

Big Data para a investigação em saúde e a ciência aberta: um contributo para a gestão do conhecimento

Big Data for health research and open science: a contribution to knowledge management

Jorge Magalhães

Investigador em Saúde Pública, Núcleo de Inovação Tecnológica de Farmangui-
nhos–NIT FAR. Coordenador Mestrado Gestão, P&D na Indústria Farmacêutica.
Fundação Oswaldo Cruz/FIOCRUZ, Ministério da Saúde do Brasil.
jorgemagalhaes@far.fiocruz.br

Zulmira Hartz

Professora catedrática convidada de Avaliação em Saúde; Global Health and
Tropical Medicine, GHTM, Instituto de Higiene e Medicina Tropical, IHMT,
Universidade Nova de Lisboa, UNL, Lisboa, Portugal.

Maria do Rosário O Martins

Professora catedrática; GHTM, Instituto de Higiene e Medicina Tropical. Global
Health and Tropical Medicine, GHTM, Instituto de Higiene e Medicina Tropical,
IHMT, Universidade Nova de Lisboa, UNL, Lisboa, Portugal.
mrfom@ihmt.unl.pt

Resumo

Este trabalho visa contribuir para a reflexão, em tempos de grandes volumes de dados a transitar na *web* diariamente, sobre a questão da saúde pública global e, no mesmo sentido, conjecturar o potencial da ciência aberta para a investigação científica e o contexto da sua complexidade na gestão do conhecimento. O volume de dados disponíveis ultrapassa 2,5 x 10¹⁸ bytes novos/dia na *web*. Esses dados apresentam-se nas mais variadas formas, como vídeos, fotos, artigos, relatórios etc. Assim, é necessário aferir a veracidade e valor dos mesmos para a correta tomada de decisão. Considerando que cerca de 43% dos dados são relativos à Saúde e que cerca de um milhão de artigos científicos publicados ao ano são da área da saúde, é mister pensar modelos não triviais para resolver problemas da saúde local com foco global. Neste cenário, a inteligência colaborativa, movimentos *web 2.0* e a Ciência Aberta, têm-se mostrado ferramentas essenciais para a solução de problemas das mazelas da humanidade e muitos mais ágeis do que a velha ciência fechada – grupos isolados sem partilhar os seus dados ou, quando compartilham, fazem-no a preços inacessíveis ao mundo em desenvolvimento ou subdesenvolvido. Portanto, sem esgotar o assunto, espera-se contribuir na importância da percepção do valor da informação para a área de saúde pública global, com uso de ferramentas e informações de acesso livre para a construção de uma ciência aberta e acessível a toda humanidade.

Palavras Chave:

Big Data em Saúde, *web 2.0*, ciência aberta, gestão do conhecimento, translação do conhecimento.

Abstract

This work aims to contribute to the discussion, in times of daily huge volumes of data for use on the web, of global public health issues. Furthermore, the potential of open science to the scientific research and the context of its complexity to the knowledge management. The volume of available data exceeds 2.5 x 10¹⁸ new bytes / day on the web. This data is presented in several forms, such as videos, photos, articles, reports etc. Thus, it is necessary to assess its veracity and value for a correct decision-making. Whereas about 43% of the data relating to health and about one million scientific articles published per year are of health area, to think new models is not trivial to solve problems with focus in local health but thinking globally. In this scenario, collaborative intelligence, web 2.0 and the Open Science movement have proven essential tools for troubleshooting the ills of humanity with much more agile than the old closed science - isolated groups without sharing any data or, when data share occurs, the prices are unaffordable to the developing world or underdeveloped. Therefore, without exhausting the subject, the paper expects to contribute to the importance of the perceived value of information for global public health using tools and free access to information for the construction of an open and accessible science to all mankind.

Key Words:

Big Data in Health, web 2.0, open science, knowledge management, knowledge translation.

Introdução

Com a entrada do século 21, o mundo vivencia uma era totalmente conectada à *internet*. Cerca de 40% da população está conectada à *internet* (“Big data”, 15:19:59; McKinsey Global Institute, 2011). Nesse sentido, O’Reilly (2007), sugeriu o termo *Big Data* – um gigantesco banco de dados atualizado em tempo real, que atinge facilmente milhares de *terabytes* de armazenamento em diversos formatos. Os sistemas tradicionais de gerenciamento de banco de dados relacional não podem lidar com essas grandes massas de dados (J. L. Magalhães & Quoniam, 2013; Quoniam, L., Lucien, A., 2010). O *Big Data* direciona uma nova geração de metodologias desenvolvidas para extrair valor econômico e estratégico de um grande e variado volume de dados (estruturados e não estruturados), permitindo alta velocidade de captura e análise (“Gray, J. and Chambers, L. and Bounegru, L., The Data Journalism Handbook, O’Reilly Media, 2012 - InfoVis:Wiki”, [s.d.]; O’Reilly, 2007a).

A análise de *Big Data* já é utilizada com sucesso em vários países, como os Estados Unidos, que incorporaram o conceito em quase todos os seus setores produtivos. Em 2014, o governo americano apresentou o relatório “Big data: Seizing opportunities. Preserving values”, com o objetivo de consultar os principais *stakeholders* americanos como a *Apple*, *IBM*, *Google*, *Bank of America*, entre outros, sobre questões acerca de oportunidades e valores no uso do *Big Data* e, assim, avaliar como este irá alterar as relações entre governo, cidadãos, empresas e os consumidores (The White House, 2016).

O *Big Data* refere-se à terceira geração da era da informação (Magalhaes, JL & Quoniam, L., 2015; Raghupathi & Raghupathi, 2014). Inicialmente, este volume exponencial de dados abordava os critérios dos 3Vs: Volume, Variedade e Velocidade (Laney, 2001), mais adiante, foram acrescentados mais 2 Vs: os atributos de Veracidade e Valor. Alguns autores ainda atribuem os últimos 3 Vs, como Veracidade, Versatilidade e Viabilidade, onde a combinação de todos os “Vs” geram o “V” de Valor (Aleixo & Duarte, 2015)

O volume é uma referência à quantidade de informações que são disponibilizadas diariamente na *web*, enquanto as diferentes fontes caracterizam o termo variedade. Já a veracidade e a velocidade estão relacionadas com o tratamento desses dados para que sejam proveitosos, com processamento em tempo real e fidedignos, respectivamente (Breternitz & Silva, 2013). O valor é o atributo em que o bom tratamento de *Big Data* poderá gerar acesso à informação essencial e economia aos cofres da organização em questão.

Segundo Minelli *et al* (2013), o *Big Data* divide-se em tempestade perfeita de dados, tempestade perfeita de

convergência e tempestade perfeita de computação, e esta última é resultante de 4 fenômenos: lei de *Moore*, computação móvel, redes sociais e computação em nuvem (*cloud computing*). Este acervo de dados deve ser tratado para apresentar informação pesquisada de forma seletiva e objetiva para aumentar a inteligência dos negócios, além de permitir uma melhoria no processo de tomada de decisão (Minelli, M., Chambers, M., & Dhiraj, A., 2013).

A velocidade e o volume de disposição de dados no mundo virtual têm mostrado a grandiosidade do *Big Data*, bem como a capacidade de utilizar os resultados dessas informações para a área da inteligência competitiva. Ou seja, produzir mais em menos tempo, superar o concorrente e economizar tempo. O resultado dos tratamentos de dados tem se mostrado um diferencial para as tomadas de decisões.

Segundo Wikipediaminer¹ *apud* Magalhães *et al* (2013), do volume diário de dados adicionados à *web*, 47% são relacionados com Saúde e, destes, 43% referenciam-se à saúde pública (Magalhães, 2014). Portanto, é *mister* pensar em novas ferramentas para identificação, extração e análise de dados do *Big Data* relativos à saúde no âmbito global e glocal (regional). Segundo Humbert (2005), o termo “glocal” está correlacionado na atitude de pensar nos problemas de forma global e agir de forma a atuar localmente, pois ações “glocais” poderão ter um impacto global. Este pensamento, reflete a globalização, onde a tecnologia trouxe menos adaptação tecnológica a um contexto único e a participação de empresas situadas em locais menos privilegiados (Humbert, M., 2003).

O relatório “The real-world use of big data” realizado pela IBM, em colaboração com a Universidade de Oxford, revelou que a análise de *Big Data* permite que as organizações sejam até 23 vezes mais propensas a superarem os seus concorrentes de mercado do que aquelas que não analisam. No entanto, alguns setores produtivos são mais adaptáveis às estratégias de *Big Data* do que outros, como os setores de saúde, financeiro, de telecomunicações, governamental e o energético.

A saúde é considerada, maioritariamente, como um bem público global: que não seja excludente, isso é, que ninguém ou nenhuma coletividade seja excluída da sua posse ou de seu consumo; e de que os seus benefícios sejam disponíveis a todos. Há, também, o aparente consenso de que a saúde não seja concorrencial, e que não haja rivalidade, isso é, a saúde de uma pessoa não pode se dar a expensas da exclusão de outras pessoas (Buse & Waxman, 2001; Haines *et al.*, 2009; Hartz, 2012; Vance, Howe, & Dellavalle, 2009).

O processo da globalização é o motor da evolução do termo “Saúde Global”, que carrega desafios e oportunidades no campo da saúde. Saúde Global pode ser compreendida ao mesmo tempo como uma condição, uma

atividade, uma profissão, uma filosofia, uma disciplina ou um movimento. Todavia, deve-se considerar que não há consenso sobre o que seja Saúde Global, nem uma única definição, e seu campo de ação tem limites imprecisos (Fortes & Ribeiro, 2014), contudo é indiscutível que vivemos a Saúde em tempos de globalização (Koplan et al., 2009).

O fenómeno da globalização traz novas dimensões espaciais, temporais e cognitivas. Modifica a nossa percepção das distâncias e barreiras das fronteiras aos contactos globais; modifica a nossa percepção de tempo, conectando a vida cotidiana com acontecimentos que ocorrem noutras partes do planeta, modificando a nossa percepção cognitiva de como nos vemos e entendemos e ao mundo que nos cerca, permitindo o engajamento com o "outro" no mundo (Bozorgmehr, 2010).

Assim, é mister buscarmos identificar, extrair e tratar o *Big Data* da Saúde existente neste mundo globalizado, a fim de focarmos a informação essencial para os tomadores de decisão do presente século. Portanto, o início do século 21, mostra que não é trivial a existência de ferramentas que possam gerenciar e "tratar" a grande quantidade de dados científicos e tecnológicos da área da saúde, como no século anterior, os métodos tradicionais funcionavam devido a quantidade muito menor. Com o pensamento de Ciência Aberta, este volume a ser "tratado" é compartilhado livremente e favorece no avanço da ciência mais rápida na resolução de problemas não triviais.

A expressão ciência aberta (*open science*), faz referência a um modelo de prática científica que, em consonância com o desenvolvimento da cultura digital, visa a disponibilização das informações em rede de forma oposta à pesquisa fechada dos laboratórios. Atualmente a expressão também se refere a geração de materiais de pesquisa que são compartilhados abertamente, sem a necessidade de patentes.

A Comunidade Europeia tem mostrado maturidade nesta área com a promoção da Ciência Aberta. Pode citar-se como exemplo a Resolução do Conselho de Ministros de Portugal:

Tornar a ciência mais aberta e acessível a todos representa um desafio coletivo, político, cultural, económico e social. A promoção e a defesa de uma prática generalizada de Ciência Aberta significam a assunção de uma política científica comprometida com um paradigma de partilha do conhecimento, de aproximação da ciência à sociedade, envolvendo as suas diversas componentes na formulação de agendas de investigação, em processos colaborativos e participativos de investigação, na procura de respostas conjuntas aos desafios e problemas que se lhe colocam. A criação de condições e mecanismos efetivos de acesso e de partilha do conhecimento democratiza-o e contribui para a igualdade na forma-

ção e na capacitação científica, possibilitando a transferência de conhecimento e estimulando a apropriação social da ciência. A implementação da Ciência Aberta envolve a incorporação de metodologias, ferramentas e práticas de natureza colaborativa e requer o compromisso dos diversos agentes implicados na produção, divulgação e utilização do conhecimento. Reforça -se deste modo a transparência, a integridade e a reprodutibilidade da ciência, potenciando ainda a prática mais eficiente e sustentável da atividade científica, designadamente ao nível das suas lógicas de publicação, disseminação e comunicação. Ciência aberta significa mais que a partilha seletiva de dados e publicações, representa a abertura do processo científico enquanto um todo, reforçando o conceito de responsabilidade social científica (Resolução Conselho de Ministros, 2016).

Desta forma, as organizações que lidam com investigação em saúde devem procurar manusear melhor a gestão do conhecimento do *Big Data* em Saúde, cada vez mais numa ciência aberta, a retratar uma inteligência colaborativa e construtiva para a sociedade.

A evolução da *Web*, *Big Data*, Inteligência Colaborativa e a Saúde global

As últimas décadas foram marcadas pelo advento da *Web 2.0*, onde a *internet* deixou de ser uma plataforma puramente estática e passou a desempenhar um papel dinâmico e interativo (O'Reilly, 2007b), permitindo que os usuários troquem uma grande quantidade de informação instantaneamente. Até ao final de 2016, a quantidade de informação criada e replicada a partir de sensores de todos os tipos, postagens em redes sociais, *upload* de fotos e vídeos, registros de transações comerciais, sinais de GPS, rastros de navegação, entre outros, alcançará a ordem de *zettabytes* - 1 bilhão de *Gigabytes* ("Big data", 15:19:59; Cattell, Chilukuri, & Levy, [s.d.]).

O infográfico dinâmico *The Internet in Real Time* mostra que cerca de 1 milhão de *Gigabytes* de dados são gerados a cada 1 minuto na *internet*, totalizando um lucro de 142 mil dólares por minuto aos gigantes deste meio, como a *Apple*, *Google*, *Microsoft*, *Facebook*, *Netflix*, *Pandora*, *Linkedin*, etc. Nesse contexto, de grande quantidade de dados gerados a todo instante, surge o termo *Big Data*.

A *web* tem evoluído desde a versão 1.0 até a versão 4.0. Cabe destacar que os termos *internet* e *web* são facilmente confundidos e geralmente são tratados como sinóni-

1 - WikipediaMiner é um conjunto de ferramentas de mineração de dados para a extração de dados semânticos codificados dentro da Wikipedia. <http://wikipediaminer.cms.waikato.ac.nz/>

mos pelos usuários (Nova Spivack, 2013). Nos últimos anos, os avanços tecnológicos e o desenvolvimento da internet têm conduzido à criação de diversos aplicativos utilizados principalmente para facilitar a interação entre as pessoas. Com isso, a sociedade moderna pôde ser apresentada aos computadores pessoais, aos *e-mails*, à *web*, ao comércio eletrônico, aos sistemas de busca, às tecnologias de rede sem fios, às músicas *on-line*, aos vídeos *on-line*, aos *smartphones* e às diversas redes sociais. Hoje a *internet* é uma rede de pessoas e comunidades, deixando de ser apenas uma rede de computadores.

No contexto da evolução da *web*, tem-se a globalização, que já não é um exclusivo das maiores empresas multinacionais do mundo. Hoje a digitalização eliminou muitas das barreiras que impediam que pequenas e médias empresas (PME), empresários e cidadãos comuns fizessem as suas conexões com clientes e fornecedores em todo o mundo. A globalização digital, por exemplo, tem implicações significativas, principalmente no que tange às empresas e economias dos países em desenvolvimento. Nessas nações, as empresas e indivíduos podem usar plataformas digitais como uma forma de superar as restrições nos seus mercados locais. Desta forma é possível identificar oportunidades, informações e ideias em qualquer lugar do mundo.

Considerando os aspectos da globalização digital, a literatura acadêmica destaca-se sobremaneira, com acesso praticamente instantâneo graças à mídia social e outras plataformas da *internet*, pois os indivíduos estão-se conectando. Estima-se que quase um bilhão de pessoas ao redor do mundo são participantes diretos de alguma forma de globalização. Análises do *Facebook*, *Twitter*, *LinkedIn* e *WeChat*, mostram que 914 milhões de pessoas têm pelo menos uma conexão internacional em uma plataforma de mídia social.

Pode-se considerar que o desenvolvimento de uma investigação não é mais linear. A velocidade dessa evolução tem um ritmo acelerado. O ritmo de mudança de paradigma, de novas ideias tem sido muito acelerado no mundo da era do conhecimento. Destaca-se que as primeiras mudanças, embora aparentemente rápidas, levaram anos para se desenvolverem. Exemplifica-se pelo sequenciamento do HIV que levou 15 anos e o sequenciamento do SARS, que levou somente 21 dias.

Dado o avanço tecnológico em todas as áreas da ciência, aliado à evolução tecnológica, é mister o tratamento do grande volume de dados, ainda mais dos dados da área da saúde, responsável por 47% de toda a informação. O volume exaustivo de dados, requer que o mesmo seja organizado e estruturado para o possível subsídio ao tomador de decisão. Nesse sentido, a utilização de



Figura 1. Análise do termo *Big Data* e “Saúde”
Fonte: extraído pelos autores em CarrotLingo4g.

ferramentas de engenharia de busca e posterior tratamento desse volume, como, por exemplo, uma plataforma livre *Carrot Search Lingo4G®*, auxilia sobremaneira o tratamento dos dados disponíveis (J. Magalhães, Bastos, & Barroso, 2016). Somente como exemplo, de modo a estimar-se a quantidade de informação disponível referente a “*big data*” e “saúde” (*health* em inglês), realizou-se inicialmente busca no *Carrot Search Lingo3G®* com o termo “*big data*”. Foram extraídos 349.000.000 resultados. Destes, após a mineração, obteve-se os termos mais “evidentes” no objeto pesquisado (que mais se repetem) e indexados somente na base *PubMed* e, posteriormente, segregando os que possuíam o termo “*health*” no primeiro resultado com o termo “*big data*”, foram obtidos 100 *clusters* que possuíam o total de 5.564 documentos essenciais, classificados de acordo com as principais incidências, como se nota na Figura 1.

Em razão da quantidade crescente de informação adicionada diariamente aos bancos de dados, analisar o estado da arte científica e tecnológica e extrair a informação essencial para tomada de decisão, torna-se tarefa quase impossível empregando-se os meios tradicionais. Nesse âmbito, a tecnologia da informação tem contribuído com ferramentas que auxiliam este processo em qualquer área da ciência. Não obstante, além dos *softwares* pagos (*textmining*, *datamining* etc.), as ferramentas de acesso livre da *Web 2.0* proporcionam uma opção para tratar volume de dados para as organizações que não possuem uma robusta facilidade de aquisição e manutenção de *softwares* e mecanismos de busca privados (licenças de uso). Desta forma, ao utilizar mecanismos de acesso livre, há razoabilidade em obter a informação

essencial para agilizar o processo de decisão, pois refinam dados, ordenam cronologicamente, agrupam temas e conferem ao tomador de decisão vantagens e praticidade para obter informações essenciais e estratégicas. Até agora, muitas análises da relação entre globalização e saúde encararam a saúde como subproduto, como consequência espontânea – positiva, segundo alguns, negativa, segundo outros – de forças globalizadoras estranhas a essa exigência e motivadas somente por outros interesses. A saúde global é uma finalidade social desejável, hoje descuidada ou deformada pela influência do fundamentalismo monetário, mas merecedora de evidência prioritária, seja pelo seu valor intrínseco, seja como símbolo do predomínio de valores humanos sobre outros interesses (Berlinguer, 1999).

No que tange a este processo de forma eficaz, é mister o trabalho cooperativo. Segundo Pierre Lévy (1994) “a inteligência coletiva é uma Inteligência distribuída por toda parte, incessantemente valorizada, coordenada em tempo real, que resulta em mobilização efetiva das competências”, que procura o reconhecimento e o enriquecimento das pessoas. O conceito da inteligência coletiva foi criado a partir de alguns debates realizados por Pierre Lévy (1994), relacionados com as tecnologias da inteligência. Caracteriza-se pela nova forma de pensamento sustentável através de conexões sociais que se tornam viáveis pela utilização das redes abertas de computação da *internet*. As tecnologias da inteligência são representadas especialmente pelas linguagens, os sistemas de signos, recursos lógicos e pelos instrumentos dos quais nos servimos. Todo o nosso funcionamento intelectual é induzido por essas representações. Os seres humanos são incapazes de pensar só e sem o auxílio de qualquer ferramenta (Bembem & Santos, 2013).

Segundo Bonabeau (2009), a inteligência coletiva contribui fortemente para a mudança de conhecimento e poder do indivíduo para o coletivo (Bonabeau, E., 2009). O código aberto da inteligência coletiva acabará por gerar resultados superiores aos conhecimentos gerados pelo *software* proprietário desenvolvido dentro das corporações. A educação e a forma como as pessoas estão a aprender a participar em culturas de conhecimento, fora os contextos de aprendizagem formais, é determinante no novo contexto global. É crucial a aprendizagem através dos meios de inteligência coletiva, pois é importante para a democratização da ciência, uma vez que está interligada com a cultura baseada no conhecimento e sustentada pela partilha de ideia coletiva, portanto, a contribuir para uma melhor compreensão da diversidade sociedade (Burke, 1991; “Collective intelligence”, 2016; Trigo, Gouveia, Quoniam, & Riccio, 2007), com dados abertos à ciência das instituições (Pordes et al., 2007; Molloy, 2011; David, 2004; Erington et al., 2014).

O contributo da Ciência Aberta

Para pensar em ciência aberta, é necessário mudar a cultura académica, de forma a amenizar o sentimento de posse que os pesquisadores têm relativamente aos seus dados. Na maioria das vezes eles são fruto de pesquisas financiadas com dinheiro público e contam com material doado por outras pessoas. Como exemplo, os pesquisadores financiados pelo *National Institutes of Health* (NIH) dos EUA, têm a obrigação de tornar os seus resultados públicos, sob pena de não voltarem a ser financiados.

Grandes corporações, utilizam também o conceito de “open innovation” como forma de proporcionar a inovação dos seus processos e/ou produtos com contributos de qualquer investigador, empresa etc. (Celadon, 2014; Michelino, Cammarano, Lamberti, & Caputo, 2015) e na área da saúde também se traduz numa oportunidade (Chaifetz, Chokshi, Rajkumar, Scales, & Benkler, 2007).

Segundo um editorial da revista *Science* (2016), embora tenha havido progresso nos últimos anos, a maior parte dos dados clínicos e genómicos ainda são coletados e estudados de forma isolada, em silos – compartimentados por doença, por instituição, por país etc. (Nature News, 2016). Os primeiros esforços de compartilhamento têm permitido desenvolver tratamento para doenças raras e algumas formas de cancro. Porém, tal benefício só atingirá toda a população quando médicos e pesquisadores puderem aceder e comparar dados de milhões de indivíduos (*The Global Alliance for Genomics and Health*, 2016).

Imagine a dificuldade de encontrar um livro específico se o conteúdo de uma dúzia de diferentes bibliotecas nacionais fosse todo reunido num único local e, em seguida, conceber uma forma de integrar as várias maneiras com que os diferentes conteúdos são arquivados, rastreados, gravados e disponibilizados. Seria muito mais fácil pedir a cada biblioteca que guardasse os seus próprios livros e que compartilhasse a informação sobre como encontrá-los em cada biblioteca. Nesta linha de pensamento, seria interessante que a partilha de dados de ciência pudesse seguir o mesmo caminho (Nature News, 2016).

Uma das iniciativas em prol da ciência aberta que se pode citar é o convénio do Brasil no Consórcio do Genoma Estrutural (www.thesgc.org), o qual tem o objetivo de gerar pequenas moléculas inibidoras de proteínas quinases e estas moléculas estarão disponíveis para qualquer grupo de pesquisa do Brasil e do mundo. Nesse contexto, novas iniciativas de publicação e revisão por pares, como o *Peerage of Science*, o *arXiv* e a *PLoS*, têm confirmado essa tendência, como o próprio exemplo da *Wikipedia*, que foi tão criticada no seu início e que hoje se traduz numa das melhores ferramentas *Web 2.0* para

difusão do conhecimento (“Collective intelligence”, 2016; Nakayama, Ito, Hara, & Nishio, 2008; WikiEdu, 2016).

Na busca dessa maior interação e aludindo ao “ano da ciência”, a Fundação de Amparo à Pesquisas do Estado de São Paulo (FAPESP), no Brasil, tem promovido eventos nesta área. Neste pensamento, pesquisadores identificaram que havia uma lacuna de 30 anos entre a ciência refletida nas fontes *Wikipédia* em português e os desenvolvimentos de ponta sendo feito nesse campo noutros idiomas (WikiEdu, 2016). A iniciativa brasileira adicionou milhares de palavras sobre temas relacionados com propriedades matemáticas da dinâmica neurais à *Wikipédia* em português. Por exemplo, expandiram o artigo em português sobre a doença de Alzheimer e temas mais complexos como a lesão do plexo braquial. Também criaram a introdução a modelos biológicos de neurónios, e criaram um vídeo que explica “Spike sorting” – uma maneira de rastrear e medir as propriedades elétricas das células – que aparece em ambas as edições (em português e inglês) da *Wikipédia*.

Em seu modelo, especialistas e pesquisadores trabalham para explicar conceitos para voluntários do Grupo de Usuários Wikimedia no Brasil. Eles vão escrever artigos com base no envolvimento de especialistas. Muitos dos autores são pesquisadores de pós-doutoramento (Chia-vegatto Filho, 2015; WikiEdu, 2016).

Uma outra Instituição brasileira que criou, e incentiva, a política de ciência aberta ao conhecimento e a informação científica é a Fundação Oswaldo Cruz (FIOCRUZ) (Pinheiro, 2014). Essa política visa contribuir para fortalecer os mecanismos de preservação da memória institucional e aumentar o acesso e o impacto da produção intelectual da Fiocruz, constituindo-se num importante instrumento que promoverá, de forma organizada e reunida, a disseminação, acessibilidade e, conseqüentemente, visibilidade do conhecimento gerado na Instituição.

O seu Repositório Institucional (Arca) é o principal instrumento de realização dessa política, atendendo ao objetivo de: reunir, hospedar, preservar, tornar disponível e dar visibilidade à produção científica da instituição. Dessa forma, é obrigatório o depósito no Repositório Institucional Arca das dissertações e teses defendidas nos Programas de Pós-graduação da Fiocruz e dos artigos produzidos no âmbito da Fiocruz publicados em periódicos científicos. Além de texto, o repositório pode conter imagem e áudio. O repositório Arca deverá ter capacidade de integração com sistemas nacionais e internacionais que possibilite automaticamente a inclusão e coleta da produção intelectual pertinente, observando em especial os protocolos e padrões definidos no modelo *Open Archives Initiative* (OAI).

A União Europeia está a dar um salto nas questões de abertura de conhecimento. Acabou de ser publicado um

estudo, na forma de livro, intitulado *Open innovation, open science, open to the world*. Este estudo foi encomendado ao Comissário Europeu para a Investigação, Inovação e Ciência pelo próprio presidente da União Europeia. A publicação apresenta a abertura do conhecimento como um processo natural advindo do uso das potencialidades das novas tecnologias da informação. Aponta que estimular a abertura do conhecimento é o caminho para melhor aproveitamento dos recursos públicos investidos em ciência e inovação, entre diversas outras vantagens, como as vantagens educacionais. Acrescentam-se, ainda, princípios para a abertura do conhecimento através de inovação aberta, ciência aberta, ciência cidadã, citando indiretamente a *Wikipédia* (EU Bookshop, 2016).

A publicação sugere a utilização de um conceito chamado “Global Research Area”. Neste conceito “pesquisadores e inovadores podem trabalhar com colegas internacionais onde pesquisadores, conhecimento científico e tecnologia circulam tão livremente quando possível”. Para o Brasil, as tecnologias desejadas para ensino de ciências e engenharias são aquelas que tem as propriedades das “Global Research Area”, ou seja, “o conhecimento científico e tecnologia circulam tão livremente quando possível”.

Cabe destacar, o evento que teve lugar a 23 de junho, realizado pela Universidade NOVA de Lisboa, nesta vertente, o seminário sobre *Big Data*, Desenvolvimento Sustentável e a Ciência Aberta, promovido pelo Instituto de Higiene e Medicina Tropical (IHMT) e o *Institut Français*, em parceria com a FIOCRUZ, no Brasil, e a Agência Nacional de Inovação de Lisboa. O seminário contou com 57 participantes, que assistiram às sessões presencialmente e através de *streaming*. Em debate estiveram dois grandes temas: a aplicação de *big data* em investigação e na utilização de ciência aberta para a transferência de tecnologia. No contexto do primeiro tema, foi analisado o uso de *big data* e de informação aberta em saúde. No segundo, promoveu-se a discussão sobre a inovação para a inclusão social e contextos de baixa-renda, assim como a análise automatizada de patentes.

Considerações finais

- O século 21 trouxe novos desafios e oportunidades devido ao crescente volume de dados novos adicionados diariamente à *web*. No campo da ciência e desenvolvimento tecnológico não é diferente, principalmente na área da Saúde. Desta forma, é iminente o desenvolvimento de novas metodologias para identificação, extração e tratamentos de dados para obter a informação essencial. Portanto, “minerar” o “*Big Data*” em saúde é premente e emergente, pois proporciona agilidade nos processos para tomadas de decisão. Uma das alternativas

para organizações a fim de auxiliar nesse processo são as ferramentas *Web 2.0*.

- Políticas de ciência aberta têm-se difundido nesta nova era do século XXI como forma de resposta ao avanço tecnológico e no afã da sociedade do conhecimento, procurar soluções para as mazelas de forma mais rápida e dinâmica. A Europa e EUA estão na vanguarda do processo.
- O trabalho de investigação individual tem perdido

espaço acadêmico nos últimos tempos, onde o trabalho coletivo, pela inteligência colaborativa, tem logrado êxito não somente para instituições privadas mas também para institutos públicos, em resposta aos anseios da sociedade. Destaque para a área da saúde, onde avanços científicos e tecnológicos foram alcançados em questão de dias, apenas com o cruzamento (*data mining*) de dados abertos ao invés de anos, se comparado com investigações tradicionais do antigo século.

Bibliografia

1. ALEIXO, J. A., & DUARTE, P. (2015). BIG DATA OPPORTUNITIES IN HEALTHCARE. HOW CAN MEDICAL AFFAIRS CONTRIBUTE? *Revista Portuguesa de Farmacoterapia*, 7, 230–236.
2. Bembem, A. H. C., & Santos, P. L. V. A. da C. (2013). Inteligência coletiva: um olhar sobre a produção de Pierre Lévy. *Perspectivas Em Ciência Da Informação*, 18(4), 139–151. <http://doi.org/10.1590/S1413-99362013000400010>
3. Berlinguer, G. (1999). Globalização e saúde global. *Estudos Avançados*, 13(35), 21–38. <http://doi.org/10.1590/S0103-40141999000100003>
4. Big data: The next frontier for innovation, competition, and productivity | McKinsey Global Institute | Technology & Innovation | McKinsey & Company. (15:19:59). Recuperado 7 de março de 2013, de http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation
5. Bonabeau, E. (2009). Decisions 2.0: The Power of Collective Intelligence. *MIT Sloan Management Review*, 45–52.
6. Bozorgmehr, K. (2010). Rethinking the “global” in global health: a dialectic approach. *Globalization and Health*, 6, 19. <http://doi.org/10.1186/1744-8603-6-19>
7. Burke, R. R. (1991). Reasoning with empirical marketing knowledge. *International Journal of Research in Marketing*, 8(1), 75–90. [http://doi.org/10.1016/0167-8116\(91\)90008-U](http://doi.org/10.1016/0167-8116(91)90008-U)
8. Buse, K., & Waxman, A. (2001). Public-private health partnerships: a strategy for WHO. *Bulletin of the World Health Organization*, 79(8), 748–754. <http://doi.org/10.1590/S0042-96862001000800011>
9. Cattell, J., Chilukuri, S., & Levy, M. ([s.d.]). How big data can revolutionize pharmaceutical R&D | McKinsey & Company. Recuperado 24 de agosto de 2016, de <http://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/how-big-data-can-revolutionize-pharmaceutical-r-and-d>
10. Celadon, K. L. (2014). Knowledge Integration and Open Innovation in the Brazilian Cosmetics Industry. *Journal of Technology Management & Innovation*, 9(3), 34–50. <http://doi.org/10.4067/S0718-27242014000300003>
11. Chaifetz, S., Chokshi, D. A., Rajkumar, R., Scales, D., & Benkler, Y. (2007). Closing the access gap for health innovations: an open licensing proposal for universities. *Globalization and Health*, 3(1), 1. <http://doi.org/10.1186/1744-8603-3-1>
12. Chiavegatto Filho, A. D. P. (2015). Uso de big data em saúde no Brasil: perspectivas para um futuro próximo. *Epidemiologia E Serviços de Saúde*, 24(2), 325–332. <http://doi.org/10.5123/S1679-49742015000200015>
13. Collective intelligence. (2016, agosto 2). In *Wikipedia, the free encyclopedia*. Recuperado de https://en.wikipedia.org/w/index.php?title=Collective_intelligence&oldid=732642532
14. David, P. A. (2004). Understanding the emergence of “open science” institutions: functionalist economics in historical context. *Industrial and Corporate Change*, 13(4), 571–589. <http://doi.org/10.1093/icc/dth023>
15. Errington, T. M., Iorns, E., Gunn, W., Tan, F. E., Lomax, J., & Nosek, B. A. (2014). An open investigation of the reproducibility of cancer biology research. *eLife*, 3, e04333. <http://doi.org/10.7554/eLife.04333>
16. EU Bookshop. (2016). *Open innovation, open science, open to the world - Research policy and organisation - EU Bookshop*. Recuperado de <http://bookshop.europa.eu/en/open-innovation-open-science-open-to-the-world-pbK10416263/>
17. Fortes, P. A. de C., & Ribeiro, H. (2014). Saúde Global em tempos de globalização. *Saúde E Sociedade*, 23(2), 366–375. <http://doi.org/10.1590/S0104-12902014000200002>
18. Gray, J. and Chambers, L. and Bounegru, L., *The Data Journalism Handbook*, O’Reilly Media, 2012 - InfoVis:Wiki. ([s.d.]). Recuperado 6 de março de 2014, de http://www.infovis-wiki.net/index.php?title=Gray,_J._and_Chambers,_L._and_Bounegru,_L._The_Data_Journalism_Handbook,_O%27Reilly_Media,_2012
19. Haines, A., McMichael, A. J., Smith, K. R., Roberts, L., Woodcock, J., Markandya, A., ... Wilkinson, P. (2009). Public health benefits of strategies to reduce greenhouse-gas emissions: overview and implications for policy makers. *The Lancet*, 374(9707), 2104–2114. [http://doi.org/10.1016/S0140-6736\(09\)61759-1](http://doi.org/10.1016/S0140-6736(09)61759-1)
20. Hartz, Z. M. A. (2012). Meta-evaluation of health management: Challenges for “new public health”. *Ciencia E Saude Coletiva*, 17(4), 832–834. <http://doi.org/10.1590/S1413-81232012000400004>
21. Humbert, M. (2003). *Globalization and glocalization: problems for developing countries and policy (supranational, national and subnational) implications*. In Cassiolato, J. E., Lastres, H. M. M. and Maciel, M. L. (eds). *Systems of Innovation and Development*. Cheltenham, UK: EE Elgar. Recuperado de https://www.e-elgar.com/shop/systems-of-innovation-and-development?__website=uk_warehouse
22. Koplan, J. P., Bond, T. C., Merson, M. H., Reddy, K. S., Rodriguez, M. H., Sewankambo, N. K., ... Consortium of Universities for Global Health Executive Board. (2009). Towards a common definition of global health. *Lancet (London, England)*, 373(9679), 1993–1995. [http://doi.org/10.1016/S0140-6736\(09\)60332-9](http://doi.org/10.1016/S0140-6736(09)60332-9)
23. Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety* (Application Delivery Strategies) (p. 4). Recuperado de <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
24. Magalhães, J., Bastos, A. C., & Barroso, W. (2016). Cenário Global e Global das Tendências Científicas e Tecnológicas em Diabetes: Uma Abordagem do Big Data em Saúde no Século 21. *Revista Gestão em Sistemas de Saúde*, 5(1), 1–14. <http://doi.org/10.5585/rgss.v5i1.191>
25. Magalhães, J. L., & Quoniam, L. (2013). Perception of the Information Value for Public Health: A Case Study for Neglected Diseases: Library and Information Science Book Chapter | IGI Global. In *Rethinking the Conceptual Base for New Practical Applications in Information Value and Quality* (p. 345). IGI Global. Recuperado de <http://www.igi-global.com/chapter/perception-information-value-public-health/84218>
26. Magalhaes, J.L., & Quoniam, L. (2015). Percepção do valor da informação por meio da inteligência competitiva 2.0 e do Big Data na saúde. In *Análise da Informação para Tomada de Decisão: desafios e soluções* (Vol. 1, p. 365). Brasil: Kira Tarapanoff (Org.).
27. McKinsey Global Institute. (2011). *Big Data: The Management Revolution - Harvard Business Review*. Recuperado de <http://hbr.org/2012/10/big-data-the-management-revolution/ar/1>
28. Michelino, F., Cammarano, A., Lamberti, E., & Caputo, M. (2015). Knowledge Domains, Technological Strategies and Open Innovation. *Journal*

of *Technology Management & Innovation*, 10(2), 50–78.

29. Minelli, M., Chambers, M., & Dhiraj, A. (2013). *Big Data, Big Analytics*. EUA: John Wiley & Sons, Inc. Recuperado de https://books.google.com.br/books/about/Big_Data_Big_Analytics.html?hl=pt-BR&id=Mg3WvT8uHV4C

30. Molloy, J. C. (2011). The Open Knowledge Foundation: Open Data Means Better Science. *PLoS Biol*, 9(12), e1001195. <http://doi.org/10.1371/journal.pbio.1001195>

31. Nakayama, K., Ito, M., Hara, T., & Nishio, S. (2008). Wikipedia Mining for Huge Scale Japanese Association Thesaurus Construction (p. 1150–1155). IEEE. <http://doi.org/10.1109/WAINA.2008.37>

32. Nature News. (2016). The ups and downs of data sharing in science. *Nature*, 534(7608), 435–436. <http://doi.org/10.1038/534435b>

33. Nova Spivack. (2013). Articles | Nova Spivack - Minding the Planet. Recuperado 28 de janeiro de 2013, de <http://www.novaspivack.com/articles>

34. O'Reilly, T. (2007a). *What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software* (SSRN Scholarly Paper No. ID 1008839). Rochester, NY: Social Science Research Network. Recuperado de <http://papers.ssrn.com/abstract=1008839>

35. O'Reilly, T. (2007b). *What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software* (SSRN Scholarly Paper No. ID 1008839). Rochester, NY: Social Science Research Network. Recuperado de <http://papers.ssrn.com/abstract=1008839>

36. Pinheiro, L. V. R. (2014). Do acesso livre à ciência aberta: conceitos e implicações na comunicação científica. *Revista Eletrônica de Comunicação, Inovação & Inovação em Saúde*, 8(2). <http://doi.org/10.3395/reciis.v8i2.629>

37. Pordes, R., Petravick, D., Kramer, B., Olson, D., Livny, M., Roy, A., ... Quick, R. (2007). The open science grid. *Journal of Physics: Conference Series*,

78(1), 12057. <http://doi.org/10.1088/1742-6596/78/1/012057>

38. Quoniam, L., Lucien, A. (2010). *Intelligence compétitive 2.0 : organisation, innovation et territoire*. France: Librairie Lavoisier. Recuperado de <http://www.lavoisier.fr/livre/notice.asp?ouvrage=2139418&pos=8>

39. Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2. <http://doi.org/10.1186/2047-2501-2-3>

40. Resolução Conselho de Ministros. Resolução do Conselho de Ministros n.º 20/2016 – Diário da República n.º 70/2016, Série I de 2016-04-11, Série I de 2016-04-11 (2016). Recuperado de <http://legislacaoportuguesa.com/resolucao-do-conselho-de-ministros-n-o-202016-diario-da-republica-n-o-702016-serie-i-de-2016-04-11/>

41. The Global Alliance for Genomics and Health. (2016). A federated ecosystem for sharing genomic, clinical data. *Science*, 352(6291), 1278–1280. <http://doi.org/10.1126/science.aaf6162>

42. The White House. (2016). *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights* (p. 29). EUA. Recuperado de https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

43. Trigo, M. R., Gouveia, L. B., Quoniam, L., & Riccio, E. L. (2007). *Using competitive intelligence as a strategic tool in a higher education context*. (D. Remenyi, Org.). Nr Reading: Academic Conferences Ltd.

44. Vance, K., Howe, W., & Dellavalle, R. P. (2009). Social Internet Sites as a Source of Public Health Information. *Dermatologic Clinics*, 27(2), 133–136. <http://doi.org/10.1016/j.det.2008.11.010>

45. WikiEdu. (2016). The Year of Science spreads to Brazil – Wiki Education Foundation. Recuperado 25 de agosto de 2016, de <https://wikiedu.org/blog/2016/08/02/brazil-year-of-science/>