

Resumo: O mundo está vivenciando uma pandemia provocada pela variante do Coronavírus nomeado SARS-CoV-2, causando a doença infecciosa COVID-19. Ações globais estão sendo realizadas no combate a esse vírus, dentre as quais a Vírus Outbreak Data Network (VODAN), que visa estabelecer uma infraestrutura de dados federada alinhada aos princípios FAIR e que apoie a coleta de dados de prontuários de pacientes infectados por vírus de alto contágio. A FIOCRUZ, que coordena a rede GO FAIR Brasil Saúde, assumiu a rede VODAN BR em parceria com a UNIRIO e UFRJ, com a participação inicial dos hospitais Gaffrée e Guinle, Municipal São José e o Israelita Albert Einstein. Este artigo objetiva apresentar a infraestrutura nacional alinhada às diretrizes e à rede internacional, para coleta e gestão de dados de prontuários de pacientes, desses hospitais, infectados com a COVID-19, para estabelecer modelo informacional que apoie o enfrentamento desta e de outras possíveis pandemias.

Palavras-chave: Coronavírus; Gestão de Dados; Princípios FAIR; SARS-CoV-2; VODAN.

Abstract: The world is experiencing a pandemic caused by the variant of the Coronavirus named SARS-CoV-2, causing the infectious disease COVID-19. Global actions are being carried out to combat this virus, including the Virus Outbreak Data Network (VODAN), which aims to establish a federated data infrastructure aligned with the FAIR principles and which supports the data collect from medical records of patients infected with high contagion. FIOCRUZ, which coordinates the GO FAIR Brazil Health network, took over the VODAN BR network in partnership with UNIRIO and UFRJ, with the participation of the hospitals Gaffrée and Guinle, Municipal São José and Israelita Albert Einstein. This article aims to present the national infrastructure aligned with the guidelines and the international network, for collecting and managing data from patient records, of these hospitals, infected with COVID-19, to establish an information model that supports the confrontation of this and other possible pandemics.

Keywords: Coronavirus; Data Management; FAIR principles; SARS-CoV-2; VODAN.

1. Contextualização

A pandemia atual está relacionada a uma nova variante do coronavírus, denominada SARS-CoV-2, causadora da doença COVID-19 (do inglês, *Coronavirus Disease* 2019). Esta é uma doença infecciosa emergente, identificada pela primeira vez na cidade chinesa de Wuhan, cujos casos iniciais foram divulgados em dezembro de 2019. Desde então, o surto da COVID-19 adquiriu dimensão pandêmica, com casos confirmados em mais de 100 países, incluindo o Brasil.

Diante deste cenário trágico que se espalhou rapidamente por todo o planeta, ações globais de diversas naturezas vêm sendo implementadas no combate a esse novo coronavírus, dando oportunidade para que as práticas da ciência aberta, tão discutida nas comunidades

científicas, atuem como um agente disseminador dos dados, informação e conhecimento de forma célere e efetiva.

A iniciativa **GO FAIR**¹ internacional, que tem a missão de disseminar e implementar os serviços e princípios FAIR, um acrônimo para *Findable, Accessible, Interoperable e Reusable*. Wilkinson *et al.* (2016), juntamente com as iniciativas **CODATA**, **RDA** e **WORD DATA SYSTEM** se juntaram e formaram a **DATA TOGETHER**², que visa a união de esforços para otimizar o ecossistema global de dados de pesquisa e identificar as oportunidades e necessidades que irão acionar as infraestruturas federadas para atender à nova realidade da ciência orientada por dados. Para isso, lançaram a Rede **Vírus Outbreak Data Network (VODAN)**³ sob a necessidade urgente de estabelecer uma infraestrutura de dados federada, alinhada aos princípios FAIR e que apoie a captura de dados de prontuários médicos de pacientes infectados por vírus de alto contágio. Segundo Mons (2020), essa rede foi concebida para dar o pontapé inicial em uma ‘comunidade’ que projeta e constrói uma infraestrutura de dados distribuídos internacionalmente e interoperáveis, que oferece suporte a respostas baseadas em evidências para os surtos de vírus.

No momento, a VODAN está voltada para a coleta e o gerenciamento dos dados de pacientes que carregam o coronavírus, de acordo com os princípios FAIR. Dessa forma, será possível tornar os dados dos pacientes contaminados disponíveis para serem reutilizados em pesquisas, interligá-los com dados de outras fontes, fazer diferentes monitoramentos, além de realizar previsões de impacto e controle voltados para as medidas de enfrentamento de epidemias, em condições bem definidas, respeitando a privacidade dos pacientes, conforme a legislação vigente.

A VODAN Internacional está desenvolvendo ainda uma ontologia e modelo conceitual simplificado a fim de permitir que os especialistas realizem o mapeamento semântico necessário e publiquem os dados em um formato padrão, o *Resource Description Framework* (RDF). Seguindo também os princípios FAIR, este formato, legível por máquina, permite o processamento automatizado dos dados, apoiando mecanismos robustos de exploração e integração de dados. Todo o material desenvolvido será "reutilizável", o que significa que o mesmo processo e infraestrutura podem ser usados para qualquer outro surto de vírus que futuramente surja. É importante ressaltar que esse modelo conceitual possibilitará que “a inteligência artificial trabalhe nos dados por meio da identificação de correlações reais e autênticas. É necessário um trabalho de análise conceitual para separar o "sinal" do "ruído de fundo", o que realmente é, do que apenas parece ser” (FATTOR, 2020).

A rede VODAN Internacional conta com a participação de dezenove países sendo eles: Alemanha, Áustria, Brasil, Checoslováquia, China, Espanha, Estados Unidos, França, Grécia, Holanda, Hong Kong, Itália, Luxemburgo, Noruega, Reino Unido, Eslovênia,

¹ Disponível em: <https://www.go-fair.org/>. [Consult. 25 dez 2020].

² Disponível em: <https://www.go-fair.org/go-fair-initiative/data-together/>. [Consult. 25 dez 2020].

³ Disponível em: <https://www.go-fair.org/implementation-networks/overview/vodan/>. [Consult. 25 dez. 2020].

Suécia, Suíça e Zimbábue. A rede VODAN África e Ásia é um exemplo de como a VODAN está sendo usada para apoio no combate ao coronavírus naqueles países. Informações adicionais como parceiros, leituras, repositórios etc, podem ser encontradas no *site* VODAN África & Ásia⁴.

2. A VODAN BR

A **Virus Outbreak Data Network Brazil (VODAN BR)**⁵, que é o braço brasileiro da rede VODAN Internacional, coordenada pela Fundação Oswaldo Cruz (FIOCRUZ), a Universidade Federal do Rio de Janeiro (UFRJ) e a Universidade Federal do Estado do Rio de Janeiro (UNIRIO), deu início a coleta de dados de pacientes confirmados com COVID-19, de forma anonimizada, seguindo o formulário recomendado pela Organização Mundial de Saúde (OMS) em três hospitais parceiros, localizados no Estado do Rio de Janeiro os hospital Universitário Gaffrée Guinle e Municipal São José, e no Estado de São Paulo o hospital Israelita Albert Einstein.

A rede VODAN BR comunga do mesmo entendimento da necessidade brasileira de colaborar com a diminuição da carência de infraestrutura tecnológica e humana para lidar com as questões de gestão de dados em saúde. No momento presente de agravo diário da pandemia COVID-19 em todo país, decisões ágeis e precisas são demandadas, evitando prejuízos econômicos e humanos para toda a nação. A gestão adequada dos dados de pacientes infectados com o coronavírus, além de possibilitar o seu reuso em novas pesquisas, facilitará a tomada de decisão baseada em evidências, evitando dessa forma, decisões equivocadas por parte dos gestores e profissionais de saúde, além de possibilitar a geração de dados mais confiáveis, contribuindo para a transparência, produção e disseminação do conhecimento científico.

Com o intuito de solucionar tais demandas no combate à pandemia, a rede VODAN BR participa desse projeto internacional na busca da melhoria da gestão, compartilhamento e reutilização dos dados desses pacientes em novas pesquisas, seguindo as tendências mundiais da Ciência Aberta.

A FIOCRUZ, a UNIRIO, a UFRJ e as instituições hospitalares envolvidas são responsáveis pelo nível de acesso que será dado aos dados brutos, os quais poderão ser disponibilizados abertamente, desde que resguardada a privacidade dos pacientes. Além dos dados autorizados pelas instituições, os dados agregados e metadados serão compartilhados, exceto aqueles sensíveis, que não fazem parte do escopo da rede. Os níveis de acesso ao conjunto de dados são três:

- Dados abertos - dados que forem considerados pela instituição coletora como possíveis de abertura (ex.: Alta/óbito, caso suspeito/confirmado, UTI/UCE, ventilação invasiva/não invasiva);

⁴ Disponível em: <https://www.vodan-totafrika.info/>. [Consult, 25 dez. 2020].

⁵ Disponível em: <https://vodanbr.github.io>. [Consult. 25 dez. 2020].

- Dados restritos às instituições parceiras - pode ser definido pela instituição coletora um acesso diferenciado para as outras unidades coletoras da Rede VODAN internacional e/ou VODAN BR;
- Dados fechados - todos os dados e conjuntos de dados restritos, somente disponíveis para a instituição coletora dos dados.

Os dados podem ser usados, pelo menos, para três finalidades:

- Pesquisa - estes dados poderão apoiar investigações atuais e futuras que permitam entender em maior profundidade a evolução dos casos de internação da COVID-19;
- Tratamento - estes dados poderão ser explorados e/ou agregados para apoiar o tratamento da doença e o enfrentamento da pandemia provocada pelo SARS-CoV-2;
- Gestão/monitoramento - estes dados poderão apoiar a gestão do hospital na programação de leitos e equipamentos, além de contribuir para o monitoramento detalhado da pandemia no Brasil.

Como parte deste esforço, a FIOCRUZ a UNIRIO e a UFRJ coordenam a rede VODAN BR na busca de capitanear esforços e recursos financeiros para a consolidação da rede brasileira. Vale ressaltar que o projeto da rede VODAN BR já foi aprovado pelo Comitê de Ética em Pesquisa no Brasil, através da Plataforma Brasil, tendo sua aprovação registrada no dia 16 de julho de 2020, pelo CEP/EPSJV-Fiocruz – CAAE: 32359220.4.0000.5241 – CONEP/CNS/MS.

3. Infraestrutura de dados

Esta rede tem por objetivo desenvolver uma infraestrutura para a coleta e gerenciamento de dados de prontuários desses hospitais e de outros que possam vir a se juntar à rede, de forma que possam ser analisados em dados vinculados ou formato CSV, ou ainda, por meio da interface de programação de aplicativos GraphDB, de acordo com um esquema de autorização de licenciamento previamente definido. Além disso, os metadados correspondentes deverão ser disponibilizados em um FAIR Data Point⁶, a fim de aumentar a recuperação desses dados em repositórios confiáveis, onde os conjuntos de dados desses pacientes serão hospedados, promovendo assim, o reaproveitamento e a sua reprodutibilidade em novas pesquisas.

Os dados serão coletados a partir do formulário de Pesquisa Clínica Uniforme, sugerido pela Organização Mundial da Saúde (OMS, 2020) nos hospitais parceiros preservando a privacidade das pessoas envolvidas. São vários os desafios a serem enfrentados no

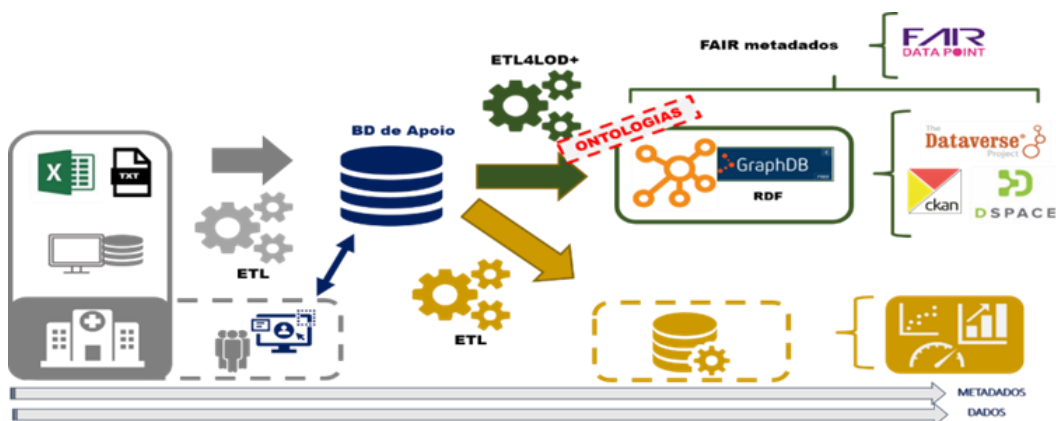
⁶ Disponível em: <https://www.go-fair.org/how-to-go-fair/fair-data-point/>. [Consult. 25 dez. 2020].

desenvolvimento desta infraestrutura genérica e flexível que dará suporte a todo o ciclo de vida dos dados e aos seus respectivos metadados.

Este é um projeto baseado em um modelo conceitual comum aos participantes da rede VODAN internacional, que garantirá que os dados sejam compreensíveis por humanos e por máquinas, em condições bem definidas. As atividades estão sendo realizadas em estreita cooperação entre as instituições envolvidas seguindo as seguintes metas de atuação:

- i.** uma estratégia de extração e coleta dos sistemas heterogêneos disponíveis nas unidades hospitalares;
- ii.** um banco de dados relacional de preparação, seguindo a estrutura de formulários da OMS para dados e metadados, enriquecido com uma ontologia de referência e outros vocabulários-padrão associados;
- iii.** um suporte baseado em workflow para o processo de FAIRification, usando a ferramenta ETL4LOD; bancos de dados de gráficos locais em cada hospital participante;
- iv.** um repositório confiável alinhado aos princípios FAIR; tratamento e alimentação dos dados coletados no banco de dados;
- v.** exportação e tratamento semântico dos dados, utilizando o modelo conceitual (ontologia de referência) da VODAN internacional;
- vi.** disponibilização dos dados agregados conforme a legislação vigente; um Web-app do formulário da OMS que permitirá a coleta de dados via celular.

Fig. 1 – Plataforma VODAN BR



Fonte: os próprios autores

Essa infraestrutura que está sendo elaborada no Brasil faz parte da rede global VODAN possibilitando que os dados gerados e os metadados associados participem de uma ampla gama de pesquisas em nível internacional. As ações em curso para a implantação da infraestrutura nacional perpassam os seguintes passos: criação do *Web-app* do Formulário

de Pesquisa Clínica Uniforme, que permitirá a coleta de dados complementares àqueles eventualmente já coletados pelos hospitais participantes e que serão desidentificados e extraídos para alimentação do sistema de tratamento e armazenamento dos dados; análise comparativa do formulário de coleta de dados do hospital participante com o Formulário de Pesquisa Clínica Uniforme adotado pela OMS; realização das customizações necessárias no sistema do hospital para a coleta/EXTRAÇÃO dos dados; capacitação dos profissionais envolvidos no preenchimento do formulário; teste do *Web-app* com o suporte dos hospitais participantes; projeto e implantação do sistema de tratamento/combinção/carga dos dados coletados e metadados de proveniência em banco de dados intermediário; coleta dos dados dos pacientes hospitalizados com suspeita ou confirmados de COVID-19; tratamento e alimentação dos dados coletados no banco de dados intermediário; exportação e tratamento semântico dos dados, utilizando o modelo conceitual (ontologia de referência) da VODAN internacional, para armazenamento em banco de dados baseado em tecnologias estabelecidas de representação do conhecimento em grafo; disponibilização dos dados agregados, conforme a legislação vigente.

4. Motivações do trabalho em rede para enfrentamento de pandemias

O mundo sofre um forte impacto e repercussão do surto do vírus na saúde, na economia e na sociedade como um todo, bem como pela falta de acesso aos dados de pacientes contaminados, devido às dificuldades legais, éticas e de disponibilidade de infraestruturas de apoio.

Nos dias de hoje, não existe uma maneira coordenada e estruturada de compartilhamento dos dados de pacientes entre países e regiões. O fato dos dados não serem FAIR (Localizável, Acessível, Interoperável e Reutilizável) prejudica significativamente seu uso, dificultando aos pesquisadores testarem hipóteses e encontrarem novas dependências ou novas correlações entre elementos de dados. Por essa razão, a adoção de uma infraestrutura de dados FAIR possibilitará a interoperabilidade entre os diferentes repositórios e o uso conjunto de seus dados. É importante ressaltar que o acesso aos dados coletados de pacientes contaminados deve respeitar as estruturas reguladoras e de governança e estar em conformidade com a Lei Geral de Proteção de Dados Pessoais (LGPD), Lei nº 13.709/2018 (BRASIL, 2018).

Esses dados poderão ser utilizados para diferentes fins. Além do uso na pesquisa, eles poderão contribuir para apoiar o monitoramento da pandemia e realizar diversos tipos de estudos, tais como: alternativas terapêuticas para a COVID-19; patogênese da história natural da doença; avaliação de atenção à saúde; prevenção e controle. Além disso, possibilitará a geração de elementos de divulgação da ciência por intermédio de publicações científicas e de outros veículos de divulgação.

5. Considerações finais

Esta pesquisa contém metas inovadoras que trazem vários benefícios tanto para a ciência, quanto para a sociedade, tais como, a criação de uma infraestrutura de gestão de dados de pesquisa "FAIR" voltada não apenas para a pandemia da COVID-19, mas também para futuras situações de crise ligadas a diferentes epidemias. Espera-se gerar elementos que

contribuam para a consolidação da rede VODAN Brasil e estreitar a participação da FIOCRUZ, UNIRIO e UFRJ em parcerias nacionais e internacionais. Esta é uma experiência significativa que comprova a importância da prática da ciência aberta voltada para a abertura e o compartilhamento de dados de pesquisa, contribuindo para maior transparência, interoperabilidade, velocidade de ampliação do conhecimento e para o avanço da ciência, com impacto célere para a sociedade.

Espera-se, ainda, apoiar o desenvolvimento da ontologia coordenada pela VODAN Internacional, a fim de permitir que os especialistas implementem o mapeamento semântico necessário e publiquem os dados em um formato RDF. Desta forma se possibilitará que os dados sejam legíveis por máquina, seguindo os princípios FAIR.

Destarte, alcançaremos o objetivo central deste projeto, que é criar infraestrutura federada de gestão de dados de forma a possibilitar que os dados dos pacientes contaminados com o vírus SARS-CoV-2 sejam passíveis de ser reutilizados em pesquisas, tratamento e monitoramento dos casos e apoiem a criação de alternativas terapêuticas para a COVID-19 e da história natural da doença; avaliação e atenção à saúde; além de prevenção e controle.

Entende-se ainda que a infraestrutura federada e toda a metodologia desenvolvida apoiarão a ciência a enfrentar de forma mais célere e eficaz novas pandemias e epidemias, possibilitando minimizar perdas imensuráveis para a sociedade, como a que estamos enfrentando no momento atual da história.

Referências bibliográficas

BRASIL. Leis, decretos, etc.

2018 *Lei nº 13.709, de 14 de agosto de 2018*. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). [Em linha]. 2018. [Consult. 8 abr. 2020]. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm.

FATTOR, Mauro

2020 Contro il Covid-19 anche gli specialisti dell'informatica : intervista a Giancarlo Guizzardi. *Alto Adige*. [Em linha]. 2020. [Consult. 9 abr. 2020]. Disponível em: <https://www.altoadige.it/cultura-e-spettacoli/contro-il-covid-19-anche-gli-specialisti-dell-informatica-1.2312141>.

MONS, Barend

2020 The VODAN IN: support of a FAIR-based infrastructure for COVID-19. *European Journal of Human Genetics*. [Em linha]. 28 (2020) 724-727. [Consult. 9 abr. 2020]. Disponível em: <https://doi.org/10.1038/s41431-020-0635-7>.

OMS

2020 *Global COVID-19: clinical platform: novel coronavirus (COVID-19): rapid version*. [Em linha]. Apr. 2020. [Consult. 20 dez 2020]. Disponível em: <https://apps.who.int/iris/handle/10665/331768>.

WILKINSON, Mark D. [et al.]

2016 The FAIR guiding principles for scientific data management and stewardship. *Nature Scientific Data*. [Em linha]. 3:160018. [Consult. 9 abr. 2020]. Disponível em: <https://www.nature.com/articles/sdata201618>.

Viviane Veiga | viviane.veiga@icict.fiocruz.br

Fundação Oswaldo Cruz (FIOCRUZ), Brasil

Maria Luiza Campos | mluiza.campos@gmail.com

Universidade Federal do Rio de Janeiro (UFRJ), Brasil

Carlos Roberto Lyra da Silva | profunirio@gmail.com

Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Brasil

Patricia Henning | henningpatricia@gmail.com

Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Brasil

João Moreira | j.luizrebelomoreira@utwente.nl

Universidade de Twente, Holanda