficiente. Deste modo, poderá ser aberta uma nova era para o diagnóstico por evidências e a Medicina de precisão. O que se pode chamar de Medicina Baseada em Dados.

REFERÊNCIAS

- ALMEIDA, M. B., 2020. *Ontologia em Ciência da Informação: teoria e método*. Curitiba: CRV.
- ALVAREZ, E. B., et al., 2016. Os Sistemas de Recomendação, Arquitetura da Informação e a Encontrabilidade da Informação. *Transinformação*. Set./dez., **28**(3), 275-286.
- ALVES, M. D. R., e M. I. F. SOUZA, 2007. Estudo de correspondência de elementos de metadados: Dublin Core e Marc 21. *Revista Digital de Biblioteconomia e Ciência da Informação*. Jan/jun., **4**(2), 20-38.
- BALDI, V., 2014. Datificação: contra um saber sem conhecimento. Em: Jussara BORGES, org. *Mundo Digital: uma sociedade sem fronteiras?* João Pessoa: Ideia, pp. 75-88.
- BALDI, V., e L. OLIVEIRA, 2014. Luzes e sombras na cultura da convergência digital. Em: R. H. TAVARES, e S. S. GOMES, org. *Sociedade, educação e redes: desafio à formação crítica*. Araraquara: Junqueira & Marin, Parte 1, pp. 125-152.
- BLOOM, G., H. STANDING, e R. LLOYD, 2008. Markets, information asymmetry and health care: towards new social contracts. *Social Science & Medicine*. **66**(10), 2076-2087.
- CHAN, L., e M. ZENG, 2006. Metadata interoperability and standardization: a study of methodology: Part. I. *D-Lib Magazine*. Jun., **12**(6), 1-19.
- CHOWDHURY, G., 2015. Usuários da Recuperação da Informação. *InCID: Revista de Ciência da Informação e Documentação*. **5**(2), 4-33.
- CLAVIER, V., 2019. Knowledge Organization, Data and Algorithms: the new era of visual representation. *Knowledge Organization*. **46**(8).
- D'ARC, T., 2021. Por que usar um sistema de recomendação com inteligência artificial. Em: *SmartHint* [Em linha] [consult. 2023-05-10]. Disponível em: <https://www.smarthint.co/sistema-de-recomendacao-inteligencia-artificial>.
- DUMBILL, E., et al., 2012. *Big Data Now: 2012 Edition*. [EUA]: O'Reilly Media, Inc.
- EBOLI, M., 2004. *Educação corporativa no Brasil: Mitos e verdades*. 2.ª ed. São Paulo: Gente.
- FUNDAÇÃO VANZOLINI, 2021. *5 Vs do Big Data*. São Paulo. 6/10/2021. Em: X [Em linha]. São Paulo. 2021-20-06 [consult. 2023-05-10]. Disponível em: <https://x.com/vanzolini>.
- GARCIA, C. A., e R. FROZZA, 2013. Sistema de Recomendação de Produtos Utilizando Mineração de Dados. *Revista Tecno-Lógica* [Em linha]. Jun., **17**(1), 78-90 [consult. 2023-05-10]. Disponível em: <https://online.unisc.br/seer/index.php/tecnologica/article/view/3283>.

- GOLDSCHMIDT, R., E. PASSOS, e E. BEZERRA, 2015. *Data Mining: Conceitos, técnicas, algoritmos, orientações e aplicações*. 2ª. ed. Rio de Janeiro: Campus.
- GRAY, J., 2009. Jim Gray on eScience. A Transformed Scientific Method. Em: T. HEY, S. TANSLEY, e K. TOLLE, ed. *The Fourth Paradigm: data-intensive scientific discovery*. Redmond, WA, USA: Microsoft Research.
- HUMBY, C., 2018. *Data is the new oil* [Em linha] [consult. 2023-04-10]. Disponível em: <https://www.franklintempleton.co.uk/articles/insti/data-is-the-new-oil>.
- LOGAN, R. K., 2012. *O Que é informação? A propagação da organização na Biosfera, na simbologia, na tecnosfera e na econosfera*. Rio de Janeiro: Contraponto.
- MAIMONE, G. D., N. C. SILVEIRA, e M. DE F. G. M. TÁLAMO, 2011. Reflexões acerca das relações entre representação temática e descritiva. *Informação & Sociedade*. **21**(1).
- MALOSSI, J., 2018. Arquitetura de Big Data para o Hospital de Clínicas: Estudo de Caso. Em: *Médium* [Em linha] [consult. 2023-05-16]. Disponível em: <https://medium.com/joaomalossi/arquitetura-de-big-data-para-o-hospital-de-cl%C3%ADnicas-estudo-de-caso-47e967b3dc7>.
- MANYIKA, J., et al., 2011. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. São Francisco: McKinsey Global Institute.
- MAYER-SCHÖNBERGER, V., e K. CUKIER, 2014. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Londres: John Murray.
- MCGARRY, K. J., 1984. *Da documentação à informação: um contexto em evolução*. Lisboa: Presença.
- MICROSOFT, 2022. *Arquiteturas de Big Data* [Em linha] [consult. 2023-04-10]. Disponível em: <https://learn.microsoft.com/pt-br/azure/architecture/databases/guide/big-data-architectures>.
- NETTO, A. V., L. BERTON, e A. K. TAKAHATA, 2021. *Ciência de Dados e a Inteligência Artificial na Área da Saúde*. São Paulo: Editora dos Editores.
- PEDRON, C. D., A. GOULARTE, e S. N. ZILBER, 2015. Big Data: uma revisão sistemática da literatura de administração. Em: *12th CONTECSI International Conference on Information Systems and Technology Management, 20-22 May 2015*.
- SANJUAN, F. I., e G. C. BOWKER, 2017. Implications of Big Data for Knowledge Organization. *Knowledge Organization*. **44**(3).
- SAYÃO, L. F., e L. F. SALES, 2012. Curadoria Digital: um novo patamar para preservação de dados digitais de pesquisa. *Informação & Sociedade: Estudos*. **22**(3), 179-191.
- SCHESTATSKY, P., 2020. *Medicina do amanhã: como a genética, o estilo de vida e a tecnologia juntos podem auxiliar na qualidade de vida*. 3.ª ed. São Paulo: Gente.
- SILVA, F. C. C., 2019. *Gestão de dados científicos*. Rio de Janeiro: Interciência.
- SOARES, F., R. Y. HAMANAKA, e B. C. M. S. MACULAN, 2021. Interoperabilidade semântica no contexto de dados da Biodiversidade: um estudo de caso sobre a utilização de padrões de Metadados. Em: *XXI Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB, 25 a 29 de outubro de 2021, Rio de Janeiro*.
- SOUSA, M. R. F., 2012. O Acesso a informações e a contribuição da arquitetura da informação, usabilidade e acessibilidade. *Informação & Sociedade: Estudos*. **22**(especial), 65-76.
- WARD, S., e A BARKER., 2013. *Undefined by data: a survey of big data definitions* [Em linha] [consult. 2023-05-16]. Disponível em: <https://doi.org/10.48550/arXiv.1309.5821>.