



University of
Texas Libraries



e-revist@s

Sumários.org



Centro Unversitário Santo Agostinho

revistafsa

www4.fsnet.com.br/revista

Rev. FSA, Teresina, v. 21, n. 12, art. 5, p. 92-111, Dez. 2024

ISSN Impresso: 1806-6356 ISSN Eletrônico: 2317-2983

<http://dx.doi.org/10.12819/2024.21.12.5>

DOAJ DIRECTORY OF
OPEN ACCESS
JOURNALS

WZB
Wissenschaftszentrum Berlin
für Sozialforschung



MIAR



Dataficação e Desinformação: Datademia na Cobertura Sobre Covid-19 em Redes Digitais

Datafication and Disinformation: Datademia in Coverage of Covid-19 on Digital Networks

Adriana Alves Rodrigues

Doutora em Ciência da Informação pela Universidade Federal da Paraíba

Pesquisadora de Pós-Doutorado pela Universidade Federal da Paraíba

E-mail: profeadrianaalves@gmail.com

Endereço: Adriana Alves Rodrigues

Universidade Federal da Paraíba, Programa de Pós-Graduação em Ciência da Informação - UFPB.
Universidade Federal da Paraíba - Campus I - Lot.
Cidade Universitária 58051900 - João Pessoa, PB - Brasil

Editor-Chefe: Dr. Tonny Kerley de Alencar Rodrigues

Artigo recebido em 18/10/2024. Última versão recebida em 30/10/2024. Aprovado em 31/10/2024.

Avaliado pelo sistema Triple Review: a) Desk Review pelo Editor-Chefe; e b) Double Blind Review (avaliação cega por dois avaliadores da área).

Revisão: Gramatical, Normativa e de Formatação



RESUMO

O artigo busca teorizar e explorar sobre o surgimento do fenômeno da *datademia*, a partir do entrelaçamento entre os conceitos de dataficação e *big data*, no que se refere à cobertura jornalística da Covid-19. Para tanto, realizou-se uma pesquisa exploratória-descritiva e bibliográfica teórico-conceitual, analisando visualizações de dados dos meios digitais. Como resultados, debate-se: a) os processos de dataficação e *big data* como desencadeadores do conceito de *datademia*; b) a dimensão dos dados no contexto pandêmico a partir das visualizações de dados e c) a tipologia dos dados advindos enquanto proeminentes na divulgação das informações do coronavírus. Considera-se que a tríade *big data*-dataficação-datademia pode configurar em práticas viabilizadoras diante do dilúvio de dados, ao mesmo tempo em que problemáticas de gestão e gerenciamento de dados neste processo podem inviabilizar a compreensão dos dados pandêmicos.

Palavras-chave: Big Data. Dataficação. Visualização de Dados.

ABSTRACT

The article seeks to theorize and explore the emergence of the datademic phenomenon based on the intertwining between the concepts of datafication and big data with regard to journalistic coverage of Covid-19. To this end, an exploratory-descriptive and theoretical-conceptual bibliographical research was carried out analyzing data visualizations from digital media. As results, it discusses: a) the processes of datafication and big data as triggers of the concept of datademic; b) the dimension of data in the pandemic context based on data visualizations and c) the typology of data arising as prominent in the dissemination of coronavirus information. It is considered that the big data-datafication-datademic triad can constitute viable practices in the face of the deluge of data, at the same time that data management and management issues in this process can make understanding pandemic data unfeasible.

Keywords: Big Data. Datafication. Data Visualization.

1 INTRODUÇÃO

O surto global de COVID-19 teve um forte impacto na vida econômica e social em vários países. No estágio inicial da prevenção da epidemia, a falta de transparência dos dados governamentais e o entendimento insuficiente sobre o vírus e vários fatores frequentes trouxeram desafios para a prevenção e controle da pandemia. De acordo com a Organização Mundial de Saúde (OMS), foi na China onde ocorreu o primeiro caso confirmado de SARS-CoV-2, em 31 de dezembro de 2019, porém a pandemia foi declarada em 11 de março de 2020. A partir dessa data, todos os países do mundo foram obrigados a fazer isolamento social, quarentena, *lockdown*, uso de máscaras e álcool em gel como medidas de combate contra o novo coronavírus.

Ao mesmo tempo, observou-se um frenesi das informações divulgadas nas redes digitais na tentativa de entender o contexto atual. A cobertura jornalística da pandemia propiciou novos modos de fazer reportagens, no que se refere à representação da realidade e ao povo brasileiro, com foco no engajamento dele (BECKER, 2021). Na atual era do *Big Data*, o grande volume de dados gerados e disseminados nas redes foi facilmente espalhado como forma de compreensão do novo coronavírus. Nesse sentido, podemos observar os dados da Covid 19 através da perspectiva de dataficação (*datafication*), isto é, “colocá-los num formato quantificado de modo que possam ser tabulados e analisados” (MAYER-SCHONBERGER; CUKIER 2013, p. 54) da cobertura midiática ao tratar e divulgar os dados epidemiológicos. Dessa forma, a abundância e, sobretudo, a disseminação pelos jornais de modo intensivo estimulou uma espécie de sobrecarga dos dados no contexto, no qual denominamos de *datademia*, expressivo aumento de grandes volumes de dados sobre o vírus com a tendência de se multiplicar exponencialmente sobre um determinado evento particular. Nessa direção, esse fenômeno é intensificado e amplificado pelas redes digitais com alto poder de divulgação e espalhamento. O surgimento da *datademia* é um processo vinculado historicamente pelo surgimento da dataficação, *Big Data*, ambos desdobramentos das tecnologias digitais da contemporaneidade.

No contexto da pandemia, é pertinente ressaltar também o conceito de infodemia, caracterizado pelo excesso de informações difíceis de lidar. Na década de 1940, tivemos uma explosão informacional. A partir de 2000, emergiu a ideia de infodemia, termo criado por David J. Rothkopf. Durante a pandemia da Covid-19, o fenômeno foi visto como infodemia ou como epidemia de informações. Da mesma forma que excesso de informações gerais pode

trazer consequências, a *datademia* em si pode trazer efeitos similares e, ao mesmo tempo, potenciais para lidar com dados complexos por meio da visualização de dados e gráficos.

Uma questão-chave na prevenção e controle de epidemias científicas é como usar o potencial da tecnologia da informação para quebrar ilhas de dados, rastrear a transmissão do vírus e obter prevenção e controle precisos. Durante esse período, a visualização de dados foi amplamente explorada pela imprensa mundial, para apresentar os dados estatísticos e difusão de dados sobre a pandemia (como casos confirmados, óbitos e recuperados, atualização contínua) a exemplo do mapa interativo¹ feito pelo Centro de Recursos de Coronavírus, da Universidade Johns Hopkins. Nessa perspectiva, diversos veículos de comunicação em nível mundial produziram mapas similares e/ou coletaram dados divulgados pela Universidade americana para incremento do material noticioso e foram os principais suportes utilizados pela imprensa mundial.

A investigação tem como caráter a abordagem qualitativa e faz uso do estudo exploratório-descritivo a partir de Gil (2008), esse método deve proporcionar maior proximidade com o tema estudado e oferece análise de exemplos que estimulem a compreensão do objeto analisado.

Para tanto, analisa as visualizações de dados sobre a pandemia em veículos de mídia no período de fevereiro a março de 2020, momento em que se iniciou a deflagração do vírus em escala planetária e, como consequência, uso mais sistemático das visualizações. Como estratégia de coleta de dados, optamos por explorar os veículos de comunicação que têm uso sistemático de *dataviz*² em suas reportagens, quais sejam: *The New York Times*, *The Washington Post*, O Globo, G1, Jornal Nexo, *El Mundo* e *La Nación*. O critério de seleção do *corpus empírico* levou em consideração os aspectos de originalidade, representatividade e diversidade, considerando que podem auxiliar a parte da seleção (Machado; Palacios, 2007). Para interpretar os dados, utilizamos o procedimento de Análise Descritiva, que consiste em descrever as principais tendências de um fenômeno a partir dos dados existentes e em observar situações que levam a novos fatos. Desse modo, tensionamos os resultados do período inicial da pandemia com uso intensivo de dados para compreender a emergência da *datademia* da produção, principalmente de gráficos e visualizações de dados. Em pesquisa posterior pretendemos explorar o pico desse uso de dados entre 2020 e 2022.

¹ Disponível em: <https://coronavirus.jhu.edu/map.html> Acesso em: 3 jul de 2022.

² Do inglês. Abreviação de visualização de dados.

2 REFERENCIAL TEÓRICO

2.1 O processo de dataficação, big data e pandemia

Os dados se tornaram tão relevantes para a sociedade da informação quanto a gasolina para a indústria dos combustíveis, também ocupam lugares centrais em diversos setores do capitalismo contemporâneo. Na era do *Big Data*, significa que os dados são uma forma de capital para tudo, desde a casa inteligente à cidade inteligente, áreas da economia, governança, produção para a distribuição, dispositivo de consumo etc. (KITCHIN, 2014). Para a cobertura jornalística não é diferente, tendo em vista o uso cada vez mais intenso de dados e visualizações de dados para as narrativas, como é o caso do Nexo Jornal.³

Para Storey e Song (2017, p.50), o *Big Data* pode ser pensado como algo que “refere-se ao alto volume, velocidade e variedade de ativos de informação que exigem formas inovadoras de processamento para a tomada de decisões e otimização dos processos”⁴, esse fenômeno não se restringe aos avanços tecnológicos. No entanto, “a verdadeira revolução não está nas máquinas que calculam os dados, e sim nos dados em si e na maneira como os usamos” (MAYER-SCHONBERGER; CUKIER, 2013, p. 5).

Inserida nesse contexto, a dataficação sugere que alguma coisa seja transformada/processada em dados quantificáveis. Isto é, a capacidade de registrar informações, quantificar, analisar essas informações e extrair valor a partir delas, tendo em vista que todos os dados podem apresentar algum valor desde que estejam em um contexto. Para além das definições, pode ser considerado como um novo paradigma na Ciência e na Sociedade (VAN DIJCK, 2017) e com impacto direto no jornalismo e nos fluxos informacionais das redes digitais.

O termo ganhou popularidade acadêmica ao ser introduzido pelos autores Mayer-Schönberger e Cukier, em 2013, na obra que aborda o fenômeno do *Big Data*⁵, analisado em várias áreas do conhecimento. Apesar de esse fenômeno impactar e transformar a logística do volume informacional, os autores ponderam e dizem que estamos nos primórdios dessa Era e que desafia a maneira como vivenciamos toda essa transformação em escala global. “O *Big Data* pode exigir que mudemos, que nos acostumemos com a desordem e incerteza”

³ <https://www.nexojornal.com.br/>

⁴ No original: “refers to the high volume, speed and variety of information assets that require innovative forms of processing for decision making and process optimization”.

⁵ Livro: *Big Data: Como extrair volume, variedade, velocidade e valor da avalanche da informação cotidiana*. Rio de Janeiro: Elsevier, 2013.

(MAYER-SCHÖNBERGER; CUKIER, 2013, p. 34). Em uma perspectiva semelhante, um dos desafios que o *Big Data* pode enfrentar é que esse processo mescla dados de diferentes origens, denominados “dados podres” (LUPTON, 2015), acarretado pela tarefa dos dados coletados de que “a capacidade de coletar *big data* não foi acompanhada pela capacidade de analisá-los e interpretá-los de maneira significativa e sofisticada” (SPAAN; THIEL, 2017, p. 2).

Os estudos dos dados digitais revelam-se como matrizes oriundas de um conjunto de disciplinas em vários níveis e, portanto, envoltos numa complexidade latente na cadeia de processualidade desses dados, quais sejam: produzir, coletar, filtrar, analisar e representar dados para diferentes finalidades e, com isso, instaurando uma espécie de controvérsia dos dados. Nessa direção, a dataficação vai implicar dois aspectos: 1) a infraestrutura externa através da qual são coletados, processados e armazenados; e 2) os processos de geração de valor, que incluem monetização, mas também meios de controle estatal, produção cultural, empoderamento cívico etc. (MEIJAS; COLDRY, 2019), que funcionam como um tipo de multicamadas que incluem os procedimentos de disseminação, acesso, armazenamento, análise e vigilância que são de propriedade ou controlados principalmente por corporações e estados.

A utilização intensiva dos dados sobre a Covid-19, desde o seu surgimento em 2020, tem se tornado ubíqua na sociedade contemporânea e nos meios digitais, embora eles prescindam de uma conferência de sentidos, tendo em vista que “os dados não são objetos puros ou naturais com uma essência própria. Eles existem em um contexto, tendo um significado no contexto e forma na perspectiva do observador” (BORGMAN, 2015, KINDLE LOCATION p. 607)⁶. Nesse sentido, o próprio processo de dataficação originou o que denominamos de *datademia*, dados ubíquos sobre a pandemia e que tiveram (e continuam tendo) forte apropriação pelos veículos de comunicação em larga escala. A dataficação propicia recursos criativos ao tornar públicos e interativos os dados epidemiológicos. Assim, o potencial criativo é amplificado para a construção de novos sentidos e conhecimentos. E, naturalmente, carrega uma responsabilidade ética. Dataficar um fenômeno significa transformá-lo em informação quantificável, que seja passível de contabilizar, e esse aspecto recursivo o torna um campo desafiador em curso. Os dados são informações que podem se apresentar de maneira editada, analisada e combinada com outros dados e, conseqüentemente,

⁶ No original: “Data have no value or meaning in isolation. They can be assets or liabilities or both. They exist within a knowledge infrastructure - an ecology of people, practices, technologies, institutions, material objects and a relationship”

podendo criar um espaço estruturado e datafocado de informações e serem usados na construção de narrativas jornalísticas, a exemplo da visualização de dados.

Nesse sentido, os dados se configuram como elementos fundamentais na produção de conhecimento na sociedade e que é gerenciada a partir deles. Nesse ponto de vista, a importância dos dados vai para além de sua simples representação da realidade do mundo, são construções *sobre* o mundo, mas que mantêm seu estatuto social na medida em que eles estão inseridos em um contexto, e é material, na medida em que ganham forma (símbolos, números etc.). A infraestrutura de dados e os bancos de dados não são simplesmente meios técnicos de montagem e compartilhamento de dados neutros, como também não são meramente produtos que armazenam dados capturados sobre o mundo, mas processos relacionais que fazem trabalho no mundo (KITCHIN, 2014). A era dos grandes volumes de dados está adquirindo grande atenção, assim como aconteceu com o *Big Science* há 15 anos. A autora explica que o dado apresenta muitos tipos de valores e que estes podem variar de acordo com o lugar, tempo e contexto. Ela ressalta que “os dados não têm nenhum valor ou significado em isolamento. Eles podem ser ativos ou passivos ou ambos. Eles existem dentro de uma infraestrutura de conhecimento - uma ecologia de pessoas, práticas, tecnologias, instituições, materiais, objetos e relações”⁷ (BORGMAN, 2015, KINDLE VERSION, LOCATION 345).

Considerando a ubiquidade dos dados na sociedade contemporânea, a autora nos alerta que seu entendimento depende do contexto no qual estão inseridos e do olhar do pesquisador. “Os dados não são objetos puros ou naturais com uma essência própria. Eles existem em um contexto, tendo um significado no contexto e forma perspectiva do observador”⁸. A autora enxerga a problemática de definição sobre o que se entende por dados e os define como formas de informação, um amplo conceito que está cada vez mais difícil de definir, ressaltando que os problemas epistemológicos e ontológicos são abundantes, resultando em muitos livros dedicados à explicação de que são dados.

⁷No original: “Data have no value or meaning in isolation. They can be assets or liabilities or both. They exist within a knowledge infrastructure - an ecology of people, practices, technologies, institutions, material objects and a relationship”.

⁸ No original: Data are not pure or natural objects with an essence of their own. They exist in a context, have meaning in the context and form the viewer's perspective.

3 RESULTADOS E DISCUSSÕES

3.1 As dimensões dos dados no contexto pandêmico

A análise exploratória permitiu que os dados epidemiológicos fossem divididos em três dimensões, quais sejam: a) Dimensão Expositiva: diz respeito a dados que apenas têm a função de explicar os dados, e, portanto, sem interação; b) Dimensão Simulativa: os dados epidemiológicos são simulados no que se refere às formas de contágio, como se prevenir, simulação do *lockdown*; e c) Dimensão Interativa-interpretativa: são as Narrativas Visuais propriamente ditas com as funções de interagir com os dados e com densidade interpretativa que agrega também as características das outras duas dimensões. Trata-se, assim, de Visualizações de Dados mais completas e demonstrando maior robustez informativa sobre os desdobramentos da Covid 19.

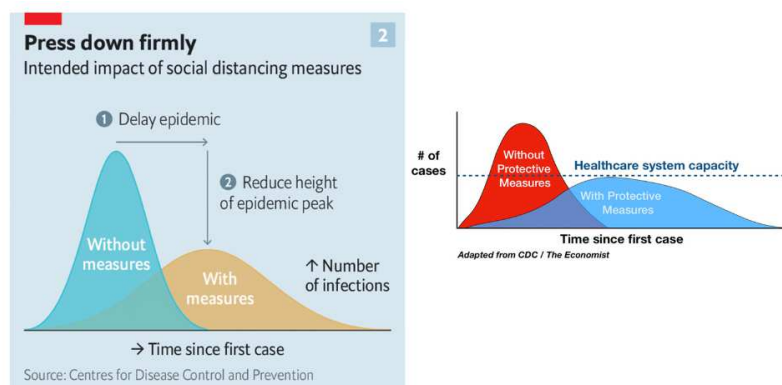
Na Dimensão Expositiva, que podemos denominar de *flatten the curve* (achatar a curva), constatamos uma profusão desse tipo de formato visual na tentativa de decodificar o impacto da pandemia. Esses tipos de visualizações se espalharam rapidamente pelos veículos da imprensa, sendo o jornal *The Economist*⁹ (Gráfico. 1), feito Rosamund Pearce, jornalista de dados do jornal a criar e o primeiro a publicar mostrando a curva do contágio e que logo foi absorvido por outros veículos de comunicação, a exemplo do *The New York Times*¹⁰. A ideia do Rosamund Pearce (jornalista de dados do TWPost) em criar e publicar esse tipo de visualização foi inspirado pelo órgão *Center for Disease Control And Prevention*,¹¹ (Gráfico. 2) com sede nos EUA, ao publicar dados do avanço e medidas preventivas do vírus da influenza, em 2017.

⁹ Disponível em <https://tinyurl.com/43md5ak4> Acesso em: 20 jul de 2020.

¹⁰ Disponível em: <https://tinyurl.com/2p8cmvu8> Acesso em: 2 jun de 2022.

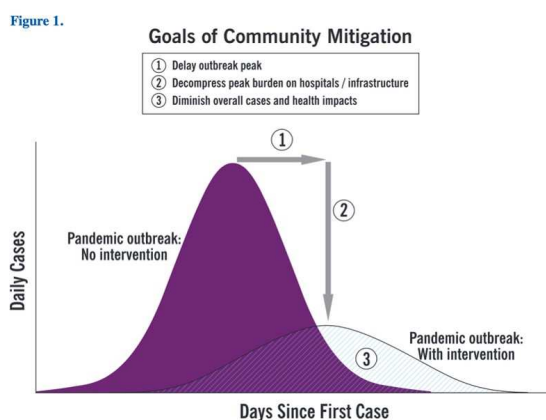
¹¹ Disponível em: <https://tinyurl.com/bdzdwc5> Acesso em: 5 jun de 2022

Gráfico 1- Gráficos Achate a curva do The Economist e uma versão adaptada no The New York Times



Fonte: The Economist e The New York Times

Gráfico 2- Achate a curva do Center for Disease Control and Prevention



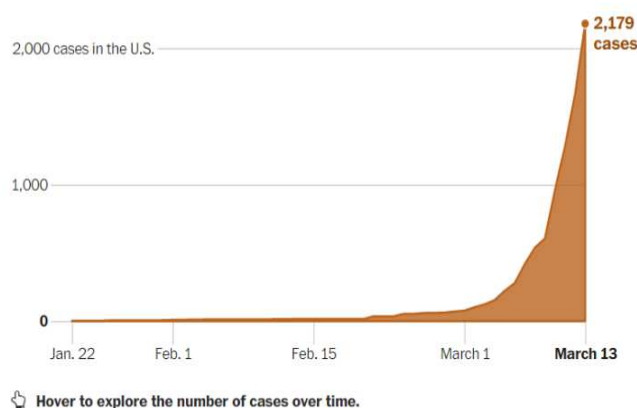
Fonte: Center for Disease Control and Prevention (CDP)

Ao observar os gráficos acima, nota-se que a intenção foi de mostrar a curva das pessoas infectadas, mas não aparece nenhum número, apenas mostra formas de achatar a curva ao diminuir o espalhamento do coronavírus. Ambas as produções foram moldadas por uma variável chamada taxa reprodutiva, ou R. Ela representa, com efeito, o número de casos adicionais que cada novo caso dará origem. A facilidade deste gráfico ao explicitar a curva do contágio se dá pela própria natureza dele, de se mostrar de simples leitura e apreensão do conteúdo. Naquele primeiro momento da pandemia, precisava-se de informações epidemiológicas para achatar a curva de crescimento, que foi o objetivo da divulgação dos dados dos centros de saúde. Aqui, nesta dimensão, não se trata especificamente de

visualização de dados com mais densidade informativa, mas de gráficos sobre a pandemia que optaram pela simplicidade da produção.

Na Dimensão Simulativa, houve uma profusão de visualizações de dados fazendo simulações sobre o isolamento social, estratégia de mitigação do contágio e disseminação do coronavírus. Publicada em 14 de março de 2020 e assinada por Harry Stevens, jornalista do *The Washington Post*,¹² foi uma das reportagens mais acessadas da história do jornal e tinha como proposta simular distintos cenários da propagação da Covid-19, bem como a adoção de medidas de restrição (quarentena), uso de máscaras e álcool gel, entre outras informações. A reportagem esclarece sobre a propagação do vírus e enaltece os cenários simulativos de distanciamento social como a melhor forma de diminuir a disseminação do contágio, mostrando, em primeiro momento, um gráfico *The Exponential Curve* (Gráfico 3), que revela o avanço da doença. Os dados são apresentados em três categorias: pessoa saudável, pessoa doente e pessoa recuperada.

Gráfico 3- Curva Exponencial (Exponential Curve) feita pelo The Washington Post



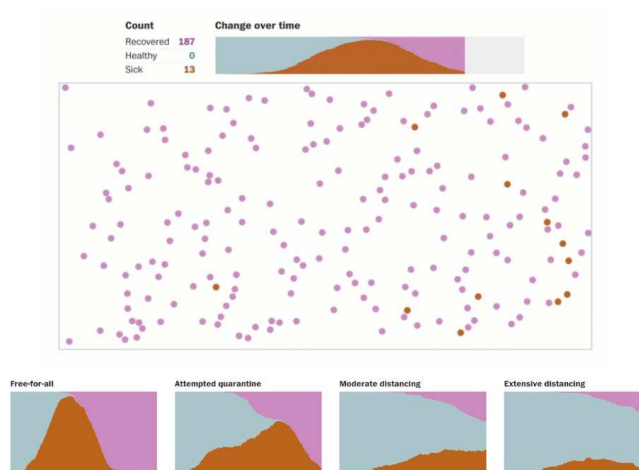
Fonte: The Washington Post

A organicidade da visualização de dados acontece em meio às simulações através dos cenários distintos como modo de prever o avanço, caso as medidas preventivas não fossem adotadas. As simulações, nesse caso, atuam como os sustentáculos dessa narrativa com dados, e, a partir deles, vão mostrando o cenário de contágio da covid, espalhamento, distanciamento social e quarentena (Gráfico 4). Hoje, passados mais de dois anos da pandemia, pode parecer

¹² Disponível em: <https://tinyurl.com/mtafstbb> Acesso em: 6 mar de 2020.

sem sentido algum simular o avanço da doença, mas, no início, foi fundamental para que pudéssemos ter a noção do agravamento do coronavírus, se munir de informações sobre o espalhamento (*spread*) do vírus, formas de prevenção etc. Nessa perspectiva, as visualizações de dados foram essenciais em assuntos como esse, de escala planetária, na tentativa de entender a magnitude do vírus e demais desdobramentos da doença. Na história dos dados epidemiológicos, as visualizações de dados atuaram na magnitude do problema através das investigações de John Snow,¹³ médico sanitário inglês que descobriu como a cólera se espalha, tendo o seu mapa mais citado na literatura, e de Florence Nightingale¹⁴, enfermeira e estatística inglesa, autora do mais famoso diagrama de área polar que mostra as causas da mortalidade no hospital militar.

Gráfico 4- Visualizações de dados simulativas feitas pelo The Washington Post



Fonte: The Washington Post

Na Dimensão Interativa-interpretativa são mostradas as visualizações de dados que apresentam em sua estrutura visual maior consistência dos dados e uma variedade de formatos gráficos. Assim, na *dataviz*¹⁵, feita pelo *The New York Times*¹⁶ em março de 2020, tem como foco mostrar o surgimento do vírus a partir do fluxo de viagens entre os países e de como ele saiu da China, num mercado de frutos do mar na cidade de Wuhan e se espalhou por essa localidade e por outros países. Ao analisar o fluxo de viagens, a narrativa traz dados de como

¹³Disponível em: <https://tinyurl.com/4zut36n7> Acesso em: 5 ago de 2022.

¹⁴ Disponível em: <https://tinyurl.com/4u9f3ubb> Acesso em: 5 ago de 2022.

¹⁵ Do inglês. Abreviação de *Data Visualization*

¹⁶ Disponível em: <https://tinyurl.com/36swjdys> Acesso em: 20 jul de 2022.

o vírus se distribuiu nas cidades de *Nova York*, *Sydney* e *Bangkok*, onde foi detectado o primeiro caso do coronavírus fora da cidade chinesa.

Gráfico 4- Visualização How the Virus Got Out feita pelo The New York Times



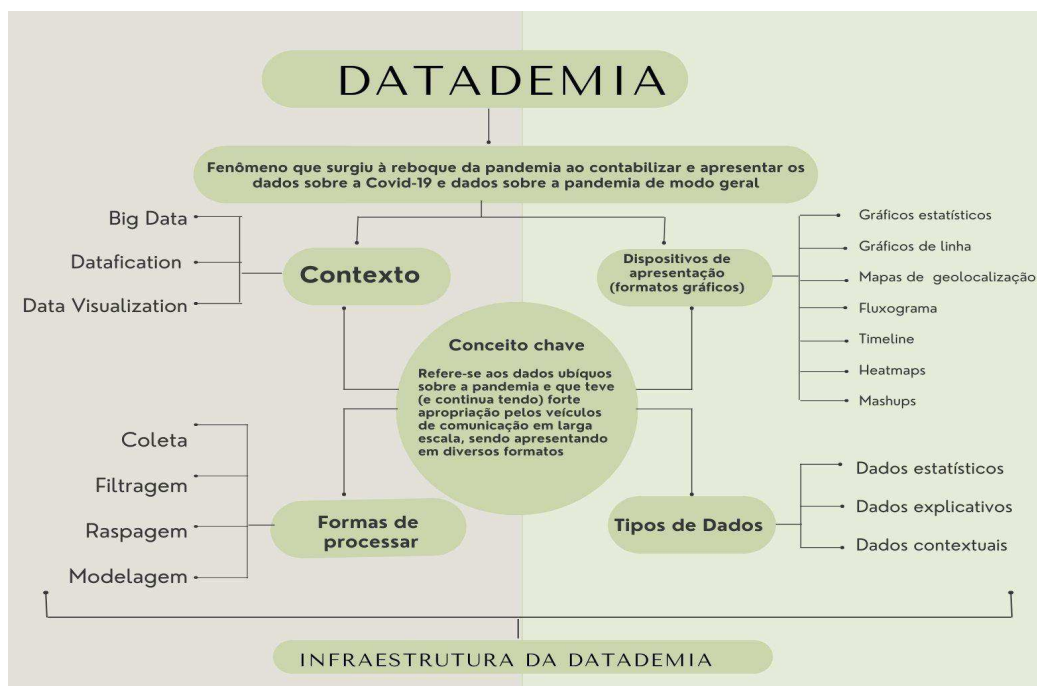
Fonte: The New York Times

Nessa perspectiva, observamos o que denominamos de *dataficação do vírus*, ou seja, no período analisado, observamos a proliferação dos dados sobre o coronavírus de modo intensivo e sistemático. Por se tratar de uma temática que por sua natureza requer números, o processo de dataficação se mostra oportuno para o momento da pandemia ao exibir números de óbitos, confirmados e recuperados. Tomando como base de que dataficação é quantificar (MAYER-SCHÖNBERGER; CUKIER, 2013), trata-se de um mecanismo essencial, tendo em vista o fluxo informacional em abundância disponível em redes digitais, que, quando bem exploradas, torna-se salutar nestas visualizações, apresentando características oxigenadoras de como esses dados são tratados, editados, apresentados e armazenados, principalmente nas narrativas jornalísticas. É nessa perspectiva que denominamos de dataficação do vírus, conforme mencionado acima, no todo esse conjunto de dados epidemiológicos ganha novos contornos e significados imersos no ambiente digital: não apenas como lugar de repositório estático de informações, mas um lugar que vai além de ter a capacidade ilimitada de armazenar.

As técnicas de *dataviz*, práticas já adotadas em diversos jornais mundiais, ocuparam o epicentro na pandemia, ou podemos falar em epicentro dos dados para revelar dados acerca das demais informações do vírus sob uma infinidade de formatos e estéticas. Nesse contexto e ao propor o fenômeno da Datademia, estamos circunscrevendo numa abordagem derivada dos processos de *big data* e da dataficação, que, para alguns autores, se configuram como paradigma (VAN DIJCK, 2014) ao demarcar esse movimento a partir das visualizações de dados intensivamente utilizadas na mídia mundial. É nessa perspectiva que repousa o nosso

objeto de estudo na ótica de posicioná-lo também como fenômeno que se reconfigura advindo de outros fenômenos em curso. Na infraestrutura da *datademia* (gráfico 1), pode-se visualizar os conceitos e demais desdobramentos do fenômeno em curso.

Gráfico 1 - Infraestrutura da Datademia



Fonte: A autora (2022)

O recorte, por tratar os dados provenientes da pandemia, fornece *insights* para compreender o movimento deles e sobre quais formas estão sendo apresentados, tendo em vista o aumento exponencial da quantidade de dados disponíveis nas últimas décadas e na reconfiguração da sociedade (MAYER-SCHÖNBERGER; CUKIER, 2014). Além do conceito chave, a infraestrutura foi pensada numa logística ampliada de como este conceito é operacionalizado. No que se refere à forma de processar os dados, tomamos como referências para este quesito os pesquisadores e estudiosos da big data (MANYKA, 2011; KITCHIN, 2014; MAYER-SCHÖNBERGER; CUKIER, 2014; MAR, 2015) e de visualização de dados (FRY, 2007; SEGEL; HEER, 2010; MURRAY, 2017; CAIRO, 2016) que, ao analisar os grandes volumes de dados, estabeleceram os procedimentos que auxiliam no processo de refinar os dados e transformá-los em estruturas visuais. Coletar, filtrar, raspar e modelar os dados se inserem na processualidade de como lidar com grandes quantidades de dados até gerar a visualização de dados que é um desmembramento do processo de modelagem.

Em dispositivos de apresentação, referem-se aos formatos visuais contidos numa visualização, quais sejam: gráficos interativos, gráficos de linha, mapas de geolocalização, fluxograma, *timeline*, *heatmaps* e *mashups*, todos formatos gráficos com dados sobre a pandemia foram os mais recorrentes na nossa análise. Importa ressaltar que esses formatos não são modelos canônicos de produzir uma visualização de dados, mas significam enfatizar a frequência deles sobre os dados epidemiológicos. Podemos apontar como fatores para essa recorrência o fato de serem, uma boa parte deles, de fácil assimilação, uma vez que no calhamaço de dados a todo o momento requer uma forma visual, em primeira instância, de fácil compreensão.

Um outro exemplo de como os mapas interativos foram explorados durante a pandemia foi o projeto “No epicentro”¹⁷, da Lupa, foi um dos vencedores da 2ª edição do *Sigma Awards*¹⁸ e da *Society for News Design (SND)*¹⁹ ambos os eventos que premiam as melhores produções em Jornalismo de Dados. O projeto, em parceria com o *Google News Initiative*, tem como ponto de partida a pergunta: “E se todos os mortos por Covid-19 fossem seus vizinhos?” e faz uma simulação de como a vizinhança ficaria se as vítimas ficassem perto de onde a pessoa mora. O usuário introduz o endereço e é exibida uma narrativa com o histórico do crescimento da Covid-19 em nível nacional. Além das premiações acima mencionadas, o projeto teve sua versão reproduzida pelo jornal americano *The Washington Post* com o nome de *At the epicenter*²⁰. Nesse sentido, esse exemplo da Lupa e a versão no jornal americano sobre a operacionalidade e visualização destes dados podem proporcionar um fluxo de trabalho dinâmico ao mesmo tempo desafiante frente aos grandes volumes de dados.

Nessa direção e tendo característica interdisciplinar, a datademia representa uma extensão criativa dos movimentos dos grandes volumes de dados (*big data*) e da dataficação, ambos processos em emergência na sociedade contemporânea, na qual essas duas tendências intensificam as formas de usos e apropriações que tendem a criar novos cenários emergentes, ao considerar as plataformas como espaços digitais potencializadores que incidem nos processos de produção, aceleração da circulação e consumo desses dados. Lemos (2021, p.195) ressalta que o processo de dataficação “é o procedimento básico da sociedade de plataformas” e que estas agem a partir de um arcabouço de dados coletados de modo indiscriminado. Ainda assim, o autor alerta para o fato de a dataficação não se restringir à

¹⁷ Disponível em: <https://piaui.folha.uol.com.br/lupa/epicentro/> Acesso em: 2 jun de 2022.

¹⁸ Disponível em: <https://sigmaawards.org/about/> Acesso em: 2 jun de 2022.

¹⁹ Disponível em: <https://snd.org/> Acesso em: 2 jun de 2022.

²⁰ Disponível em: <https://tinyurl.com/29dzzmf7> Acesso em 5 jun de 2022.

conversão do analógico para o digital, “mas da modificação de ações, comportamentos e conhecimentos baseados na performance dos dados elaborada por sistemas de inteligência algorítmica”. Com esse entendimento, a datademia pode ser enxergada, principalmente, como um fenômeno sociotécnico em construção que apresenta características e potencialidades que remodelam a forma de coletar, filtrar, refinar e apresentar esses dados.

3.2 Tipologia dos Dados da Datademia

No contexto da datademia, a gestão dos dados torna-se uma tarefa diária e muitas vezes árdua para as empresas, governos e organizações em como gerenciar. Dessa forma, duas atividades são imprescindíveis, a saber a) Gestão de Análise dos dados: usos e explorações do ferramental tecnológico pelas empresas que deem conta do calhamaço de dados em alta performance, e b) Gestão de Armazenamento dos Dados: requer a compreensão de que, ao lidar com grandes volumes de dados, faz-se necessário um ambiente adaptativo, tanto do ponto de vista profissional, quanto tecnológico, que assegure o armazenamento, recuperação e manipulação destes dados. A gestão dos dados é um procedimento central em um ambiente com altos volumes de dados e requer, dentre outros atributos, os implementos de políticas e plano de gestão no intuito de manter a integridade, segurança e utilização desses dados.

De um modo geral, os dados variam em conformidade com sua forma (qualitativa ou quantitativa), estrutura (estruturados, semiestruturados ou não), fonte (capturado, derivado, exaustão, transitória), produtor (primário, secundário, terciária) e tipo (indexical, atributo de metadados). Eles também podem apresentar muitos formatos, incluindo números, símbolos, imagens, texto etc. (KITCHIN, 2014). Na infraestrutura da datademia, identificamos três tipos específicos de dados epidemiológicos (gráfico 2), a saber.

Gráfico 1 - Tipologia dos Dados no contexto da Datademia



Fonte: A autora (2022)

Ao esquematizar a tipologia dos dados, sendo os resultados da pesquisa exploratória, definimos cada tipo de dado encontrado, a saber:

- a) **Dados Estatísticos:** Consistem em registros numéricos. Têm a característica de serem extensos e se relacionam com os atributos dos fenômenos (altura, distância, peso, área, volume, comprimento), ou são representativos e estão vinculados com as características não-físicas de fenômenos (como a classe social, nível educacional, privação social, a qualidade de rankings de vida);
- b) **Dados Qualitativos:** São dados não-numéricos, como textos, figuras, imagens, áudio, imagens 3D, 360°. Estes podem ser transformados em dados quantitativos. A análise de dados qualitativos acontece sob dados originais, na busca de construir significado e compreensão;
- c) **Dados Contextuais:** São dados que intercalam entre os dados qualitativos e os dados estatísticos, podem ser organizados e deslocados em uma estrutura visual mesclando com números, textos, ilustrações, mapas, entre outros formatos gráficos para conferir uma amplitude de compreensão do que está sendo apresentado a partir dessa fusão de ambos os tipos de dados.

Os esforços empreendidos ao esquematizar a tipologia dos dados recaem no entendimento de que, dada a complexidade dos estudos dos dados, em que há uma profusão de definições, não se configura como um modelo estático e padronizado, e, portanto, apresentando rigidez conceitual. Pelo contrário, trata-se, nesta investigação, de um recorte temporal (pandemia), um prognóstico flutuante de um universo ainda maior que requer aprofundamento de pesquisa com outras formulações de hipóteses. Os Tipos de Dados explorados nesta pesquisa têm um foco de detalhar as especificidades dos dados epidemiológicos, mas que eles se intercambiam dependendo das propostas da visualização, uma vez que “os dados podem ser totalmente descritos através de representações formais e estruturais” (SETZER, 2015, *online*).

Assim, é preciso ressaltar a natureza e heterogeneidade dos dados encontrados nos resultados. Em primeiro lugar, grande parte da origem dos dados observados se refere a dados e bancos de dados da saúde oriundos de empresas e centros de saúde oficiais do governo que divulgam os dados pandêmicos. Neste período, uma profusão de dados é divulgada indiscriminadamente sobre a Covid-19 nas redes, os quais logo se espalham com grande velocidade, porém, os veículos de comunicação reforçaram os cuidados em explorar fortemente as fontes oficiais para as criações gráficas. Não que os dados governamentais sejam, de fato, verídicos por serem oficiais, mas, de certa forma, é como se fossem um carimbo na visualização de dados que conduz à confiabilidade dos dados apresentados, e assim, a relevância da peça gráfica, atuando como grandes norteadores. Em segundo lugar, a heterogeneidade dos dados transformados em visualizações revela como, dentro de ambientes digitais dinâmicos, a flexibilidade e a mutabilidade das formas e formatos podem ser benéficas e podem trazer novas oxigenações, demarcar tendências visuais de dados e formatos expressivos e múltiplos. Por se tratar de um assunto não tão atrativo por si só, há formatos gráficos eminentemente criativos e de excelência na estética e conteúdo, onde há essa sincronicidade ao representar os dados pandêmicos.

4 CONSIDERAÇÕES FINAIS

O artigo apresentou o conceito de datademia, a partir da mescla conceitual dos processos de *big data* e dataficação, e analisou a cobertura midiática sobre a Covid-19 através da pesquisa exploratória. Desse modo, a articulação da tríade *big data*-dataficação-datademia pode ser uma condição contínua inserida nos processos digitais dinâmicos na cobertura jornalística e que passam por alguns parâmetros, quais sejam: a) Movimento dos Dados

Abertos (*open data*): Considerando que dados abertos “são estruturados para facilitar a leitura em máquina e que devem ser livremente acessíveis” (BORGMAN, 2015, *KINDLE VERSION, LOCATION* 345), *Open data* depende de tecnologias abertas e que são desenvolvidas para compartilhar, expandir as possibilidades de disseminação da informação que tem impacto na gerência dos dados abertos; 2) Gestão de Dados (análise): ferramenta tecnológica para análise dos grandes volumes de dados, não somente a realização de coleta, mas que apresente poderio analítico; e 3) Gestão de Dados (armazenamento): Dispor de uma infraestrutura de armazenamento da diversidade dos dados coletados, como imagens em movimento, áudios, vídeos etc., com acesso e facilidade de manipulá-los.

Em decorrência dos grandes volumes de dados e diversidade das fontes, os desafios de se obter dados não significativos são altos, tendo o dilúvio de dados encontrados e o desafio de minerar recai na confiabilidade dos dados, uma vez que fontes de grandes volumes de dados podem se mostrar que, não sendo tão confiáveis, são propensos a erros e interrupções. Nesta direção, os enfrentamentos desafiadores da tríade *big data*-dataficação-datademia podem se tornar aliados nos processos de viabilizar e apresentar os dados epidemiológicos porque trazem tentativas de minimizar as problemáticas diante do dilúvio de dados, a partir de objetivos superiores que é de acesso livre dos dados a toda sociedade, sejam eles em quais formatos forem apresentados e/ou disponibilizados.

A infraestrutura da datademia pode ser compreendida como um compêndio teórico/esquemático, apresentando uma interface com as características e demais aspectos que moldam as visualizações de dados epidemiológicos. Através dele, a tipologia dos dados estabelecida é uma tentativa de compreender o fenômeno, bem como o seu universo, mas que não se mostra um esquema estático. Ao contrário, consideramos um conceito fluido, cujos primeiros contornos conceituais foram empreendidos nesta investigação e podem ser mais bem alicerçados com outras pesquisas/análises empreendidas.

A tríade *big data*-dataficação-datademia pode configurar em práticas viabilizadoras diante o dilúvio de dados, ao mesmo tempo em que problemáticas de gestão e gerenciamento de dados neste processo podem inviabilizar a compreensão dos dados. Em um outro momento, pode-se analisar esse conceito em um universo de amostra de produções gráficas realizadas em universidades brasileiras e estrangeiras e de projetos laboratoriais com essa finalidade.

REFERÊNCIAS

- BECKER, B. Reconfigurações do Jornalismo Audiovisual: um estudo da cobertura do Fantástico sobre a pandemia da Covid-19. **Lumina**, v. 15, n. 3, p. 6–22, 2021. Disponível em: <https://periodicos.ufjf.br/index.php/lumina/article/view/35300> .Acesso em: 6 ago. 2022.
- BORGMAN, C. **Big Data, little data, no data**. Scholarship in the networked world. Cambridge-London: The MIT Press, 2015. (Kindle version).
- CAIRO, A. **Truthful Art, The: Data, Charts, and Maps for Communication**. New Riders, 2016.
- SPAANJ, R; THIEL, A. (2017) Big data: critical questions for sport and society. **European Journal for Sport and Society**. Disponível em: <https://tinyurl.com/yc2y9bk3>. Acesso em: 3 mar de 2022.
- FRY, B. **Visualizing Data**. Publisher(s): O'Reilly Media, 2017.
- KITCHIN, R. Big Data, new epistemologies and paradigm shifts. **Big Data & Society** April–June, 2014. Disponível em: <https://tinyurl.com/3auxb9s6> Acesso em: 5 abr 2020.
- MACHADO, E. PALACIOS, M. Um modelo híbrido de pesquisa: a metodologia aplicada pelo GJOL. In: LAGO, C; BENETTI, M. **Metodologia de pesquisa em Jornalismo**. Petrópolis, RJ: Vozes, 2007 p.199-222.
- MANIKA, J *et al.* 2011. **Big Data: The next frontier for innovation, competition, and productivity**. Disponível em: <https://goo.gl/EHAhVV> acesso em: 14 jul. 2014.
- MAYER-SCHONBERGER, V; CUKIER, K. **Big Data: Como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana**. Rio de Janeiro: Elsevier, 2013
- MEJIAS, U. A; COULDRY, N. (2019). Datafication. **Internet Policy Review**, 8(4). Disponível em: <https://policyreview.info/concepts/datafication> Acesso em: 1 mai de 2022.
- MURRAY, S. **Interactive Data Visualization for the Web**. Sebastopol: O'Reilly Media, 2013
- LEMONS, A. (2021). Dataficação da vida. **Civitas - Revista De Ciências Sociais**. Disponível em: <https://doi.org/10.15448/1984-7289.2021.2.39638>
- SEGEL, E; HEER, J. Narrative visualization: Telling stories with data. **Visualization and Computer Graphics**, IEEE Transactions on, 16(6):1139–1148, 2010. Disponível em: <https://goo.gl/TmNgFu> Acesso em: 3 set. 2018
- STOREY, V. C.; SONG, I. Big data technologies and management: what conceptual modelling can do. **Data & Knowledge Engineering**, 2017. Disponível em: <https://tinyurl.com/mrxnkdvh> Acesso em: 25 jun. 2018.

VAN DIJCK, J. **Datafication, dataism and dataveillance**: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 2014. Disponível em: <https://tinyurl.com/mw43ycns> Acesso em: 3 set. 2018.

Como Referenciar este Artigo, conforme ABNT:

RODRIGUES, A. A. Dataficação e Desinformação: Datademia na Cobertura Sobre Covid-19 em Redes Digitais. **Rev. FSA**, Teresina, v. 21, n. 12, art. 5, p. 92-111, Dez. 2024.

Contribuição dos Autores	A. A. Rodrigues
1) concepção e planejamento.	X
2) análise e interpretação dos dados.	X
3) elaboração do rascunho ou na revisão crítica do conteúdo.	X
4) participação na aprovação da versão final do manuscrito.	X