



Jornalismo e Dados: a apuração a partir das bases de dados do Sistema de Informação e Agravos e Notificações do DATASUS

Luan Moraes Romero¹

Universidade Federal de Santa Maria.

Resumo: O estudo propõe uma aproximação teórico-metodológica sobre o uso de bases de dados públicas para a produção jornalística problematizando a utilização do Sistema de Informação de Agravos e Notificações (SINAN) como fonte jornalística, no contexto dos estudos de Jornalismo Guiado por Dados (JGD) (TRASEL, 2014). Assim, contextualizam-se as discussões teóricas em torno do termo JGD; discute-se a constituição do SINAN para refletir as configurações das bases de dados abertos sobre saúde; desenvolvem-se observações empíricas no SINAN para problematizar as rotinas de apuração no JGD e aponta-se caminhos e limitações para os usos dos dados contidos no SINAN como fonte jornalística. Por fim, concebe-se que a formação das bases de dados possui atravessamentos de fatores de diversas ordens, como fatores políticos, econômicos, sociais e epidemiológicos, que devem ser considerados na apuração jornalística.

Palavras-chave: base de dados, Jornalismo Guiado por Dados, apuração, fonte jornalística.

1. Introdução

A emergência das práticas do jornalismo de dados vem se popularizando na última década. Em decorrência desta popularização, aumenta a demanda de profissionais especializados no uso de bases de dados para a produção jornalística. Neste contexto, as bases de dados abertas seriam mais uma fonte documental (GEHRKE, 2018) disponível para o uso dos jornalistas em sua produção diária de notícias. Desse movimento, algu-

¹ Estudante de mestrado de Comunicação da Universidade Federal de Santa Maria (UFSM). E-mail: luan_155@hotmail.com

mas questões emergem para o campo, especialmente relacionadas à ética de acesso a dados e às técnicas de leitura e comparação de bases de dados.

Parte-se dessa problemática para propor uma aproximação teórico-metodológica acerca do uso de bases de dados públicas para a produção jornalística. De modo particular, analisa-se o Sistema de Informação de Agravos e Notificações (SINAN), ligado ao Ministério da Saúde, buscando questionar seus potenciais e limites como fonte jornalística, no contexto dos estudos de Jornalismo Guiado por Dados (TRASEL, 2014).

2. Fundamentos Teóricos

A computadorização da cultura, conforme Manovich (2001), possibilita não só a emergência de novas formas de linguagem de mídia, como também reconfigura as já existentes. O autor concebe as bases de dados - uma coleção estruturada de dados que é organizada para a procura rápida de informações - como uma das formas culturais emergentes com o advento da Internet. Na discussão particular acerca de como as bases de dados podem ser usadas na confecção de produtos jornalísticos, consideram-se pertinentes as contribuições dos estudos sobre o Jornalismo Guiado por Dados, como explicadas por Träsel (2014, p. 119). O pesquisador se atém às formas com que as tecnociências podem ser inseridas nas rotinas produtivas do Jornalismo.

Entre as considerações que Träsel (2014) aponta está a crença sobre a possibilidade de substituição de fontes humanas por fontes numéricas, o que mostra indícios da difusão do imaginário cibernético na cultura jornalística. Tal substituição poderia, supostamente, fazer com que as rotinas produtivas jornalísticas eliminassem os erros humanos. Como discute Träsel (2014, p. 228) a tecnofilia “permitiria um aprimoramento da apuração das notícias pelos repórteres, que, através da Internet e de ferramentas de análise numérica, poderiam estabelecer uma base factual mais sólida, do ponto de vista da objetividade”. Outra possibilidade, também nesta visão, seria a inversão: as fontes humanas não pautariam mais o noticiário, mas os jornalistas, de posse de dados, poderiam ir a elas e confrontá-las a partir de uma apuração prévia.

Contudo, é importante considerar as críticas apontadas por pesquisadores como Boyd e Crawford (2012). Em artigo sobre a discussão em torno da problemática das pesquisas em Big Data, os autores propõem provocações que podem apontar armadilhas

óbvias aos cientistas sociais, porém surpreendentes para investigadores de outras áreas e, podemos destacar de modo particular para personagens não necessariamente relacionados com a pesquisa científica - no caso dos jornalistas. É importante frisar que as reflexões propostas são direcionadas para os estudos feitos com Big Data retirados de redes sociais digitais, contudo alguns pontos são relevantes, porque trazem à tona problemáticas correlatas aos usos de base de dados em outros contextos. Apresenta-se de maneira resumida tais provocações, e depois discute-se a partir delas as imbricações relevantes para este trabalho.

A primeira provocação discorre sobre como o Big Data muda a definição de conhecimento. Para Boyd e Crawford (2012) a emergência de um aparato diferente para a construção do conhecimento faz com que haja também a mudança dos objetos estudados. Contudo, é preciso ser feita a análise crítica sobre tais aparatos, porque os números não falam por eles mesmos, refletem escolhas realizadas durante a montagem das estruturas das bases de dados e escolhas metodológicas quanto a coleta de informações para o preenchimento das planilhas. Outra questão trazida por Boyd e Crawford (2012) é que a exacerbada quantidade de dados não faz necessariamente com que sejam melhores dados. Os autores destacam a necessidade de colocar as informações coletadas em perspectiva, buscar saber qual a sua origem e identificar os vieses incutidos na subjetividade do pesquisador quando seleciona os dados, uma discussão empreendida a muitos anos por cientistas sociais segundo os autores.

Uma terceira consideração relevante, indicada por Boyd e Crawford (2012), é que ao retirar-se os dados coletados do contexto em que foram elaborados, eles podem perder significado (ou sofrer alterações de sentidos). Tal concepção tem implicações óbvias, como a de que dados sobre a saúde, que foram coletados sobre doenças, não podem ser confundidos com dados sobre economia, que se relacionam a renda, por exemplo. No entanto, há outras implicações mais complexas, como a articulação de dados de saúde com dados econômicos, em que é preciso ter um cuidado extremo ao se fazer determinadas inferências, já que sem conhecimento prévio é possível cair em certos determinismos já superados pelas ciências sociais. Por exemplo, ao se relacionar a renda com a incidência maior de algumas doenças em uma população, não se pode cair em um “economicismo”, de redução das realidades sociais a fatores econômicos.

Ao levarmos em consideração tais críticas, percebemos que há desafios e limitações para a utilização de bases de dados por jornalistas. Assim, concebemos como problemática de pesquisa propor uma aproximação teórico-metodológica sobre o uso de bases de dados públicas para a produção jornalística partindo de uma problematização sobre a utilização do Sistema de Informação de Agravos e Notificações (SINAN) como fonte jornalística no contexto dos estudos de Jornalismo Guiado por Dados (TRASEL, 2014). Apresenta-se a seguir o percurso metodológico empreendido.

3. Metodologia empregada

A pesquisa se inspirou em algumas das discussões propostas por Franciscato (2007) sobre a feitura de pesquisa aplicada em jornalismo. Com relação aos procedimentos metodológicos, o autor os distribui em três etapas: a preparatória à pesquisa, o desenvolvimento da pesquisa e a análise de resultados. A problemática que permeia esta investigação se dá em torno da questão “como os dados do SINAN (Sistema de Informação de Agravos de Notificação) podem ser usados como fonte jornalística?”.

Na fase preparatória efetuou-se uma pesquisa sobre as indicações do processo de apuração no jornalismo clássico e no Jornalismo Guiado por Dados. Em um segundo momento, descreveu-se a constituição das bases de dados analisadas, e como tais fatores contribuem para a apuração jornalística. O experimento centrou-se na escolha de caso ou situação específica para validação dos dados obtidos - o tema escolhido, com relevância noticiosa no momento da pesquisa, foi a febre amarela. Ao fim, retratou-se os resultados obtidos no experimento, assim como se problematizou as incongruências encontradas na manipulação dos dados sobre a febre amarela nas bases do SINAN.

3.1 A apuração no Jornalismo Guiado por Dados

Para Domingo et al. (2008), os processos de produção das notícias passariam por 5 estágios: acesso ou observação, seleção ou filtragem, processamento ou edição, distribuição e interpretação. Para os autores, o primeiro processo seria a captura da informação inicial, e do material necessário para a geração de uma notícia. O estágio de seleção seria quando decisões começaram a ser tomadas para a concretização do que será noticiado. Já a edição seria o momento que a história começa a tomar forma e é finalizada,

para passar ao processo de distribuição, em que ela é levada aos leitores, e, por fim, à etapa da interpretação, em que a notícia publicada vira alvo de comentários e discussões.

As rotinas produtivas propostas para a operacionalização do Jornalismo Guiado por Dados, como proposta por Bradshaw (2011), se divide em 5 processos: compilação, limpeza, contextualização, combinação e comunicação. Na fase de compilação, o autor fala que: “ou você tem uma questão que necessita de dados, ou um banco de dados que necessita ser questionado”² (BRADSHAW, 2011, s.p.). A segunda etapa - limpeza - seria para remover erros humanos que podem afetar a consolidação dos dados e converter os dados obtidos em uma forma possível de ser analisada. Já a contextualização é o momento de questionar a origem dos dados, a forma com que eles se organizam, ou seja, compreender a metodologia por trás da montagem das bases de dados. A quarta fase seria a combinação, em que se busca cruzar diferentes bancos de dados - como, por exemplo, colocar informações coletadas sobre algumas cidades em um mapa. E por fim, na fase de comunicação aconteceria a publicação do que foi reunido.

Considera-se que a fase inicial compreendida desde a formulação da pauta, coleta de informações até ao primeiro esboço do material a ser noticiado, pode ser denominada como o processo de apuração, pois “apurar pode resumir-se a um jogo de evidências confrontadas a outras.” (PEREIRA JÚNIOR, 2006, p.72). Dessa forma, esta investigação intenciona fazer sua intervenção em torno do processo de apuração, e não se pretende a finalização de um produto jornalístico.

3.2 Sistema de Informação de Agravos de Notificação (SINAN): o real versus o ideal

Dando prosseguimento aos passos metodológicos, apresentam-se as primeiras aproximações com o objeto empírico - o Sistema de Informação de Agravos de Notificação (SINAN), através das informações pertinentes sobre o Ministério da Saúde (MS), sobre o Departamento de Informática do SUS (DATASUS) e sobre a história do sistema. Depois fala-se, sob o ponto de vista jurídico, como o SINAN é idealmente constitu-

² “Data journalism begins in one of two ways: either you have a question that needs data, or a dataset that needs questioning” (BRADSHAW, 2011, s.p.)

ído e discute-se as implicações dessas escolhas, contrapondo com o que se pode verificar na realidade. Por fim, aborda-se brevemente a escolha metodológica sobre o caso analisado - os dados sobre febre amarela - e os procedimentos empreendidos para a validação de tais informações.

3.2.1 História e constituição jurídica

Segundo Lima (2009), o DATASUS passou a existir na estrutura do Ministério da Saúde com a constituição da Fundação Nacional de Saúde (FUNASA), em 1991. O sistema herdou as responsabilidades do Dataprev - mecanismo ligado a coleta de dados para fins previdenciários. Assim, no âmbito governamental, os dados relacionados à saúde serão organizados de maneira sistemática, desde essa época, para que possam ser utilizados em outras instâncias, para além da esfera previdenciária, como por exemplo, na gestão de recursos internos do Ministério da Saúde.

O DATASUS pode ser considerado um repositório de dados relacionados a saúde brasileira. Parte dele, o Sistema de Informação de Notificações e Agravos (SINAN) é um banco de dados de caráter epidemiológico, constituído a partir de 1993. Do ponto de vista jurídico, alguns conceitos relacionados ao SINAN são definidos pela portaria de consolidação nº 4, de 28 de setembro de 2017, emitida pelo Ministério da Saúde. Entre eles, é regulamentado o que vem a ser agravo, autoridades de saúde, doença, epizootia, entre outros. Também é previsto pela portaria de consolidação a notificação compulsória pelos profissionais de saúde ou responsáveis pelos serviços público e privados de saúde. Além disso, o decreto prevê quais são as doenças consideradas de notificação nacional, qual a periodicidade e para quem as unidades de saúde devem se reportar. Assim, doenças como coqueluche, dengue, febre amarela, HIV/AIDS, raiva, entre outras, são enfermidades de interesse nacional. Também são definidos os parâmetros de alimentação dos dados do SINAN, que relacionam o não cumprimento das atividades por parte dos agentes regionais e locais a implicações de repasses dos recursos econômicos do bloco de vigilância em saúde³.

³ Disponível em: <http://portalarquivos2.saude.gov.br/images/pdf/2017/novembro/17/Orientacoes-sobre-os-recursos-transferidos-no-Bloco-de-Vigilancia-em-Saude.pdf> Acesso 14/05/2018

É importante que o jornalista conheça, ainda que de forma indicial, os termos estabelecidos em lei para que possa fazer a “tradução” correta das informações para outros públicos. O termo “epizootia”, por exemplo, é um vocábulo do jargão das ciências biológicas que corresponde a doença ou morte de animal ou de grupo de animais que possa apresentar riscos à saúde pública, e deve ser traduzido e melhor descrito nas reportagens. Se olharmos para o caso específico da febre amarela, caso de estudo nesta pesquisa, a “epizootia” pode ser reconhecida, dependendo do caso, a partir de relatos sobre “a morte de bugios”⁴.

Por fim, a discussão sobre a estruturação do SINAN sob o ponto de vista histórico e jurídico dá indícios de como o sistema deve funcionar idealmente. Mas como poderemos argumentar próximo entretítulo, em especial a partir dos relatórios de gerenciamento do sistema, a prática de funcionamento do Sinan não necessariamente acontece a partir dessa previsão ideal.

3.2.2 O cotidiano médico e o gerenciamento do sistema

Ao se considerar a legislação em torno do SINAN, podemos inferir que o sistema possui estrutura suficiente para apresentar dados consolidados. No entanto, se a lei molda o que seria ideal, é no cotidiano dos profissionais de saúde e de tecnologia da informação do DATASUS que a realidade se mostra. Assim, como um dos passos metodológicos, analisou-se os planos diretores de tecnologia da informação elaborados pelo DATASUS.

A equipe do DATASUS costuma elaborar periodicamente o Plano Diretor de Tecnologia da Informação (PDTI). O primeiro relatório disponível no site do DATASUS é o elaborado para o biênio 2014-2015, contudo há uma versão anterior não disponível no site, do triênio 2010-2013.

Nos relatórios são feitos diagnósticos da gestão de recursos e dos processos de tecnologia da informação. Além disso, é elaborado um planejamento visando indicar quais são as áreas prioritárias de trabalho para a equipe, considerando que os dados in-

⁴ No contexto de estudo sobre a febre amarela, a morte de bugios pela infecção pela doença, é um indício de que o vírus está se espalhando na mata.

seridos no sistema devem ser usados para que diferentes órgãos institucionais possam tomar decisões e estabelecer políticas públicas.

Como exemplo das informações encontradas nestes documentos destacamos o último relatório, publicado em 2017, em que a equipe revela onze dificuldades encontradas através da análise feita. Dentre os diagnósticos indicados, os que mais podem afetar os usos das bases de dados para fins jornalísticos são a falta de integralização e disponibilidade dos dados, assim como sua desatualização (MINISTÉRIO DA SAÚDE, 2018, p.48). Além disso, outras dificuldades internas também são reveladas no relatório, como a descontinuidade de ações voltadas para a melhoria do DATASUS em virtude da troca recorrente de gestores; a falta de preocupação com o tratamento da informação, tendo em vista como se dão os processos internos do Ministério da Saúde; a dificuldade de informar sobre como os serviços são executados, o que indica a falta de transparência. Pelas métricas utilizadas no relatório fica evidente que a maioria das metas estabelecidas para o melhoramento dos serviços feitos pela equipe de informática do DATASUS ao longo dos anos não foram atingidas. Dessa maneira, mesmo tendo legislação própria, a efetivação do sistema e a consolidação dos dados depende tanto de fatores políticos (como a troca constante de gestores), quanto da cultura organizacional (disponibilidade dos profissionais em saúde para se adaptarem).

Além disso, para se entender as implicações relacionadas ao cotidiano dos profissionais da área da saúde foi feita uma entrevista semi-estruturada (GIL, 2008) com o médico e professor Marcos Lobato da Universidade Federal de Santa Maria. Lobato possui mestrado em saúde coletiva pela Universidade Federal do Rio Grande do Sul (UFRGS), assim como já atuou como secretário municipal de saúde nas cidades gaúchas de Estância Velha, Gravataí e Charqueadas, e ainda atua como consultor em gestão municipal da saúde. Além disso, colaborou na criação de um índice de vulnerabilidade social aplicado a políticas públicas do SUS. O índice já foi utilizado pelo governo do estado do Rio Grande do Sul como parâmetro para estabelecimento de políticas públicas em saúde.

Para Lobato, com relação às legislações, a obrigatoriedade por informar os nascimentos, óbitos e doenças infecciosas sempre existiu. O que mudou, na percepção do médico, é o uso dessas informações, já que agora elas passaram a ser abertas e publici-

zadas, e anteriormente eram restritas aos profissionais da saúde. No entanto, para além da obrigação em informar esses dados, Lobato destaca o cuidado ao publicizá-las, para que não se crie pânico na população. Ele cita um exemplo prático, da possível criação de pânico quando uma criança contrai meningite em uma escola, que os pais já passam a não querer enviar seus filhos, sem ao mesmo saber qual o tipo de meningite.

Outro ponto levantado pelo médico é como se dá a mobilização entre os profissionais da saúde em situações emergenciais, como em casos de epidemias. Há doenças que, com a simples notificação de um caso confirmado, já mobilizam um grande número de profissionais para investigar a origem e conter a propagação, como o caso da raiva, em que um caso identificado já constituiria uma epidemia. Assim, é importante que os jornalistas saibam que as especificidades de cada doença são levadas em consideração para o estabelecimento do que vem a ser uma epidemia, para que não se divulgue apressadamente que há uma quando este não for o caso, do ponto de vista técnico. O preconceito também é indicado como fator que tem relação com a notificação das doenças, um fator de ordem moral. O professor fala que algumas doenças são mais notificadas no setor público que no privado. E isso se deve, entre outros fatores, ao estigma social que certas doenças possuem, como doenças sexualmente transmissíveis em geral.

Como explica o professor Lobato:

Na questão dos agravos de notificação a gente tem muito problema. Problemas de todos os tipos, porque tem uma implicação social neles. Doenças infecciosas tem essa implicação social, ou de pânico, ou de preconceito. E até do ponto de vista político, do uso político, no sentido de tentar atingir ou deslegitimar um governo, um executivo.. “olha, não fazem ações de prevenção, proteção e não sei o quê, e por isso teve tal coisa”. Isso aconteceu recentemente com a febre amarela.

Evidencia-se que tanto a estruturação das bases de dados do DATASUS quanto a alimentação delas são perpassadas pelos fatores técnicos, organizacionais, políticos, econômicos e sociais. Dessa maneira, jornalisticamente essa percepção deve ser levada em consideração na elaboração de matérias e reportagens.

4. Febre Amarela: a tentativa da validação de dados

Após se ter feito a discussão sobre como o SINAN é idealmente constituído, e o contrapondo com o que se pode verificar na realidade, deu-se prosseguimento com a tentativa de validação dos dados dispostos no sistema. Para fazer a escolha metodológica acerca de quais dados utilizar para efetuar a validação, recorreu-se a uma análise das publicações veiculadas pelo site do Ministério da Saúde, através da tag “Agência Saúde”⁵. A coleta inicial ocorreu entre os dias 1 de janeiro de 2018 e 2 de maio de 2018, totalizando 5 meses de cobertura sobre o tema, sendo que todos os hiperlinks das notícias foram dispostos em uma planilha. Dos 613 hiperlinks, para fins de recorte metodológico, foram retiradas as notícias que não tinham relação com os dados contidos no SINAN, restando 70 notícias. Destas, fizemos uma classificação de temas: febre amarela (32 notícias), hanseníase (30 notícias), HIV/Aids (2 notícias), meningite e HPV (2 notícias), sarampo (2 notícias), gripe (1 notícia), hepatite C (1 notícia), tuberculose (1 notícia). Nota-se que dentre as doenças mais tratadas nas notícias veiculadas no período, a febre amarela se destacou. Dessa maneira, os dados a serem tratados foram escolhidos com base nesse levantamento prévio.

Do ponto de vista médico⁶, quando chega a uma unidade de saúde o paciente é avaliado de acordo com os sintomas apresentados. Assim, para fins de notificação são consideradas pessoas suspeitas de possuírem a febre amarela aquelas que: 1) tiverem um “quadro febril agudo (até 7 dias), acompanhado de icterícia e/ou manifestações hemorrágicas, não vacinado contra febre amarela ou com estado vacinal ignorado”; 2) quem tiver um “quadro febril agudo (até 7 dias), residente ou que esteve em área com transmissão viral (ocorrência de casos humanos, epizootias ou de isolamento viral em mosquitos) nos últimos 15 dias, não vacinado contra febre amarela ou com estado vacinal ignorado”. Já as formas de confirmar se a pessoa está ou não com febre amarela são: através do diagnóstico laboratorial ou se o caso suspeito evoluir “para óbito em menos de 10 dias, sem confirmação laboratorial, no início ou curso de surto ou epidemia, em que outros casos já tenham sido comprovados laboratorialmente”.

⁵ Quando fez-se a observação dos canais de comunicação mantidos pelo Ministério da Saúde, observou-se que as declarações oficiais do Ministério são encontradas em seu site indexadas pela tag “Agência Saúde”.

⁶ Disponível em: <http://portal.arquivos2.saude.gov.br/images/pdf/2014/janeiro/23/doen-infecciosas-guia-bolso-8ed.pdf> . Acesso 08/05/2018

Considerando que a febre amarela é uma doença que deve ter sua notificação em até 24 horas para a Secretaria de Saúde Municipal, é importante perceber que o caso notificado pode evoluir tanto para a confirmação de que seja realmente febre amarela, quanto para a uma notificação negativa, em que se constata que a pessoa não possui a doença. Os dados disponíveis no SINAN e analisados neste trabalho são de casos confirmados.

O acesso às bases de dados SINAN pode ser feito de diversas formas, contudo, verificou-se três caminhos possíveis⁷, mas todos eles houve dificuldades de encontrar as informações. Além disso, é preciso entender que as bases de dados do SINAN estão divididas em dois períodos de tempo: de 2001 até 2006, e de 2007 até 2016. Assim, para coletar os dados de 2001 até 2016 é necessário efetuar as duas coletas e depois efetuar o agrupamento das bases de dados. Ao buscar-se os dados brutos sobre febre amarela no DATASUS encontra-se a possibilidade de selecionar variáveis para linhas e colunas. Em termos matemáticos, essa representação é do tipo matricial, pois é possível representar as tabelas em função de suas variáveis da seguinte maneira: (“variável para linha”, “variável para coluna”). Ou, em termos leigos pode-se dizer que os dados são encontrados no formato de tabelas.

Os cruzamentos efetuados foram realizados entre as bases de dados relacionadas a “casos confirmados de febre amarela” de 2001 a 2006 e de 2007 a 2016, já que entre 2006 e 2007. Para fins de sistematização, indicam-se os dados cruzados através da notação matricial (linha, coluna). Os cruzamentos obtidos foram (evolução, sexo); (evolução, mês 1º sintoma(s)); (evolução, ano 1º sintoma(s)); (evolução, faixa etária); (município de notificação, ano 1º sintomas); (município de notificação, mês 1º sintoma(s)).

Optou-se por “município de notificação” pois, como a identificação efetuada pelo IBGE sobre os nomes das cidades é informada pelo sistema, é possível fazer o cruzamento e descobrir o estado e região de cada município. Além disso, a escolha por “ano 1º sintoma(s)” se deu para que seja possível analisar um grande período de tempo, assim com “mês 1º sintoma(s)” permite analisar ao longo de um ano como a doença se

⁷ Considerou-se os caminhos para encontrar as informações partindo de: 1)Portal SINAN (<http://portalsinan.saude.gov.br/>); 2)Partindo do DATASUS (<http://www.datasus.gov.br/>); e 3) Partindo do portal Sala de Apoio à Gestão Estratégica (Sage) (<http://sage.saude.gov.br/>)

“comportou”. Já “evolução” foi selecionada para se ter noção da letalidade da doença, e “sexo” foi escolhido por mera curiosidade, já que, se houver alguma alteração significativa entre a incidência da doença entre homens e mulheres, essa seria uma informação que poderia ser melhor investigada.

As informações contidas nas notas de rodapé do SINAN facilitam o manuseio do sistema, porque indicam caminhos para coletar dados para mais de um ano, por exemplo. Sem lê-las, pode se cair no infeliz caminho de fazer a coleta, ano por ano, para depois juntá-las. Além disso, nas notas de rodapé são indicadas as atualizações mais recentes dos dados.

Com posse de todos os cruzamentos para coleta dos dados efetuou-se o entrelaçamento dos dois períodos disponíveis. Nos dados que envolviam “municípios de notificação”, além de reuni-los na mesma planilha, reorganizaram-se as informações, excluindo cidades duplicadas. Além disso, destacaram-se com cores os dados referentes aos dois períodos, verde para os de 2001 até 2006 e branco para os de 2007 até 2016. Esse destaque é importante para que sejam efetuados os cálculos de soma dos totais, como também para cruzar as informações entre, por exemplo, (município de notificação, ano 1º sintoma(s)); (município de notificação, mês 1º sintoma(s)).

4.1 Inferência possíveis com base nos dados brutos

Com os dados já tratados, apresentam-se as inferências possíveis de serem efetuadas, posteriormente discute-se os resultados encontrados. Assim, temos:

1) do cruzamento da matriz (município de notificação, ano sintoma(s)) com a classificação de estado e região do IBGE: a febre amarela afetou 105 municípios, 14 estados, e todas as regiões de 2001 a 2016. Os cinco municípios mais atingidos em termos de casos confirmados absolutos foram: Serro (MG) - 21; Sabrinópolis (MG) - 19; Brasília (DF) - 17; Manaus (AM) - 13; Belo Horizonte (MG) - 11. O estado que mais teve municípios afetados foi Minas Gerais, seguido de São Paulo. Os anos que mais tiveram casos notificados foram: 2016 (52), 2003 (52), 2009 (47); 2008 (44), 2001 (35)

2) da matriz (município de notificação, mês sintoma(s)): as meses em que há mais casos confirmados são os meses de verão: dezembro (66), janeiro (91), fevereiro (29) e março (54).

3) da matriz (evolução, sexo): dentre os números analisados, enquanto 234 homens sofreram da doença, somente 44 mulheres tiveram febre amarela. Desses 102 homens ficaram curados, e 121 morreram; já entre as mulheres 23 foram curadas e 20 morreram vítimas da doença.

4) da matriz (evolução, mês 1º sintoma(s)): ao considerarmos os índices de letalidade, que são calculados através da seguinte expressão: índice de letalidade = número de óbitos*100/número de casos confirmados.

| Evolução caso | Jan | Fev | Mar | Abr | Mai | Jun | Jul | Ago | Nov | Dez | Total |
|---------------|-------|-------|-------|-------|-------|-------|-------|------|-------|-------|-------|
| | 46,15 | 34,48 | 44,44 | 42,86 | 37,50 | 100,0 | 66,67 | 0,00 | 66,67 | 66,67 | 50,18 |

Tabela 1: índice de letalidade mensal

É importante perceber que, a olhos leigos, o 100% de junho poderia indicar que este é o mês mais letal. Poderia se criar a manchete “Junho é o mês em que pessoas com febre amarela mais morrem”, por exemplo. De certa forma, o título é correto, se levarmos em consideração somente o índice de letalidade. Contudo, é preciso compreender que este “junho” representa os 16 “junhos” da série histórica, e além disso não houveram muitos casos confirmados, em números absolutos, durante este mês. O possível pânico gerado a partir do título proposto é considerável, já que nem todas as pessoas vão ter acesso aos dados brutos para discernir o que este “junho” representa. Dessa forma, o índice de letalidade pode ser indicador de informações importantes, mas também indicar percepções que, para pessoas leigas, geram confusão.

Constatou-se que os municípios de Maués (AM) e Jataí (GO) não estavam presentes na matriz (município notificação, mês 1º notificação). Não se tem informações suficientes para falar sobre por que esses dois municípios não aparecem. No entanto, tal fato implica por extrapolação em um descompasso no restante das informações sobre febre amarela presentes no sistema. Com por exemplo, o número total de casos confirmados não é o mesmo: na matriz (município de notificação, ano 1º sintoma(s)) é 278, e na matriz (município de notificação, mês 1º sintoma(s)) é 273. Ainda, analisando as implicações desse descompasso, pode-se observar que o erro se mantém em outras ma-

trizes, pois no cruzamento das variáveis “ano 1º sintoma(s)” e “mês 1º sintoma(s)” com “evolução” o descompasso se manteve.

5. Considerações Finais

Manovich (2001) indica que as bases de dados são a principal forma de expressão cultural de nossos tempos. Podemos compreender, assim, que a formação das bases de dados investigadas possui atravessamentos de diversas ordens, como fatores políticos, econômicos, sociais e científicos. Investigar as implicações desses fatores, e de outros que possam emergir, na formação de outras bases de dados é fundamental para qualificar o trabalho jornalístico.

Nesta pesquisa, partimos dessa premissa geral para compreender as particularidades do DATASUS, a partir de sua história, sua gestão, os critérios de sua alimentação, e por suas características. O experimento realizado buscou dar conta de executar os cruzamentos possíveis dos dados brutos, tendo como caso a febre amarela. Demonstramos quais inferências eram possíveis de fazer, assim como tencionamos aquelas que podem gerar distorções. Por fim, indicamos as incongruências encontradas, que possuem reflexos no fazer jornalístico, dentre as quais a falta de exposição dos dados a partir de 2016 e a incongruência entre diferentes eixos de pesquisa – em alguns cruzamentos em matriz encontramos a ausência de dados sobre determinados municípios.

Do ponto de vista político-econômico, as políticas públicas de saúde são elaboradas tomando como referência os dados contidos nessas bases de dados. Do ponto de vista jornalístico, esses dados não são confiáveis para serem utilizados sem uma prévia análise. Estatisticamente é possível conceber modelos matemáticos que levem em consideração discrepâncias como essas, contudo, se aumenta o percentual da margem de erro. Por fim, é importante destacar que é possível continuar com a investigação a partir dos dados disponíveis no SINAN. Há informações sobre outras doenças, que podem ser analisados separadamente, e posteriormente fazer reflexões sobre como a especificidade de cada doença pode ser apreendida pelas bases de dados. Além disso, é possível investigar também os outros sistemas inseridos no DATASUS, como o Sistema de Informações sobre Nascidos Vivos (SINASC).

Referências bibliográficas

BARBOSA, Suzana. Jornalismo digital em base de dados (JDBD): Um paradigma para produtos jornalísticos digitais dinâmicos. 2007. 329 f. Tese (Doutorado) - Curso de Comunicação, Universidade Federal da Bahia, Salvador, 2007. Disponível em: <https://www.facom.ufba.br/jol/pdf/tese_suzana_barbosa.pdf>. Acesso em: 22 mar. 2018.

BRADSHAW, Paul. The inverted pyramid of data journalism. Online Journalism Blog, 7 jul. 2011. Disponível em: <http://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism>. Acesso: 20 fev. 2018.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. Sistema de Informação de Agravos de Notificação – Sinan: normas e rotinas / Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Vigilância Epidemiológica. – 2. ed. – Brasília : Editora do Ministério da Saúde, 2007. Disponível em http://bvsmms.saude.gov.br/bvs/publicacoes/07_0098_M.pdf . Acesso 05/03/2018

BOYD, Danah; CRAWFORD, Kate. CRITICAL QUESTIONS FOR BIG DATA. Information, Communication & Society, [s.l.], v. 15, n. 5, p.662-679, jun. 2012. Informa UK Limited. <http://dx.doi.org/10.1080/1369118x.2012.678878>.

CASTELLS, Manuel. A sociedade em rede. 11. ed. São Paulo: Paz e Terra, 2008.

_____. A galáxia da Internet. Rio de Janeiro: Jorge Zahar, 2003

DOMINGO, D., Quandt, T., Heinonen, A., Paulussen, S., Singer, J. & Vujnovic, M. (2008). Participatory Journalism Practices in the Media and Beyond: An International Comparative Study of Initiatives in Online Newspaper. Journalism Practice, 2(3), pp. 326- 342. doi: 10.1080/17512780802281065

FRANCISCATO, Carlos Eduardo. Delimitando um modelo de pesquisa aplicada em jornalismo. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO DA REGIÃO NORDESTE, 9., 2007, Salvador. Anais... . Salvador: Intercom, 2007. p. 1 - 15.

Disponível em:
<<https://ri.ufs.br/bitstream/riufs/1859/1/ModeloPesquisaAplicada.pdf>>. Acesso em: 1 mar. 2018.

GEHRKE, Marília. O uso de fontes documentais no jornalismo guiado por dados. Porto Alegre: Repositório Digital Lume, 2018. Disponível em: <<http://www.lume.ufrgs.br/bitstream/handle/10183/172614/001060430.pdf?sequence=1>>. Acesso em: 05 jun. 2018.

GIL, Antonio Carlos. Métodos e técnicas de pesquisa social. 6. ed. São Paulo: Atlas, 2008.

LIMA, Claudia Risso de Araujo et al. Departamento de Informática do SUS – DATASUS: A experiência de Disseminação de Informações em Saúde. In: Ministério da Saúde. A experiência brasileira em sistemas de informação em saúde. Brasília: Editora do Ministério da Saúde, 2009. p. 109-128. Disponível em: <https://www.measureevaluation.org/our-work/health-information-systems/health-information-system-strengthening-in-lac-region-2005-2010/experiencia_brasileira_sistemas_saude_volume1.pdf#page=109>. Acesso em: 02 mar. 2018.

LIMA JUNIOR, Walter Teixeira. Big Data, Jornalismo Computacional e Data Journalism: estrutura, pensamento e prática profissional na Web de dados. Estudos em Comunicação nº 12, 207-222. Lisboa: UBI, 2012.

MANOVICH, Lev. The language of new media. Cambridge: MIT Press, 2001.

MINISTÉRIO DA SAÚDE. Plano Diretor de Tecnologia da Informação 2014-2015. Brasília: Editora do Ministério da Saúde, 2015. Disponível em: <http://datasus.saude.gov.br/images/PDTI_2014-2015_Vs_Atualizada_jul2015.pdf>. Acesso em: 05 maio 2018.

MINISTÉRIO DA SAÚDE. Plano Diretor de Tecnologia da Informação 2017-2018. Brasília: Editora do Ministério da Saúde, 2018. Disponível em:

<http://datasus.saude.gov.br/images/PDTI_2014-2015_Vs_Atualizada_jul2015.pdf>. Acesso em: 05 maio 2018.

PEREIRA JÚNIOR, Luiz Costa. A apuração da notícia: métodos de investigação na imprensa. Petrópolis: Vozes, 2006.

SECRETÁRIA DE VIGILÂNCIA EM SAÚDE. Portaria nº 47, de 3 de maio de 2016. Brasília, Disponível em: <http://portalsinan.saude.gov.br/images/documentos/Portarias/Portaria_47_de_03052016.pdf>. Acesso em: 02 abr. 2018.

SILVER, Nate. What the fox knows. FiveThirtyEight, 17 mar. 2014. Disponível em: <http://fivethirtyeight.com/features/what-the-fox-knows/>. Acesso: 18 abr. 2018.

TRÄSEL, Marcelo Ruschel. Entrevistando planilhas: estudo das crenças e ethos de um grupo de profissionais de jornalismo guiado por dados no Brasil. 2014. 314 f. Tese (Doutorado) - Curso de Comunicação, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2014. Disponível em: <<http://tede2.pucrs.br/tede2/bitstream/tede/4590/1/461784.pdf>>. Acesso em: 1 mar. 2018.