Inovação em saúde: A implementação de um data lake para o armazenamento, sistematização e disponibilização de dados em saúde no Brasil

Autoria

Daniel do Prado Pagotto - danielppagotto@gmail.com
Prog de Pós-Grad em Admin – PPGA / UnB - Universidade de Brasília

Denise Santos de Oliveira - deniseadm@hotmail.com
57 / UnB - Universidade de Brasília

Wanderson da Silva Marques - wdsmarques@gmail.com
Pesquisa / UFG - Universidade Federal de Goiás

Vicente da Rocha Soares Ferreira - vicenterocha@ufg.br
Programa de Pós-Graduação em Administração - PPGADM / UFG - Universidade Federal de Goiás
Programa de Pós-Graduação em Administração - PPGADM / UFG - Universidade Federal de Goiás

Vinicius Nunes de Azevedo - viniciuscoach@gmail.com
Programa de Pós-Graduação em Administração - PPGADM / UFG - Universidade Federal de Goiás

Cândido Vieira Borges Junior - candidoborges@gmail.com
Programa de Pós-Graduação em Administração - PPGADM / UFG - Universidade Federal de Goiás

Agradecimentos

Este artigo é parte dos resultados de projeto de pesquisa objeto de acordo de cooperação firmado entre a Universidade Federal de Goiás e a Secretaria de Gestão do Trabalho e da Educação na Saúde/Ministério da Saúde (TED 179/2019, Processo 25000206114201919/FNS)

Resumo

Este artigo tem como objetivo apresentar o problema relativo ao armazenamento, sistematização e disponibilização de dados em saúde no Brasil e uma solução inovadora, a implementação de um data lake com dados do setor de saúde. O data lake foi construído a partir de três etapas: (1) planejamento e priorização das bases de dados a serem importadas para o repositório; (2) extração, carregamento e tratamento dessas bases com o apoio das ferramentas Apache Airflow e Dremio; e (3) aplicação do uso. Os resultados evidenciam a capacidade da plataforma de armazenar um grande volume de dados (Big Data), bem como de propiciar uma navegação intuitiva, facilitando a compreensão e manuseio dos dados por analistas em saúde. Constata, ainda, que gestores públicos e pesquisadores reconhecem as contribuições da ferramenta para as suas decisões e a potencialidade desta para o desenvolvimento de outras soluções de inteligência para análise de dados da área de saúde. A solução apresentada visa contribuir para a gestão e o planejamento de políticas de saúde, permitindo o acesso, de modo rápido e amplo, a dados diversos, que suportam a tomada de decisões na área de saúde com mais agilidade, segurança e qualidade.

Inovação em saúde: A implementação de um *data lake* para o armazenamento, sistematização e disponibilização de dados em saúde no Brasil

Resumo: Este artigo tem como objetivo apresentar o problema relativo ao armazenamento, sistematização e disponibilização de dados em saúde no Brasil e uma solução inovadora, a implementação de um *data lake* com dados do setor de saúde. O *data lake* foi construído a partir de três etapas: (1) planejamento e priorização das bases de dados a serem importadas para o repositório; (2) extração, carregamento e tratamento dessas bases com o apoio das ferramentas *Apache Airflow* e *Dremio*; e (3) aplicação do uso. Os resultados evidenciam a capacidade da plataforma de armazenar um grande volume de dados (*Big Data)*, bem como de propiciar uma navegação intuitiva, facilitando a compreensão e manuseio dos dados por analistas em saúde. Constata, ainda, que gestores públicos e pesquisadores reconhecem as contribuições da ferramenta para as suas decisões e a potencialidade desta para o desenvolvimento de outras soluções de inteligência para análise de dados da área de saúde. A solução apresentada visa contribuir para a gestão e o planejamento de políticas de saúde, permitindo o acesso, de modo rápido e amplo, a dados diversos, que suportam a tomada de decisões na área de saúde com mais agilidade, segurança e qualidade.

Palavras-chave: Dados em saúde. *Data lake*. Ciência de dados. Gestão da saúde.

Introdução

A disponibilidade de dados confiáveis é crucial para que gestores de saúde possam tomar melhores decisões (Moutselos & Maglogiannis, 2020); assim, o tratamento, o gerenciamento e a análise adequada dos dados permitem a obtenção de informações fundamentais para a gestão dos serviços de saúde (Dash et al., 2019). Em meio aos avanços tecnológicos, foram identificados aumentos expressivos no volume de dados registrados nos sistemas de informação em saúde (Shortreed et al., 2019). Embora isso tenha representado muitas vantagens, tais como a vigilância em saúde, análise da distribuição de profissionais entre os territórios e a prestação de cuidados de saúde, também gerou grandes desafios aos gestores, entre os quais a dificuldade de gerenciamento desse grande volume de dados (Kroezen et al., 2018; Moutselos & Maglogiannis, 2020).

Geralmente, esses dados são disponibilizados em repositórios isolados (Gamache et al., 2018). No Brasil, por exemplo, as bases do Cadastro Nacional de Estabelecimentos de Saúde (CNES) dispõem de dados sobre estabelecimentos de saúde, suas infraestruturas e os profissionais vinculados a eles. Ademais, o Sistema de Informação de Agravos e Notificações (SINAN) e o Sistema de Informações sobre Mortalidade (SIM) fornecem dados de natureza epidemiológica, enquanto o Sistema de Informações Hospitalares (SIH), o e-SUS Atenção Primária (e-SUS APS) e o Sistema Nacional de Regulação (SISReg) são utilizados para o gerenciamento dos serviços de assistência à saúde. Por sua vez, as bases de projeções populacionais do Ministério da Saúde e do Instituto Brasileiro de Geografia e Estatística (IBGE) dispõem de dados sobre a demografia relativa ao sistema de saúde. Assim, observa-se que o problema a ser enfrentado pelo gestor, em uma análise macro, é encontrar, tratar, sistematizar, sintetizar e relacionar esse grande volume de dados, advindo de diferentes fontes (Kroezen et al., 2018) e em diferentes formatos.

Para solucionar esse tipo de problema é necessário reunir os dados em um único local. A integração dos dados de maneira estruturada permitiria aos gestores, planejadores e pesquisadores da saúde acessar, de modo rápido e fácil, um amplo conjunto de dados, bem como melhor visualizá-los, compreendê-los, e realizar análises a partir do correlacionamento entre eles (Dash et al., 2019; Gamache et al., 2018). Nesse sentido, o objetivo do presente artigo

é apresentar e descrever todo o processo de organização e implementação de um *data lake* da área da saúde no Brasil.

O *data lake* é uma estrutura de armazenamento de dados que reúne informações de diversas fontes e aplica modelos analíticos para fornecer uma nova abordagem de interpretação, gerenciamento e análise aos usuários (Maini et al., 2018). A partir dele, gestores podem obter *insights* que permitam maior eficiência na gestão do sistema de saúde, em suas mais amplas ou específicas e complexas nuances.

A relevância deste estudo se sustenta na apresentação de todo o percurso de desenvolvimento do *data lake* assim como na descrição de todo o fluxo utilizado para a implementação de uma infraestrutura de dados estratégica e valiosa. Outra vantagem é a possibilidade de que os mesmos procedimentos podem ser adotados em contextos diferentes daqueles para os quais foi estruturado o *data lake* apresentado neste estudo. Para fins práticos, o resultado deste trabalho é um avanço, uma vez que objetiva consolidar em uma fonte única e de fácil acesso a dados públicos outrora pulverizados em múltiplas fontes, que não possibilitam análises conexas nem a exploração completa de todo o seu potencial.

Sistemas brasileiros de informação e a consolidação dos dados

Nas últimas décadas, um conjunto de sistemas de informação foi implantado ou expandido no Brasil, o que ampliou a disponibilidade de informações para a gestão em saúde (Batista et al., 2019). Os gestores e pesquisadores dispõem de uma rede de informações composta por dados demográficos, epidemiológicos, de monitoramento de programas de saúde, de quantidade de profissionais disponíveis, entre outros (Correia et al., 2014).

Muitos sistemas de informação que abordam dados em saúde de modo finalístico são administrados pelo Ministério da Saúde, tais como o e-SUS Notifica, o Sistema de Informações sobre Mortalidade (SIM), o Sistema de Informações Hospitalares do SUS (SIH/SUS), o Sistema de Informações sobre Nascidos Vivos (Sinasc), o Sistema de Informação de Agravos de Notificação (Sinan), o Sistema de Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico (Vigitel), entre tantos outros. Tais sistemas tendem a retratar as condições de saúde da população e fornecer base aos gestores e pesquisadores para o levantamento da demanda assistencial, bem como amplo acesso a pesquisadores, profissionais de saúde e sociedade em geral (Batista et al., 2019; Correia et al., 2014). A Tabela 1 apresenta a finalidade de cada um desses sistemas.

Tabela 1 - Finalidades dos sistemas de saúde

Bases de dados	**Finalidades**
e-SUS Notifica	Dispõe de notificações de síndrome gripal suspeita e notificação de covid-19.
SIM	Expõe dados sobre mortalidade.
SIA	Contempla dados sobre a prestação de serviços ambulatoriais
SIH/SUS	Apresenta informações de internações hospitalares da rede própria ou conveniada ao SUS.
Sinasc	Expõe informações sobre nascimentos.
Sinan	Apresenta dados de notificação compulsória de doenças e agravos.

Bases de dados	Finalidades
Vigitel	Apresenta dados acerca de doenças crônicas não transmissíveis, tais como diabetes, câncer e doenças cardiovasculares.
CNES	O Cadastro Nacional de Estabelecimentos de saúde é uma base que contempla dados sobre estabelecimento de saúde, profissionais atuantes, infraestrutura desses locais, dentre outros aspectos.

Fonte: Batista et al. (2019) e Brasil (2021).

Além desses sistemas, existem outros que, apesar de não tratarem de saúde de modo finalístico ou direto, são importantes para o seu gerenciamento. O Instituto Brasileiro de Geografia e Estatística (IBGE), por exemplo, apresenta projeções populacionais e características dos municípios que podem auxiliar no processo de gestão. Por sua vez, o Censo da Educação Superior, realizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), permite identificar a formação de recursos humanos em saúde (Machado & Ximenes Neto, 2018).

Assim, as iniciativas de sistematização de dados são diversas, mas esses dados são apresentados de modo fragmentado e, muitas vezes, são de difícil compreensão e acesso para o gestor (Correia et al., 2014). Sabe-se, contudo, que a forma como os dados são apresentados influencia a compreensão e a formação de *insights* (Fernandes et al., 2020). Nesse sentido, o uso de ferramentas de monitoramento das informações, tais como *dashboards*, torna-se relevante. Os *dashboards* são painéis que exibem informações diversas a partir de gráficos claros, objetivos, completos e passíveis de customizações, por meio da utilização de filtros, que podem ser incorporados, permitindo a visualização de um conjunto de dados de forma simples, o que proporciona auxílio à tomada de decisões (Knaflic, 2015). Promovem, ainda, a apresentação de tendências e correlações entre diferentes variáveis, simplificando a análise e potencializando a construção ou reconstrução de estratégias de modo preciso, customizado e ágil (Nijkamp & Kourtit, 2022).

Embora os *dashboards* auxiliem na visualização de dados, a sua apresentação de modo fragmentado, isto é, a partir de diferentes sítios, pode produzir dificuldades na sua devida identificação pelo gestor, bem como desfavorecer a devida análise das informações. Dessa forma, o agrupamento de informações de fontes diversas em um único sítio pode auxiliar os gestores e pesquisadores de saúde na sua identificação, bem como na obtenção de uma visão mais ampla dos dados (Dash *et al*., 2019; Gamache *et al*., 2018). Algumas iniciativas para consolidação de dados em sítio único são evidenciadas no Brasil, tais como os esforços da organização não governamental Base dos Dados e da Rede Nacional de Dados em Saúde (RNDS). A primeira é uma iniciativa iniciada em 2019 que visa facilitar o acesso a dados públicos, de modo geral, algumas vezes já tratados, por meio de bibliotecas das linguagens de programação *R* e *Python* ou por intermédio da ferramenta *BigQuery* (Base dos Dados, 2022). A segunda, RNDS, é um programa instituído pelo Ministério da Saúde, desde maio de 2020, para integrar dados de diversos atores de todo o País em uma plataforma única nacional (interoperabilidade de sistemas de saúde), incluindo dados de laboratórios, centros de pesquisa e desenvolvimento, farmácias, profissionais de saúde, atendimento de urgência e emergência, entre outros, permitindo o compartilhamento de informações da assistência à saúde nos setores público e privado (Coutinho *et al*., 2021; Brasil, 202X).

Embora a RNDS seja direcionada à gestão em saúde, a rede ainda não foi completamente implementada. No cenário de pandemia da covid-19, um projeto piloto foi desenvolvido para auxiliar a situação emergencial de saúde pública (Brasil, n.d.), contudo, a consolidação de dados diversos para a tomada de decisão em todos os níveis de atenção

permanece deficiente. Ainda, o acesso à RNDS é limitado a gestores que adquiram certificado digital, possuam conta no gov.br e solicitem acesso aos serviços da rede (Brasil, 202X).

Logo, a implementação de um *data lake* da área da saúde pode contribuir para suprir as limitações identificadas, visto que este armazena, em tempo real, dados de fontes diversas em seu formato bruto. Cada fonte está associada a um identificador único, e a construção de um *data lake* demanda um repositório de metadados capaz de registrar um alto nível de informações sobre entidades de dados. Nele, é possível apresentar dados a partir de diversos *dashboards* para auxiliar a sua compreensão. Além do registro de dados, é possível aplicar modelos analíticos para associação entre os dados de uma mesma base ou de bases diferentes (Maini *et al*., 2018).

Método

A construção da solução se deu ao longo do ano de 2021 e contou com três etapas, conforme ilustrado na Figura 1.

Figura 1 - Etapas para construção do *data lake*

A primeira etapa foi destinada ao planejamento do projeto bem como à priorização das bases de dados públicas a serem importadas para o *data lake*. Esta última atividade foi realizada por meio da consulta a três pesquisadores - um mestre e dois doutores - com experiência em análise de dados em saúde e responsáveis por liderar três projetos diferentes acerca de gestão do trabalho e da educação na saúde. Tais especialistas foram acessados no intuito de listar bases de dados públicas úteis para projetos nos quais estavam envolvidos. Como resultado, foi formada uma lista de 27 bases oriundas de diferentes fontes, como Datasus, Instituto Brasileiro de Geografia e Estatística (IBGE), Ministério da Educação (Mec), Ministério do Trabalho e Emprego (MTE), Agência Nacional de Saúde (ANS), entre outras. Identificada a relação de bases, os pesquisadores indicaram quais seriam aplicadas nos respectivos projetos, o que permitiu estabelecer-se uma ordem de priorização para inclusão no *data lake*. Aquelas, por exemplo, que seriam utilizadas nas três iniciativas, seriam priorizadas na etapa 2.

A segunda etapa correspondeu à condução de ciclos de extração, tratamento e carregamento das bases priorizadas. Portanto, para cada base, os dados foram extraídos de seus sítios originais, tratados de modo a padronizarem-se os diferentes formatos de dados em arquivos do tipo *parquet*, formato este com maior grau de compressão de volume se comparado a outros tipos, tais como *comma separated values* (csv). Além disso, para as bases com atualização periódica, foram construídos *scripts* de automatização da captura, tratamento e carregamento dos dados na plataforma de orquestração de fluxo de dados *Apache Airflow*. Finalmente, os dados armazenados foram disponibilizados em código livre *Dremio*.

A Figura 2 ilustra a arquitetura do *data lake*, com as tecnologias que mantêm o funcionamento da ferramenta, conforme já explicado. Os dados foram acessados a partir de múltiplas fontes, sob diferentes formatos, e carregados em estado bruto (*raw data*) em uma

primeira camada de dados. Uma segunda camada - *analytics layer* - foi desenvolvida com vista a permitir que o usuário realize tratamento e consultas de dados e as salve, o que garante a transparência na construção de consultas e seu reaproveitamento. Por fim, ainda existem uma camada de catálogo dos dados que descreve os dados presentes no *data lake* e uma camada de segurança (autenticação, controle de usuários, registro de logs, etc).

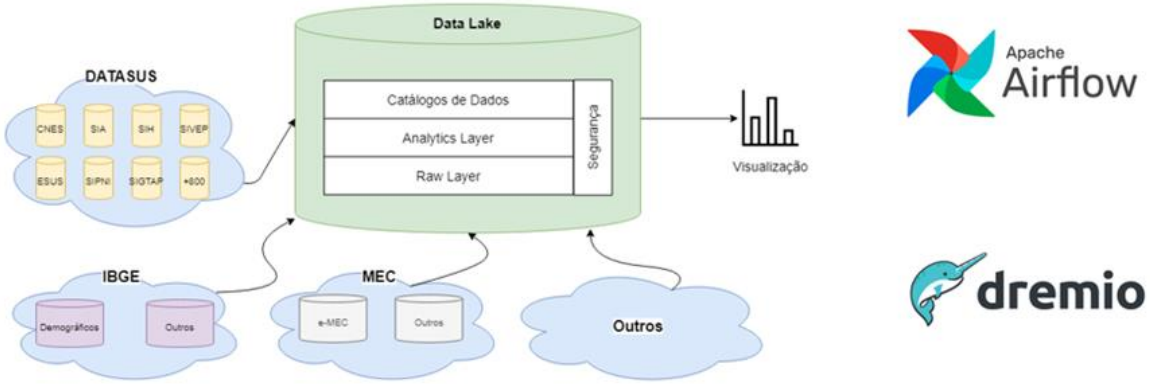

Figura 2 - Estrutura de funcionamento do *Data Lake*

A última etapa consistiu na disponibilização da ferramenta para usuários dos três projetos mencionados para o teste e relato de erros para, em sequência, consumo. Nessa etapa, também foi realizado um treinamento, que contou com a participação de 10 pesquisadores com experiência em análise de dados e em projetos sobre gestão da educação e do trabalho na saúde, no intuito de apresentar-se o funcionamento da ferramenta, bem como as formas de acesso via linguagens de programação R e Python e a interface do *Dremio*.

Uma vez difundido o uso do *data lake*, a plataforma passou a ser utilizada como referência de acesso a banco de dados, permitindo um uso mais eficiente tanto para cálculos rotineiros quanto para a construção de plataformas de *Business Intelligence*, uma vez que o *Dremio* possui a funcionalidade de integração direta a softwares desta natureza, como o *Microsoft Powerbi* e o *Tableau*.

Resultados

Ao final do primeiro ano do projeto, o *data lake* já contava com mais de 250 gigabytes de dados, 23 bases de dados, subdivididas em 239 tabelas. A Figura 3 ilustra a interface do *Dremio* com a relação de bases que podem ser acessadas. Ao todo, três projetos e 22 pesquisadores que atuam em iniciativas sobre gestão do trabalho e educação na saúde consomem dados diretamente do *data lake* para o projeto.

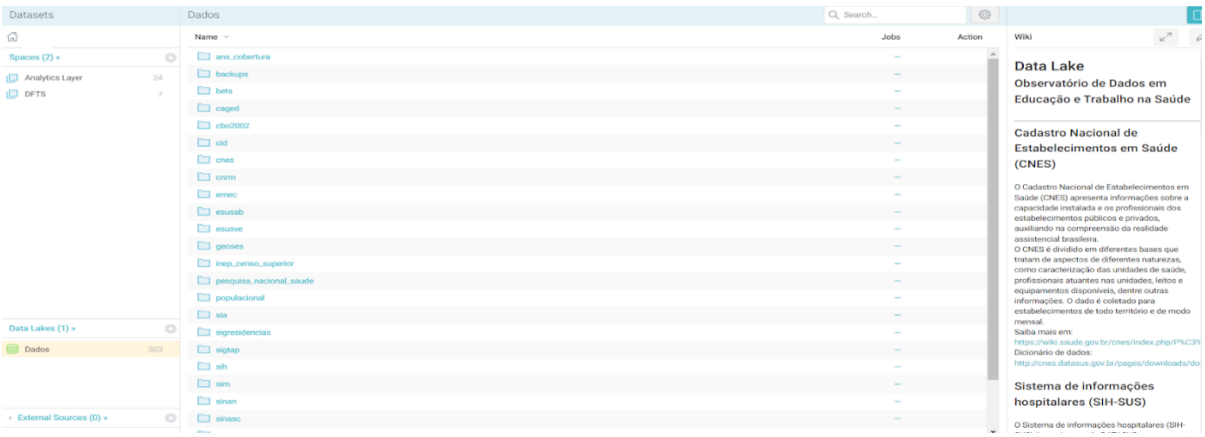

Figura 3 - Interface de lista de dados

O *data lake* trouxe vantagens, como a padronização na forma de acessar dados e a eficiência decorrente de tratamentos prévios realizados por meio de linguagem SQL (*Standard Query Language)*, utilizando-se a interface do *Dremio* (Figura 4) e linguagens de programação. O depoimento abaixo registra o comentário de um dos seus usuários.

> O *data lake* trouxe muitos benefícios para quem trabalha e pesquisa sobre os temas relacionados às suas bases de dados. Primeiro, é uma fonte única de dados. Existem bases que podem ser acessadas via protocolo de transferência de arquivo (ftp, do inglês *file transfer protocol*) do Datasus, por meio da ferramenta para tabulação de dados Tabnet, ou por meio de um site do Ministério da Saúde. Por alguma questão de tratamento, estas fontes podem apresentar alguns números - ainda que pequenos - diferentes. Portanto, uniformizar a fonte de acesso foi o primeiro ganho. Além disso, o *data lake* padronizou todos os arquivos em um formato único que pode ser acessado via comandos SQL. Assim, o pesquisador não precisa baixar diversos arquivos em diferentes formatos, como csv ou planilhas eletrônicas. Basta acessar via SQL e extrair aquelas variáveis/atributos que necessitará para suas análises. Decorrente desta última vantagem está o uso do SQL que permite um tratamento prévio dos dados de um modo mais eficiente do que por meio do acesso às bases brutas e o tratamento via bibliotecas *tidyverse* ou *pandas* das linguagens R e Python, respectivamente.

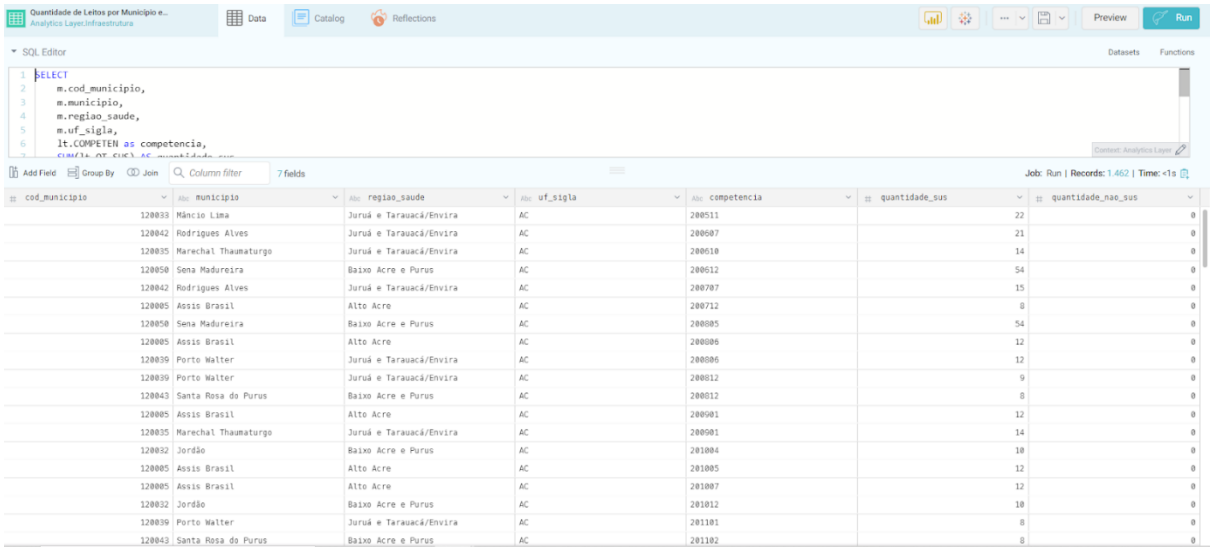

Figura 4 - Interface de acesso a dados via Dremio com editor SQL

Os resultados do projeto foram apresentados em duas ocasiões diferentes para gestores de nível estratégico, o reitor da universidade a qual o projeto está vinculado e um gestor de nível tático da administração pública federal. Diante da potencialidade e importância reforçada nesses encontros, foi proposta a continuidade da iniciativa e a incorporação de outras bases de dados, difundindo-se informações no formato de *dashboards*.

Diante do exposto e considerando que um dos propósitos do *data lake* foi centralizar dados a fim de aplicá-los em sistemas, análises estatísticas e painéis de visualização de dados, a equipe do projeto avançou na criação de *dashboards* sobre dados em saúde. Tais pressupostos fomentaram a criação da Plataforma de Inteligência em Gestão da Educação do Trabalho em Saúde (Figura 5 e 6).

Figura 5 - Protótipo de Plataforma

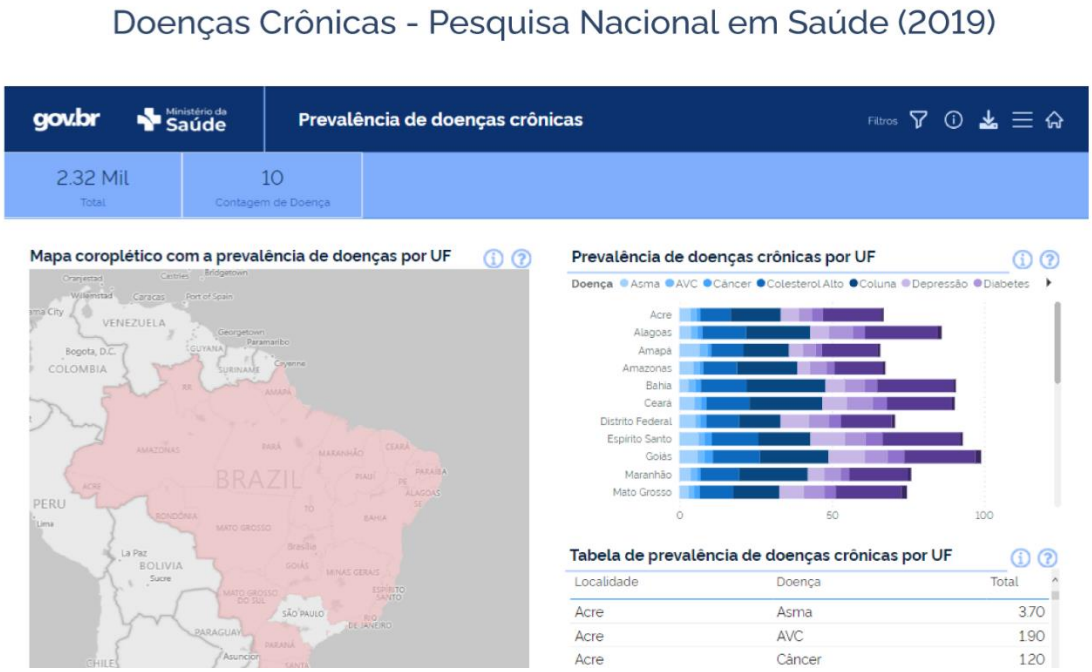

Figura 6 - Protótipo de painel da plataforma

Discussão

O uso de dados secundários oriundos de diferentes bases de dados é amplamente utilizado na área de saúde, seja na formulação de políticas de saúde, na gestão dos serviços ou no desenvolvimento de pesquisas (Ferreira et al., 2020). No entanto, por vezes, tais dados são apresentados de modo fragmentado (Correia et al., 2014; Coelho Neto; Chioro, 2021). Superar a fragmentação de dados na área de saúde é um desafio não só do Brasil (Pinto et al., 2018).

Uma das formas mais utilizadas de acesso aos dados é por meio do Tabnet/Tabwin. No entanto, por mais que tenha sido um avanço para a publicidade de dados em saúde, é uma ferramenta desenvolvida há cerca de trinta anos (Coelho Neto; Chioro, 2021) e que possui limitações importantes, incluindo a impossibilidade de conjugação de bases ou aplicação de

agregações não compatíveis às funcionalidades da plataforma. Nesse sentido, algumas experiências são importantes para a disseminação de microdados (dados brutos, com menor granularidade), como o desenvolvimento do pacote da linguagem R microdatasus (Saldanha; Bastos & Barcellos, 2020).

Como mais uma ferramenta de acesso a múltiplas fontes de microdados, o *data lake* apresentado e descrito neste estudo permite que variados dados sejam acessados a partir de uma única fonte, permitindo, assim, o aumento da confiabilidade e usabilidade das plataformas de acesso aos dados. Além disso, a inclusão de camada analítica com a consolidação de consultas e análises, constitui-se um passo fundamental para garantir maior transparência por meio de um paradigma de dados e materiais abertos (Miguel et al., 2014), o que permite maior facilidade para o trabalho de gestores e pesquisadores da área de saúde.

Diante da ampla variedade de dados disponíveis, é importante que se busquem ferramentas que sintetizem e sistematizem tal volume, garantindo-lhes acessibilidades e navegações intuitivas, condições fundamentais para a efetiva garantia de transparência proposta com a disponibilização do acesso aos dados. Nesse sentido, a visualização de dados por meio de painéis interativos pode ser uma estratégia que facilita a absorção de informações (Dash et al., 2018), contribuindo, em última instância, para uma melhor tomada de decisão pelos gestores (Ifhitkar et al., 2019), garantindo eficiência, eficácia e efetividade ao trabalho de gestores e pesquisadores.

Conclusão

Há um avanço constante sobre a disponibilidade de dados públicos decorrentes de sistemas governamentais. Apesar de avanços decorrentes de esforços públicos e do terceiro setor, iniciativas de difusão de dados e informações geradas a partir deles devem ser incentivadas. Nesse sentido, o presente artigo descreve o processo de construção de um artefato tecnológico que contribui para a difusão de dados, a transparência de análises, o apoio gerencial e apoio à pesquisa.

Além dos benefícios listados acima inerentes ao produto, o artigo descreve todo o processo de construção da plataforma, apresentando ferramentas e processos sistematizados adotados, o que permite o desenvolvimento de estruturas semelhantes para outras finalidades.

Os dados secundários públicos são amplamente utilizados em pesquisas e na tomada de decisão. Todavia, isso não exime os gestores de sistemas de informação e bases de dados de buscarem melhorias constantes nesse universo. Assim, métodos que assegurem uma constante melhoria na qualidade dos dados podem ser incorporados, como aprimoramento do preenchimento na fonte, tratamento e uso de instrumentos de coleta que assegurem a captura de dados relevantes.

Além disso, mais esforços devem ser empregados para a disponibilização de microdados de certas bases, como dos componentes do e-SUS AB. Tais questões vão além do escopo dessa ferramenta, mas, por tangenciar o principal insumo deste trabalho - os dados - não podem deixar de ser alertadas. Espera-se que essas medidas, em conjunto com a maior abertura no acesso a dados, permitam ampliar as pesquisas no campo da saúde, bem como aprimorar o gerenciamento de organizações e serviços de saúde, além de grande apoio ao trabalho de pesquisadores da área de saúde pública.

Referências

Base dos dados (2022). Quem somos. Recuperado de: <https://basedosdados.org/quem-somos> em 09 de abr de 2022.

Batista, A. G., Santana, V. S., & Ferrite, S. (2019). Registro de dados sobre acidentes de trabalho fatais em sistemas de informação no Brasil. *Ciência & Saúde Coletiva*, *24*, 693-704. https://doi.org/10.1590/1413-81232018243.35132016

Brasil. Ministério da Saúde (n.d.). Rede Nacional de Dados em Saúde - RNDS. Recuperado de <https://www.gov.br/saude/pt-br/assuntos/rnds> em 07 de abr de 2022.

Brasil. Ministério da Saúde (2021). Sistemas de informação em saúde. Recuperado de: <https://www.gov.br/saude/pt-br/composicao/svs/vigilancia-de-doencas-cronicas-nao-transmissiveis/sistemas-de-informacao-em-saude> em 07 de abr de 2022.

Coelho Neto, G. C., & Chioro, A. (2021). Afinal, quantos Sistemas de Informação em Saúde de base nacional existem no Brasil?. *Cadernos de Saúde Pública*, *37*, e00182119. https://doi.org/10.1590/0102-311X00182119

Correia, L. O. D. S., Padilha, B. M., & Vasconcelos, S. M. L. (2014). Métodos para avaliar a completitude dos dados dos sistemas de informação em saúde do Brasil: uma revisão sistemática. *Ciência & Saúde Coletiva*, *19*, 4467-4478. https://doi.org/10.1590/1413-812320141911.02822013

Coutinho, L. R., Neves, H. P. O. D. E., & Lopes, L. C. (2021). Abordagens sobre computação na nuvem: uma breve revisão sobre segurança e privacidade aplicada a e-saúde no contexto do Programa Conecte SUS e Rede Nacional de Dados em Saúde (RNDS). *Brazilian Journal of Development*, *7*(4), 35152-35170. https://doi.org/10.34117/bjdv7n4-127

Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*, *6*(1), 1-25. https://doi.org/10.1186/s40537-019-0217-0

Ferreira, J. E. D. S. M., de Oliveira, L. R., Marques, W. S., de Lima, T. S., da Silva Barbosa, E., Castro, R. R., & Guimarães, J. M. X. (2020). Sistemas de Informação em Saúde no apoio à gestão da Atenção Primária à Saúde: revisão integrativa. *Revista Eletrônica de Comunicação, Informação e Inovação em Saúde*, *14*(4). https://doi.org/10.29397/reciis.v14i4.1923

Fernandes, A. M. R., Henrique, A. S., Liebel, G., Dazzi, R. L. S., & Mezadri, T. (2020). A Relevância dos Dashboards para a Gestão da Saúde na Pandemia Causada pelo COVID-19. *Brazilian Journal of Development*, *6*(6), 39263-39274. https://doi.org/10.34117/bjdv6n6-462

Gamache, R., Kharrazi, H., & Weiner, J. P. (2018). Public and population health informatics: the bridging of big data to benefit communities. *Yearbook of medical informatics*, *27*(01), 199-206. https://doi.org/10.1055/s-0038-1667081

Iftikhar, A., Bond, R., McGilligan, V., J Leslie, S., Rjoob, K., Knoery, C., & Peace, A. (2019, September). Role of dashboards in improving decision making in healthcare: Review of the literature. In *Proceedings of the 31st European Conference on Cognitive Ergonomics* (pp. 215-219).

Knaflic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons.

Kroezen, M., Van Hoegaerden, M., & Batenburg, R. (2018). The Joint Action on Health Workforce Planning and Forecasting: Results of a European programme to improve health workforce policies. *Health Policy*, *122*(2), 87-93. https://doi.org/10.1016/j.healthpol.2017.12.002

Machado, M. H., & Ximenes Neto, F. R. G. (2018). Gestão da Educação e do Trabalho em Saúde no SUS: trinta anos de avanços e desafios. *Ciência & Saúde Coletiva*, *23*, 1971-1979. https://doi.org/10.1590/1413-81232018236.06682018

Maini, E., Venkateswarlu, B., & Gupta, A. (2018). *Data Lake-An Optimum Solution for Storage and Analytics of Big Data in Cardiovascular Disease Prediction System*. International Journal of Computational Engineering & Management, 21(6).

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., ... & Van der Laan, M. (2014). Promoting transparency in social science research. *Science*, *343*(6166), 30-31. https://doi.org/10.1126/science.1245317

Moutselos, K., & Maglogiannis, I. (2020). Evidence-based Public Health Policy Models Development and Evaluation using Big Data Analytics and Web Technologies. *Medical Archives*, *74*(1), 47. https://doi.org/10.5455/medarh.2020.74.47-53

Nijkamp, P., & Kourtit, K. (2022). Place-Specific Corona Dashboards for Health Policy: Design and Application of a 'Dutchboard'. *Sustainability*, *14*(2), 836. https://doi.org/10.3390/su14020836

Pinto, L. F., Freitas, M. P. S. D., & Figueiredo, A. W. S. A. D. (2018). Sistemas Nacionais de Informação e levantamentos populacionais: algumas contribuições do Ministério da Saúde e do IBGE para a análise das capitais brasileiras nos últimos 30 anos. *Ciência & Saúde Coletiva*, *23*, 1859-1870. https://doi.org/10.1590/1413-81232018236.05072018

Saldanha, R. D. F., Bastos, R. R., & Barcellos, C. (2019). Microdatasus: pacote para download e pré-processamento de microdados do Departamento de Informática do SUS (DATASUS). *Cadernos de Saúde Pública*, *35*(9). https://doi.org/10.1590/0102-311X00032419

Shortreed, S. M., Cook, A. J., Coley, R. Y., Bobb, J. F., & Nelson, J. C. (2019). Challenges and opportunities for using big health care data to advance medical science and public health. *American journal of epidemiology*, *188*(5), 851-861. https://doi.org/10.1093/aje/kwy292