



Universidade do Estado do Rio de Janeiro

Centro de Educação e Humanidades

Faculdade de Comunicação Social

Julia Barroso da Silveira

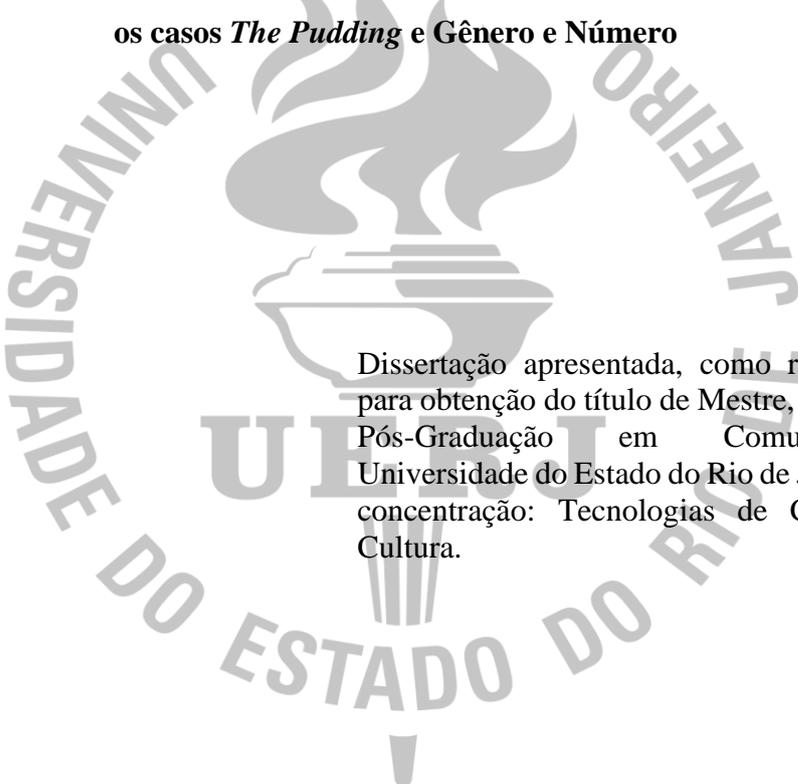
**Transparência como ferramenta na aprendizagem de jornalismo de dados:
os casos *The Pudding* e Gênero e Número**

Rio de Janeiro

2024

Julia Barroso da Silveira

**Transparência como ferramenta na aprendizagem de jornalismo de dados:
os casos *The Pudding* e Gênero e Número**



Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Comunicação, da Universidade do Estado do Rio de Janeiro. Área de concentração: Tecnologias de Comunicação e Cultura.

Orientadora: Leticia Cantarela Matheus

Rio de Janeiro

2024

CATALOGAÇÃO NA FONTE
UERJ / REDE SIRIUS / BIBLIOTECA CEH/A

S587 Silveira, Julia Barroso da.
Transparência como ferramenta na aprendizagem de jornalismo de dados:
os casos The Pudding e Gênero e Número/ Julia Barroso da Silveira. – 2024.
83 f.

Orientadora: Leticia Cantarela Matheus.
Dissertação (Mestrado) – Universidade do Estado do Rio de Janeiro.
Faculdade de Comunicação Social.

1. Comunicação – Teses. 2. Jornalismo – Teses. 3. Transparência – Teses.
I. Matheus, Leticia Cantarela. II. Universidade do Estado do Rio de Janeiro.
Faculdade de Comunicação Social. III. Título.

br

CDU 316.77

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação, desde que citada a fonte.

Assinatura

Data

Julia Barroso da Silveira

**Transparência como ferramenta na aprendizagem de jornalismo de dados:
os casos *The Pudding* e Gênero e Número**

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Comunicação, da Universidade do Estado do Rio de Janeiro. Área de concentração: Tecnologias de Comunicação e Cultura.

Aprovada em 28 de fevereiro de 2024.

Banca Examinadora:

Prof.^a Dr.^a. Leticia Cantarela Matheus (Orientadora)

Faculdade de Comunicação Social - UERJ

Prof.^a Dr.^a. Fátima Cristina Regis Martins de Oliveira

Faculdade de Comunicação Social - UERJ

Prof. Dr. Fábio Souza Vasconcellos

Escola Superior de Propaganda e Marketing

Prof.^a Dr.^a. Larissa de Moraes Ribeiro Mendes

Universidade Federal Fluminense

Rio de Janeiro

2024

AGRADECIMENTOS

Agradeço aos meus pais e aos meus avós pelo suporte e amor de sempre, e por serem minha torcida organizada. Ao meu irmão, Pedro, e aos meus amigos Lucas, Baltha e Luiza pela escuta paciente e pelas opiniões (quase sempre) ponderadas. À pessoa que me incentivou a pesquisar jornalismo de dados, Daniela Oliveira, que foi minha professora na graduação, orientadora do TCC e, recentemente, minha colega de turma.

Agradeço ao PPGCom-Uerj, aos professores que conheci nesses dois anos e aos amigos que fiz aqui – vivi intensamente as aulas, os grupos de pesquisa e os estágios docentes, e amei cada experiência. À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), cuja bolsa me permitiu focar na jornada acadêmica.

Agradeço à minha orientadora, Leticia Matheus, pela disponibilidade e atenção ao acompanhar a escrita da dissertação. Aos professores Fátima Regis e Fábio Vasconcellos, que participaram também da banca de qualificação e deram contribuições valiosas para a pesquisa, e à professora Larissa Mendes, que aceitou o convite para completar a banca de defesa.

RESUMO

SILVEIRA, Julia Barroso. *Transparência como ferramenta na aprendizagem de jornalismo de dados: os casos The Pudding e Gênero e Número*. 2024. 83 f. Dissertação (Mestrado em Comunicação) – Faculdade de Comunicação Social, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2024.

O jornalismo de dados, popularizado no Brasil a partir da Lei de Acesso à Informação (LAI), promulgada em 2011, ainda enfrenta dificuldades em sua implementação nas diretrizes curriculares da graduação em Jornalismo. Desde a falta de treinamento dos docentes, até a resistência dos alunos em se aproximar de técnicas computacionais e conceitos estatísticos e matemáticos, o jornalismo de dados é frequentemente visto sob um olhar tecnofílico ou tecnofóbico, que distanciam o trabalho com dados das práticas jornalísticas tidas como tradicionais. Como consequência, profissionais e estudantes que desejam se aproximar da área precisam recorrer ao autodidatismo ou a cursos extracurriculares. Pensando nesse contexto, esta dissertação parte da ideia de que o jornalismo de dados é, na verdade, parte da rotina atual dos jornalistas, devido ao uso cotidiano de dados na apuração, e que, por conta das mudanças cognitivas e produtivas a partir da popularização dos computadores pessoais, é possível utilizar ferramentas de transparência para aprender o jornalismo de dados. A transparência, um termo amplo que é frequentemente utilizado para falar do compartilhamento de dados públicos, serve de insumo para pautas, mas também funciona como caminho para a manutenção de processos democráticos e para aumentar a credibilidade de jornais, revelando para audiência os procedimentos de construção das reportagens, fontes de financiamento e estruturas organizacionais. Mais praticada por jornalistas que trabalham com dados, a transparência é separada nesta pesquisa em três categorias: transparência institucional, editorial-metodológica e tecnológica. Com o objetivo de sistematizar o aprendizado de jornalismo de dados a partir da transparência, a dissertação analisa o *The Pudding* e a *Gênero e Número*, dois jornais digitais que produzem reportagens com dados, de acordo com o Protocolo de Transparência no Jornalismo Guiado por Dados (Gehrke, 2021) e são destacados os parâmetros que correspondem à categoria da transparência editorial-metodológica, que tratam da produção das reportagens. Em seguida, são selecionadas uma matéria de cada jornal e, a partir delas, são evidenciadas as fases do trabalho com dados.

Palavras-chave: Jornalismo de dados. Transparência. *The Pudding*. *Gênero e Número*.

ABSTRACT

SILVEIRA, Julia Barroso. *Transparency as a tool in learning data journalism: The Pudding and Gênero e Número cases*. 2024. 83 f. Dissertação (Mestrado em Comunicação) – Faculdade de Comunicação Social, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2024.

Data journalism, popularized in Brazil with the Access to Information Law (*LAI*, in Portuguese), enacted in 2011, still faces difficulties in its implementation in the curricular guidelines for undergraduate journalism courses. From the lack of training of teachers, to the resistance of students in approaching computational techniques and statistical and mathematical concepts, data journalism is often seen from a technophilic or technophobic perspective, which distances working with data from journalistic practices seen as traditional. As a consequence, professionals and students who want to get closer to the area need to resort to self-education or extracurricular courses. With this context in mind, this dissertation starts from the idea that data journalism is, in fact, part of the current routine of journalists, due to the daily use of data in investigation, and that, due to the cognitive and productive changes following the popularization of personal computers, it is possible to use transparency tools to learn data journalism. Transparency, a broad term that is often used to talk about the sharing of public data, serves as input for journalism, but also works as a way to maintain democratic processes and increase the credibility of newspapers, revealing to the audience the processes of newsmaking, funding sources and organizational structures. Most practiced by journalists who work with data, transparency is separated in this research into three categories: institutional, editorial-methodological and technological transparency. With the aim of systematizing the learning of data journalism based on transparency, the dissertation analyzes *The Pudding* and *Gênero e Número*, two digital newspapers that produce reports with data, in accordance with the Transparency Protocol in Data-Driven Journalism (Gehrke, 2021) and the parameters that correspond to the category of editorial-methodological transparency are highlighted, which deal with the production of reports. Then, an article from each newspaper is selected and, from there, the phases of working with data are highlighted.

Keywords: Data journalism. Transparency. *The Pudding*. *Gênero e Número*.

LISTA DE FIGURAS

Figura 1 – Página inicial do <i>The Pudding</i>	34
Figura 2 – Página inicial da Gênero e Número.....	52
Figura 3 – Exemplo de gráfico da Gênero e Número	54
Figura 4 – A pirâmide invertida do jornalismo de dados.....	70

LISTA DE QUADROS

Quadro 1 – Tutoriais publicados no <i>The Pudding</i>	41
Quadro 2 – <i>Talks & presentations</i>	47
Quadro 3 – Parâmetros na transparência institucional.....	56
Quadro 4 – Parâmetros na transparência editorial-metodológica.....	57
Quadro 5 – Parâmetros na transparência editorial-metodológica na matéria da Gênero e Número.....	66
Quadro 6 – Parâmetros na transparência editorial-metodológica na matéria do <i>Pudding</i> ..	67
Quadro 7 – Comparando as etapas do jornalismo de dados.....	69

SUMÁRIO

INTRODUÇÃO.....	9
1 A TRANSPARÊNCIA COMO ALIADA DO JORNALISMO.....	18
1.1 Iniciativas brasileiras de democratização da informação.....	23
1.2 Protocolo de Transparência no Jornalismo Guiado por Dados	26
2 O JORNALISMO DE DADOS E AS TRANSPARÊNCIAS.....	30
2.1 <i>The Pudding</i>.....	32
2.2 Gênero e Número.....	50
2.3 Uma análise comparativa.....	55
3 O APRENDIZADO DO JORNALISMO DE DADOS A PARTIR DA TRANSPARÊNCIA EDITORIAL-METODOLÓGICA.....	60
3.1 A transparência das etapas do trabalho com dados.....	64
CONCLUSÃO.....	76
REFERÊNCIAS.....	79

INTRODUÇÃO

Com um *ethos* que combina elementos da cibercultura, da ética *hacker* e da cultura profissional jornalística (Träsel, 2014b), o jornalismo de dados tem se popularizado no Brasil a partir da Lei de Acesso à Informação (LAI), aprovada em 2011, e da adesão nacional à Declaração de Governo Aberto (*Open Government Partnership*, ou *OGP*), também em 2011. Considerando o aumento da disponibilidade de dados públicos e a presença de ferramentas digitais no cotidiano das redações, se tornou necessário rever técnicas de apuração, produção e difusão de notícias (Araújo, 2016).

Enquanto a abordagem tecnofóbica posiciona o jornalismo de dados como uma colonização indesejada da informática (Träsel, 2014b), há, em contraponto, um imaginário tecnofílico comum desde o século XIX de que a mecanização da produção e o progresso da racionalidade técnica simbolizam melhorias nas condições de vida e proporcionam avanços sociais (Régis, 2012). Como defendido nesta dissertação, o jornalismo de dados não se encontra em nenhum desses extremos – na verdade, ele possui processos produtivos similares aos do jornalismo visto como tradicional e segue o objetivo de informar.

Os dados podem ser tanto a fonte quanto as ferramentas em uma notícia, e ter conhecimento sobre busca, limpeza e visualização de dados impacta diretamente no exercício da reportagem (Gray; Bounegru; Chambers, 2012), mas não há uma transformação de fato na prática jornalística. Cabe, porém, uma diferenciação entre o jornalismo **de** dados, em que os dados se mostram fundamentais para a pauta e motivam a construção da reportagem, e o jornalismo **com** dados, que utilizam dados de forma ilustrativa, como proposto por Mancini e Vasconcellos (2016). A dissertação tem como foco o jornalismo **de** dados, que seria a categoria a exigir maior aprofundamento técnico do profissional, por exigir, também, análises mais profundas das bases de dados.

Nesse contexto, além de ser importante que jornalistas sejam capacitados para trabalhar com dados devido ao grande número de informação disponível, utilizar bases de dados como insumo tem, também, vantagens produtivas. As bases de dados geram fontes de valor-notícia por tempo indefinido (Träsel, 2014b) e se tornam de interesse público (Träsel, 2018), porque podem ser reutilizadas a partir de abordagens diferentes, motivadas por objetivos e recortes distintos. Análises de dados podem, também, render recortes mais precisos para a matéria, ao mesmo tempo que seguem técnicas jornalísticas tradicionais na forma de construir as reportagens guiadas por tais análises, como ressaltam Craig, Ketterer e Yousuf (2017).

Devido ao seu histórico – as origens no jornalismo de precisão, a relação com os ideais da cibercultura¹ e sua popularização a partir da disponibilidade de dados governamentais –, o jornalismo de dados se conecta à transparência como valor jornalístico. Uma ideia ainda concentrada de forma individual nos profissionais, em vez de uma prática das redações, como será discutido no primeiro capítulo, a transparência tem tido mais destaque com os jornalistas que trabalham com dados.

Esta pesquisa parte da abordagem de Christofolletti, Becker e Paul (2021) da transparência como conceito amplo aplicado a governos e a corporações. O arcabouço teórico também inclui a visão de Christofolletti (2022) e Gehrke (2020) da transparência jornalística como forma de enfrentar a atual crise de credibilidade, porque revelar procedimentos produtivos, possíveis vieses e fontes é um caminho para reforçar a conexão com as audiências.

A transparência é especialmente importante no jornalismo de dados, que une o mito da objetividade jornalística com o mito de que “os números não mentem”. Considerando que o ensino do jornalismo não envolve conhecimentos matemáticos, trabalhar com números pode, sim, resultar em análises erradas mesmo que não haja a intenção de manipulação dos dados. O letramento matemático durante o aprendizado de jornalismo de dados evita falhas comuns, como valores de média sendo utilizados mesmo quando não são representativos da amostra e a leitura de correlações como causalidades.

Além disso, a escolha cuidadosa da visualização pode evitar interpretações enganosas – o famoso gráfico de pizza não é indicado para comparações, por exemplo, devendo ser substituído por gráficos de barra. Nesses casos, novamente a objetividade se revela um mito. Por isso, informar aos leitores os passos que levaram às conclusões é importante. Indicar a fonte de dados, ou oferecer acesso às bases utilizadas, significa abrir espaço para investigações secundárias, permitindo que o público possa fazer suas próprias análises ou que outros jornalistas encontrem pautas partindo do mesmo material.

Christofolletti, Becker e Paul (2021) registraram as diferentes preocupações com a transparência nos países latino-americanos, em que apenas o código de ética argentino a destaca como um valor prioritário, em contraponto aos países europeus, que incluíram práticas de transparência em sua deontologia, mas explicam que essas práticas podem ser assimiladas de

¹ Apesar da ideia de cibercultura estar datada, ainda conectada à uma visão otimista do que seria construído a partir dos desenvolvimentos tecnológicos, o termo é utilizado nesta pesquisa porque é justamente esse ideal que se relaciona às propostas de autodidatismo e inteligência coletiva atreladas ao imaginário dos jornalistas de dados. A substituição do termo por cultura digital removeria esses sentidos necessários ao contexto.

forma individual, sem a explicitação em documentos coletivos. Assim como a transparência ainda não é a norma no campo profissional, o jornalismo de dados não foi totalmente assimilado na graduação.

Com práticas pedagógicas diluídas e esparsas, o jornalismo de dados é uma área ainda pouco lecionada no Brasil, (Ventura; Paulino, 2020). Além da falta de formalização das Diretrizes Curriculares na graduação, como registrado por Ventura (2021), Berret e Phillips (2016) afirmam que alguns professores não tratam o jornalismo de dados como vital para a carreira e que, geralmente, essa disciplina é oferecida de forma eletiva, atraindo apenas alguns alunos. Como consequência, muitos dependem de instruções fora da educação formal.

Para Araújo (2016), os profissionais de imprensa ainda utilizam a *internet* em nível de usuários, não aproveitando totalmente o potencial de obter informações para pautas inéditas por meio de ferramentas para a mineração de dados, por exemplo. Mas aprender jornalismo de dados tem sido do interesse de estudantes e profissionais da área. Um curso sobre jornalismo e visualização de dados, gratuito e virtual, do Centro *Knight* de Jornalismo nas Américas, da Universidade do Texas em Austin, foi lançado em novembro de 2019 e o Brasil foi o país com o maior número de alunos inscritos – 2.297 pessoas –, seguido pelos Estados Unidos – 1.756 pessoas – e pelo México – 775 pessoas (Gehrke, 2020).

O autodidatismo e o aprendizado por meio de cursos extracurriculares têm sido o relato comum na área, visto que o jornalismo de dados ainda não é parte formalizada da graduação em Jornalismo. Por meio dos currículos, das atividades acadêmicas e das propostas pedagógicas, as universidades têm a oportunidade de impactar a produção midiática a longo prazo, conforme as novas gerações de estudantes chegarem ao mercado de trabalho (Fittipaldi, 2019), mas, especificamente no jornalismo de dados, tem sido necessário pensar em recursos não formalizados para o aprendizado.

Partindo de pesquisa documental e coleta de informações a partir de um questionário, Träsel (2019) identificou a oferta de ensino de jornalismo de dados em 32 Instituições de Ensino Superior (IES) – 15 públicas, 17 privadas e uma estrangeira, a já mencionada Universidade do Texas com o curso do Centro *Knight* –, aproximadamente 10% das IES que possuem cursos de Jornalismo, e sete instituições de outras naturezas, como a Associação Brasileira de Jornalismo Investigativo (Abraji) e o *Google News Lab*. Entre as IES nacionais, o autor mapeou, ao todo, 52 disciplinas ou cursos de extensão, sendo 41 de cursos de graduação. Apesar de técnicas relacionadas ao jornalismo de dados terem sua presença mais antiga registrada em 2006,

segundo o levantamento, Träsel (2019) identificou que 2014 foi o ponto de inflexão na oferta de treinamentos.

Da mesma forma, Bazzo (2019) aborda em sua dissertação a falta de pesquisa aprofundada sobre a emergência do jornalismo de dados no ensino superior de Jornalismo no Brasil ou discussão válida de métodos executáveis de ensino. Com uma análise qualitativa dos Projetos Pedagógicos de Curso (PPCs) de universidades públicas, a autora observou que o jornalismo de dados é visto como método, técnica ou especialidade, e não como uma área ou subcampo. Dentre os 32 PPCs registrados, oito ofertam a disciplina Jornalismo de Dados e seis abordam as práticas do trabalho com dados em outras disciplinas.

Isso se conecta com a presença ainda tímida do jornalismo de dados nas redações brasileiras. Vasconcellos e Mancini (2016) observaram que as reportagens com dados não eram amplamente produzidas no Brasil ao analisarem a versão digital de 45 veículos jornalísticos. Os autores registraram que menos de 30% das reportagens, em uma amostra de 2.296, praticavam o jornalismo de dados que inclui a criação, estruturação e visualização de dados.

Anos depois, Santos (2019) encontrou 52 organizações que produzem matérias baseadas em dados. Partindo da ideia de jornalismo de dados como prática que se relaciona com valores do mundo *hacker*, o autor coletou informações a partir de questionários e pesquisas no *Twitter* e concluiu que o jornalismo de dados ainda não está consolidado no país, considerando que as organizações registradas se concentram em São Paulo (23) e no Rio de Janeiro (13). A presença da mídia tradicional também teve impacto na conclusão, correspondendo a 19 das 52 organizações analisadas, enquanto as 27 nativas digitais não se restringem a iniciativas independentes, mas incluem, também, nomes como UOL e G1, que pertencem a grandes grupos midiáticos.

De acordo com Mielniczuk e Träsel (2017), os estudos acadêmicos voltados especificamente ao ensino do jornalismo de dados ainda são escassos. Diante deste cenário, a pesquisa parte da hipótese de que os recursos de transparência disponibilizados por veículos jornalísticos podem ser sistematizados como ferramentas de aprendizado no jornalismo de dados. Como transparência é um termo amplo, que engloba desde a descrição do processo produtivo até as fontes de financiamento, esta dissertação considera três categorias descritas no primeiro capítulo para facilitar o recorte didático – a transparência institucional, a editorial-metodológica e a tecnológica.

Com o objetivo de pensar nessa aplicação dos recursos de transparência, especificamente a editorial-metodológica, foram analisados o *The Pudding* e a *Gênero e*

Número, dois nativos digitais especializados em jornalismo de dados, por meio do Protocolo de Transparência no Jornalismo Guiado por Dados, desenvolvido em Gehrke (2021) para a análise de veículos e produtos jornalísticos publicados. Na formulação do Protocolo, está previsto que alguns dos 15 parâmetros estabelecidos podem ter maior ou menor peso de acordo com o objetivo da análise.

Como o foco desta dissertação é o aprendizado do jornalismo de dados, ou seja, das etapas do trabalho com dados para a produção de reportagens, os 15 parâmetros foram comparados com as categorias de transparência criadas na pesquisa e aqueles que se alinhavam com a transparência editorial-metodológica foram, posteriormente, ferramenta de análise de matérias selecionadas para testar a hipótese de que é viável sistematizar os recursos como ferramenta de aprendizado.

A dissertação busca, também, colaborar com as pesquisas brasileiras sobre jornalismo de dados, visto que, de acordo com o mapeamento feito por Bazzo, Martins e Barbosa (2020), foram publicados, a partir de 2007, oito artigos, sete dissertações e seis teses relacionados a jornalismo de dados, com registros até 2018, 2017 e 2016, respectivamente. A análise foi feita considerando periódicos que obtiveram Qualis A1, A2 e B1 no quadriênio 2013-2016 e produções registradas na Biblioteca Digital de Teses e Dissertações (BDTD) e no Catálogo de Teses e Dissertações da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), um recorte justificado pelos autores por representar, de forma sociológica e não matemática, a produção mais influente sobre o campo.

Acrescentando a esse mapeamento e observando as produções publicadas no Catálogo de Teses e Dissertações da Capes, foram registradas mais 11 dissertações e seis teses. A região Sul do país se destaca nas pesquisas que envolvem jornalismo de dados, acumulando oito das 17 produções, seguida pela região Nordeste, com quatro, Centro-Oeste, com três, e Sudeste, com duas. Na análise realizada por Bazzo, Martins e Barbosa (2020), as dissertações e teses publicadas tinham origem principalmente nas regiões Sudeste – seis – e Sul – cinco –, enquanto as outras duas eram do Nordeste.

Quatro das dissertações têm foco na visualização de dados (Freitas, 2018; Caires, 2019; Huf, 2021; Lopes, 2022), e Bazzo (2019) aborda a incorporação do jornalismo de dados na graduação de Jornalismo no Brasil. A pesquisa de Gehrke (2018) trata do uso de fontes documentais, enquanto Gonçalves (2019) centraliza a LAI como fonte e ferramenta de apuração no jornalismo de dados nacional. A pesquisa de Ribeiro (2022) partiu da ótica dos jornalistas de dados para compreender a transparência nos processos produtivos, considerada uma

estratégia para aumentar a confiabilidade e como norma orientadora do trabalho jornalístico. A autora conclui que a transparência é estabelecida como prática, valor, conduta a ser implementada e contraste.

Mastrella (2019), por meio da análise de discurso de profissionais e organizações que praticam e promovem o jornalismo de dados, observou que a legitimação da área engloba a incorporação de ideologias de outras áreas, como a Ciência da Computação, e a reafirmação de valores canônicos do jornalismo, alterando, ao mesmo tempo, convenções tradicionais no ainda restrito meio profissional.

Analisando a produção de jornalismo de dados e a transparência com um recorte de produções cearenses, O Povo e Sistema Verdes Mares, criados em 2018, Lavor (2021) observou a falta de práticas de transparência nos núcleos de dados. Além disso, a autora apontou para a simples reprodução de bases de dados em detrimento de coleta própria, ou seja, não há a independência do trabalho jornalístico nessas organizações.

Também observando o jornalismo de dados a partir de organizações específicas, Silva (2022) analisou o funcionamento do consórcio de veículos de imprensa criado durante a pandemia que tinha o objetivo de garantir a apuração diária dos dados de casos e mortes da Covid-19, concluindo que a colaboração trouxe contribuições práticas para o jornalismo de dados, o jornalismo digital e à profissão de forma geral.

Passando para as teses encontradas, nota-se a clara predominância de defesas ocorridas em 2021 – cinco das seis totais. Entre as temáticas a transparência tem destaque. A análise de Brenol (2019) ressalta a transparência governamental e o uso de dados públicos no jornalismo, que reforça o papel social do campo e a ideia de interesse público, impactando em decisões governamentais e fortalecendo o exercício profissional apesar da era dos algoritmos e do excesso de informação.

Gehrke (2021) abordou a transparência como valor ético e recurso discursivo de legitimação do jornalismo e propôs um Protocolo de Transparência no Jornalismo Guiado por Dados, que é amplamente utilizado nesta pesquisa como ferramenta de análise. Construído a partir da análise de elementos de transparência observados na cobertura midiática da Covid-19, a tese separa a transparência nos níveis organizacional e operacional e parte de informações básicas, como a assinatura das matérias e a disponibilização de contatos do jornal, e se desdobra até elementos mais específicos, como graus de incerteza das informações consultadas.

Comparando dois projetos nacionais, a Agência Pública e a Gênero e Número, que será mencionada nesta pesquisa, com uma iniciativa inglesa, *The Bureau of Investigative*

Journalism, Cláudia Rodrigues (2021) retrata em sua tese as diferenças das rotinas produtivas e conclui, a partir de um *corpus* de 135 reportagens, que os jornais possuem, respectivamente, alta, média e baixa transparência.

Pensando nas narrativas construídas com dados, Kelly Rodrigues (2021) analisou quatro jornais latinoamericanos que possuem equipes dedicadas ao trabalho com dados e observou que a construção do discurso nas reportagens de dados influenciam caracterizações e projeções de personagens retratados. A autora acrescenta que, dependendo da contextualização das informações, a importância de dados específicos pode ser redimensionada.

Com o objetivo de analisar a carreira dos jornalistas de dados brasileiros, Lima (2021) realizou uma pesquisa exploratória de abordagem qualitativa, com entrevistas abertas em profundidade, observação participante e análise das histórias de vida, sendo possível, assim, compreender a prática de forma geral a partir dos percursos individuais. A autora encontrou como ponto comum entre os entrevistados a visão do jornalismo de dados como ideário de jornalismo, e, analisando as carreiras em arcos temporais, ressaltou que a escolha profissional é baseada principalmente em gostos pessoais em relação às práticas da área, divergindo do que ocorre no jornalismo de forma geral, em que mitos e imagens da profissão têm grande influência.

Tendo em mente o cenário acadêmico, o primeiro capítulo desta dissertação consiste em uma exploração da transparência, considerando o contexto da Lei de Acesso à Informação (LAI) e da transparência governamental como fontes de pautas e ferramentas de manutenção dos processos democráticos (Christofoletti, 2022). A transparência também se mostra um importante recurso no enfrentamento da crise de credibilidade no jornalismo e das instituições, como trazido por Mielniczuk e Träsel (2017), Zamith (2019) e Gehrke (2020), mas ainda há entraves para a implementação de suas práticas nas redações, como questões mercadológicas e a ideia de que haveria um aumento da carga de trabalho para os jornalistas (Christofoletti; Becker; Paul, 2021; Christofoletti, 2022).

Ainda no primeiro capítulo são estabelecidas três categorias da transparência no jornalismo, por esse ser um termo muito abrangente e com diversas aplicações. Apesar de não-canônicas, as categorias – transparência institucional, editorial-metodológica e tecnológica – têm o objetivo de facilitar o recorte didático das técnicas a serem discutidas ao longo da pesquisa. Em seguida, são observadas algumas iniciativas brasileiras que colaboram com a democratização da informação, a fim de captar uma variedade de interações entre a

transparência e o jornalismo por meio de portais que têm a transparência como insumo e, ao mesmo tempo, praticam e incentivam a transparência.

Um subcapítulo é dedicado ao Protocolo de Transparência no Jornalismo Guiado por Dados (Gehrke, 2021) e seus 15 parâmetros de transparência, que, posteriormente, foi utilizado para analisar o *The Pudding*, a Gênero e Número, e as reportagens que exemplificam a sistematização de recursos da transparência editorial-metodológica como ferramenta de aprendizado do jornalismo de dados. Esse Protocolo funciona para verificar se os veículos e as reportagens analisados são transparentes ou não, de acordo com os parâmetros.

No segundo capítulo, a transparência é pensada especificamente no cenário do jornalismo de dados, que, apesar de não ser uma área completamente afastada do jornalismo tido como tradicional, como defendido na pesquisa, envolve técnicas computacionais que ainda são vistas como novidade por alguns profissionais. São analisados o *The Pudding* e a Gênero e Número e, na investigação de seus recursos de transparência, é possível notar que são jornais muito diferentes – em termos de como praticam a transparência, escolha de pautas e formas de trabalhar o jornalismo de dados. Após as análises individuais, é realizada uma análise comparativa com base no Protocolo, para confirmar que os objetos selecionados estão alinhados com o objetivo da pesquisa.

O terceiro e último capítulo é dedicado a pensar o aprendizado do jornalismo de dados a partir da transparência editorial-metodológica. Partindo de uma revisão bibliográfica sobre o atual cenário do ensino do jornalismo de dados no Brasil, que enfrenta dificuldades em sua implementação na graduação (Ventura, 2021), observamos, também, como as transformações tecnológicas impactam a produção e a cognição. Em um contexto em que a excessiva oferta de informações muda as formas de consumo (Alonso, 2018), são desenvolvidas competências cognitivas como o autodidatismo e a criatividade (Régis, 2022), que se relacionam profundamente com a ética *hacker* que ainda faz parte do imaginário dos jornalistas de dados (Träsel, 2014b).

Aprender jornalismo de dados é, na verdade, aprender processos cognitivos para lidar com etapas de produção a partir de bases de dados (Mielniczuk; Träsel, 2017), desenvolver pensamento crítico (Gehrke, 2021) e a capacidade de inovar (Salaverría, 2011). Considerando a transparência editorial-metodológica e a análise do *The Pudding* e da Gênero e Número, o capítulo termina com a sistematização dos parâmetros desta categoria de transparência como ferramenta de aprendizado, mostrando como produtos muito diferentes ainda assim podem ser aproveitados tanto por profissionais autodidatas como em sala de aula, desde que cumpram os

critérios da transparência – assim, é possível encontrar na reportagem todas as etapas do trabalho de dados que fundamentam o jornalismo de dados.

1 A TRANSPARÊNCIA COMO ALIADA DO JORNALISMO

A transparência, de acordo com Christofolletti, Becker e Paul (2021), é um conceito presente no debate público desde a década de 1980 mas que se intensificou nos anos 2000, e transpassa tanto a esfera governamental quanto o ambiente corporativo numa tentativa de retomar a credibilidade das instituições. No Brasil, a transparência governamental possui dois grandes símbolos: o Portal da Transparência, lançado em 2004 pela Controladoria-Geral da União (CGU), e a Lei de Acesso à Informação (LAI), promulgada em 2011 e em exercício desde 2012. Enquanto o primeiro é relativo apenas ao poder executivo federal e representa a transparência ativa, quando os dados são oferecidos por iniciativa de quem os possui, a LAI se aplica aos três poderes e prevê também a transparência passiva, ou seja, permite a solicitação de dados que não estão abertos para o público.

Tornando o sigilo a exceção, a LAI significou um aumento de dados de interesse público disponíveis, impactando diretamente o jornalismo. Os profissionais se viram diante de conjuntos massivos de dados, colaborando para a expansão do jornalismo de dados no Brasil (Gehrke, 2020). Além de colaborar com a manutenção dos processos democráticos (Bazzo; Martins; Barbosa, 2020; Christofolletti, 2022), a transparência governamental é uma fonte de pautas inéditas. Observando a atuação do jornalismo em trazer a público as informações mais relevantes obtidas por meio da LAI, pode-se considerar contraditório que o jornalismo em si não pratique a transparência que exige de instâncias governamentais ou outros setores privados, como ressalta Christofolletti (2022).

A transparência tem relação com a democracia, com o aumento da participação social e com o fortalecimento da cidadania, sendo geralmente associada ao acompanhamento de políticas públicas e a decisões de pessoas que ocupam cargos governamentais (Christofolletti, 2022). Nesta pesquisa, porém, é feito um recorte das relações da transparência com o jornalismo, em suas diversas camadas e nuances, visto que este conceito é amplo e abarca diferentes práticas, que serão discutidas na dissertação.

Como elaborado por Gehrke (2021), a transparência é a abertura e a comunicação de condutas e procedimentos, funcionando de forma análoga à ponta de um *iceberg*, ou seja, o que o jornalista ou o veículo decidem tornar visível. Transparência no jornalismo, de acordo com Christofolletti (2022), é mostrar como é a rotina de trabalho, o que está por trás da informação apresentada, apontar os valores, a missão, o compromisso ético e a política editorial da empresa, revelando, também, as operações financeiras por trás do veículo de comunicação.

A relação da transparência com a objetividade é popular entre teóricos da área, tendo sido mencionada por Träsel (2019) e por Gehrke (2020; 2021), que, ao afirmar que objetividade é sobre revelar métodos e procedimentos e ser honesto com a audiência, aproxima o jornalismo da ciência. Essa conexão pode ser atribuída ao Jornalismo de Precisão, criado na década de 1970 por Philip Meyer, que propôs aplicar técnicas quantitativas das ciências sociais às rotinas de reportagem, sugerindo que este seria o caminho para um possível retorno da objetividade às redações (Träsel, 2014b). Descendente do Jornalismo de Precisão, o jornalismo de dados tem, de certa forma, ideais parecidos.

Porém, como posto em Schudson (1978), a objetividade, antes vista como o antídoto para vieses, passou à posição de um viés em si, porque o jornalismo tido como objetivo reproduzia uma visão da realidade social que não questionava as estruturas de poder. É importante ressaltar que esse debate da objetividade não se restringe ao jornalismo. Essa é uma discussão tradicional no campo da teoria do conhecimento, como destaca Moretzsohn (2001), considerando que, como a apreensão do mundo depende da subjetividade, não é possível formular um conhecimento objetivo.

Moretzsohn (2001) explica que a crítica à objetividade busca indicar que o jornalismo não é o discurso *da* realidade e sim um discurso *sobre* a realidade. Assim como parece distante a superação da ideia de verdade universal, o imaginário de um jornalismo objetivo também não parece próximo de seu fim. A transparência, então, pode ajudar a desmistificar a ideia da objetividade jornalística. Cada escolha, desde a seleção de pautas até a edição final de reportagens, é subjetiva, e evidenciar essas escolhas aponta os caminhos tomados pelo jornalista responsável.

Observando o *Digital News Report 2023*, do Reuters Institute (2023), a confiança nas notícias no Brasil está em queda: de 62% em 2015, o índice chegou a 48% em 2022 e, após o ano de eleições e o período de pandemia da Covid-19, alcançou 43%, se mantendo na 14ª posição entre os 46 países analisados. Essa crise é provocada por motivos multifatoriais, como ressaltam Mielniczuk e Träsel (2017), e algumas das razões listadas são a crise da economia mundial, a quebra com o modelo verticalizado de circulação de informações e o protagonismo da audiência, que permite questionar o jornalismo das redações tradicionais, sua credibilidade e suas conexões com os setores político e econômico.

Na pesquisa realizada por Gehrke (2020) com jornalistas de dados brasileiros, as principais razões apontadas para a adoção da transparência como um valor importante na prática jornalística são a necessidade de aumentar a credibilidade no jornalismo e combater a

desinformação. A autora também observou que, para os profissionais entrevistados, a transparência deve ser uma prática diária e não um recurso que só aparece em reportagens especiais.

As práticas de transparência, um assunto mais recorrente entre os jornalistas de dados, têm se mostrado benéficas para o jornalismo, de forma geral, como forma de enfrentamento da crise de credibilidade, como tem sido debatido no campo (Christofoletti, 2022; Gehrke, 2020; Zamith, 2019). Apesar de ser relativamente comum revelar fontes entrevistadas, quando não há entraves éticos para tal, explicitar métodos ainda não é amplamente aceito nas organizações jornalísticas. Mesmo com as mudanças de produção e no fluxo de circulação, poucos profissionais e empresas estão dispostos a correr riscos com suas propriedades intelectuais.

Tornar uma empresa jornalística auditável é reforçar os ideais éticos do meio, mas, como mencionam Christofoletti, Becker e Paul (2021), ainda há entraves mercadológicos, classistas, culturais e estruturais para implementar a transparência, e esse distanciamento da transparência também é refletido nos códigos deontológicos da profissão. Analisando a presença de valores como a transparência e a *accountability* nos códigos de ética nos 20 maiores mercados consumidores de notícias na América Latina, nota-se que, apesar da disposição discursiva em torno da transparência, não há incentivo para que os jornalistas descrevam os procedimentos de construção das matérias (Christofoletti; Becker; Paul, 2021). Os autores ainda ressaltam que, da mesma maneira que acontece com outros valores deontológicos, a transparência pode apenas habitar o imaginário valorativo e não ser traduzida em prática.

Em contraponto às estruturas organizacionais tradicionais, os nativos digitais têm sido mais abertos com as audiências em relação aos processos jornalísticos e, inclusive, em relação às fontes de financiamento, mas a transparência tende a ser acionada como discurso de legitimação (Gehrke, 2021). Temos, então, o seguinte panorama: a transparência tem sido utilizada como discurso, por ser um valor que aumenta a credibilidade dos jornais diante das audiências, mas não necessariamente há uma tradução desse discurso para a rotina de produção.

Como defendido por Fittipaldi (2019), discutir princípios em todos os níveis das redações é mais importante do que colocá-los no papel, e, nesse sentido, cabe às organizações noticiosas a realização de treinamentos internos direcionados a questões substanciais, havendo um esforço organizacional e não só o empenho pessoal de cada profissional. A transparência, de forma geral, ainda está concentrada em esforços superficiais que estão distantes de servirem como prestação de contas (Christofoletti; Becker; Paul, 2021). Além disso, o movimento de aplicar a transparência nos produtos tem partido da dimensão individual e não organizacional,

visto que não há incentivo das empresas e que faltam diretrizes objetivas nos manuais de redação, mantendo a transparência em um nível abstrato ou, no máximo, restrito à indicação de fontes utilizadas.

Um dos entraves para a ampla adoção da transparência é a pouca aderência que este conceito tem no setor privado. Mas, como observado por Christofolletti (2022), os departamentos de *compliance* têm surgido e, nos últimos 20 anos, a transparência faz parte de estratégias organizacionais que buscam convencer o público de que as empresas têm compromisso social. Outros pontos que seriam barreiras para a adoção de práticas de transparência são o aumento da carga de trabalho e a limitação de espaço ou tempo para a inclusão das informações processuais, mas, como destacou Fernanda Campagnucci, da *Open Knowledge* Brasil, há soluções tecnológicas para essas questões, como o uso de repositórios públicos para armazenar códigos e dados, possibilitando a reutilização de matéria-prima e aumentando as possibilidades de colaboração (Christofolletti, 2022).

Apesar dos pontos favoráveis à adoção de práticas de transparência levantados até aqui, há momentos em que a transparência não é recomendada. Christofolletti (2022) lista ocasiões em que a segurança – dos profissionais, da informação e das pessoas envolvidas ou retratadas na matéria – esteja em jogo: momentos em que a fonte possa ser colocada em perigo; quando a equipe está em risco; quando a investigação ainda estiver em curso e revelar informações possa interromper ou interditar os processos jornalísticos.

Nesta pesquisa, a fim de comparar as práticas de transparência de dois jornais digitais e pensar possíveis aplicações para o aprendizado do jornalismo de dados, serão utilizadas as categorias de transparência institucional, editorial-metodológica e tecnológica. Estas categorias não são canônicas do campo, mas serão elaboradas nesta dissertação por cumprirem funções diferentes e com o objetivo de facilitar a compreensão das diversas camadas desta prática. É fundamental reforçar que essa divisão visa fins didáticos, visto que a definição de transparência em que é baseada esta pesquisa é a de Christofolletti, Becker e Paul (2021): uma abertura, que pode partir de pessoas ou instituições, de processos, métodos e decisões a respeito de documentos, dados e registros.

Partindo desse conceito amplo, é possível definir as divisões sugeridas anteriormente. A transparência institucional, que se posiciona anteriormente às outras, é o compartilhamento de particularidades da instituição ou da empresa. Podemos considerá-la mais abrangente que as outras expressões da transparência porque inclui informações como missão, visão e valores da organização, que são comumente declaradas, e revela, também, as formas de financiamento.

Em seguida, temos a transparência editorial-metodológica. Esta é a que normalmente é mencionada quando se pensa em transparência no jornalismo, definida por Gehrke (2020) como a abertura de métodos e procedimentos de reportagem, partindo da obtenção da informação, passando por etapas de verificação e acompanhando o trajeto percorrido até a publicação. Para esta categoria, além da declaração dos processos metodológicos na construção de cada matéria, considera-se que há, nas publicações, a indicação de quem é o responsável, e, no caso de revisões, correções ou atualizações, isso é indicado de forma clara no mesmo conteúdo que foi editado.

Aqui, é necessário interromper as definições didáticas de diferentes faces da transparência para falar sobre o uso da expressão “metodologia” para se referir a procedimentos jornalísticos, considerando que o jornalismo não é baseado no método científico. A seção ao fim das matérias de dados dedicada a explicar os procedimentos da reportagem, como acontece na *Gênero e Número* e no *The Pudding*, sites de jornalismo de dados a serem analisados no próximo capítulo, leva esse nome, por explicar os métodos de obtenção, apuração e análise dos números apresentados.

Novamente é possível notar a presença de ideais do jornalismo de precisão no trabalho com dados, que é visto como mais próximo de uma ciência por utilizar números e dados supostamente objetivos. O jornalismo de precisão surgiu com o ideal de uma imprensa que fosse capaz de mostrar a realidade social livre de preconceitos, vieses e ideologias, utilizando o pensamento tecnológico nas rotinas produtivas (Träsel, 2014b), mas sem considerar que mesmo a ciência e as tecnologias possuem vieses e a neutralidade não é alcançável. O vínculo que existe entre a transparência e a ciência, porém, é a possibilidade da avaliação dos processos realizados e da reprodutibilidade (Christofoletti, 2022).

Retornando para as categorias propostas da transparência, nota-se que a principal delas é a editorial-metodológica, visto que falar sobre métodos é uma das formas mais eficazes do exercício da transparência (Christofoletti, 2022). Por ser explicitada a cada publicação, ajudar na superação do mito da objetividade e proporcionar o aumento da credibilidade, é a que mais se aproxima do leitor e, considerando que é uma expressão do cotidiano do jornal, é a mais relevante para o aprendizado. A última forma de transparência é a tecnológica, que pode ser definida como a explicação detalhada do que está por trás da construção das matérias, incluindo a menção das ferramentas, aplicações utilizadas e até caminhos para aprender a utilizar tais ferramentas.

A conexão entre o jornalismo e a ética *hacker* destacada por Träsel (2014b), ou seja, a disposição em compartilhar conhecimentos e a facilidade de apropriação de tecnologias nas práticas cotidianas, se conecta com a transparência tecnológica e é diretamente refletida na formação de novos profissionais da área, visto que, para que mais jornalistas saibam trabalhar com números e visualizações, é fundamental a colaboração e a democratização de ferramentas que são externas ao ensino tradicional do jornalismo. Por isso, este capítulo vai oferecer um panorama de recursos que buscam democratizar a informação, fomentar a colaboração entre usuários e propagar diversas manifestações da transparência, contribuindo para práticas jornalísticas.

O capítulo também inclui uma seção dedicada ao Protocolo de Transparência no Jornalismo Guiado por Dados, desenvolvido por Gehrke (2021) para a análise de produtos jornalísticos publicados de acordo com 15 parâmetros de transparência, incluindo as esferas organizacional e operacional. Esse Protocolo reúne de forma sistematizada as diferentes práticas mencionadas anteriormente e detalha critérios que ilustram o impacto da transparência para o público e para outros profissionais, colaborando com as análises a serem feitas no próximo capítulo.

1.1 Iniciativas brasileiras de democratização da informação

Como observado, a transparência é um termo amplo e que engloba diversas práticas. Focando nas relações entre a transparência e o jornalismo, temos a valorização da transparência governamental, que permite a fiscalização e o monitoramento do poder público, e a transparência das organizações jornalísticas em si, que, ao revelarem seus processos de apuração, fontes de financiamento e estruturas hierárquicas, têm a chance de aumentar sua credibilidade e criar novas conexões com as audiências. Como os autores vistos até aqui apontam, a transparência cobrada pelos jornalistas não necessariamente é praticada por eles, principalmente quando pensamos na transparência como prática organizacional e não individual.

Mesmo que ainda concentrada no campo do discurso, de forma geral, a transparência tem lugar de destaque em organizações que advogam pela manutenção dos processos democráticos. Nesta seção, vamos explorar diversas iniciativas brasileiras que, por meio da transparência, colaboram com a democratização da informação e com o jornalismo, ao servirem

de fonte ou produzirem por conta própria reportagens que partem da transparência governamental.

É possível dizer que essas iniciativas até mesmo praticam uma “metatransparência”. Além de utilizarem como insumo dados obtidos por meio de recursos da transparência, principalmente a governamental, também praticam a transparência ao compartilharem seus processos de apuração, divulgarem dados brutos para verificação e reprodução e frequentemente abrirem informações sobre patrocínios e financiamentos. Ou seja, essas iniciativas são transparentes enquanto defendem a transparência em sua amplitude de significados.

A **Transparência Brasil**² é uma organização não-governamental que surgiu em 2000 e teve papel de destaque na aprovação da LAI em 2011, tendo participado da primeira redação da lei. Com enfoque no combate à corrupção, também foi responsável pelo projeto Excelências, ativo entre 2001 e 2017, um banco de dados que reunia informações de processos jurídicos e do desempenho parlamentar de membros do Congresso Nacional e que impactou diretamente na Lei da Ficha Limpa. Claudio Weber Abramo, jornalista e matemático que dá nome ao principal prêmio de jornalismo de dados no país, esteve à frente da Transparência Brasil por quase 15 anos.

De acordo com o estatuto da ONG sem fins lucrativos, o objetivo é promover a defesa do interesse público. Alinhando discurso e prática, o *site* aponta todas as fontes de financiamento desde 2016, com os montantes doados, balanços, relatórios de auditorias realizadas e declarações de rendimentos. Os associados ao Transparência Brasil podem participar e votar na Assembleia Geral, em que são escolhidos os Conselhos Fiscal e Deliberativo, fazer sugestões sobre processos institucionais, ocupar cargos eletivos nos Conselhos e acessar antecipadamente informações sobre atividades internas e projetos que ainda não são de acesso público. São aceitos o apoio tanto de pessoas jurídicas quanto de pessoas físicas, mas com uma diferenciação importante: as pessoas jurídicas apoiadoras não têm direito a voto ou cargo nos Conselhos.

O Transparência Brasil, então, representa um caso em que a transparência governamental é monitorada e que explora os recursos disponibilizados pelos três poderes para encontrar irregularidades e propor políticas públicas, tendo servido de fonte para veículos como Poder 360, Estadão, BBC Brasil, UOL, Jornal Nacional e Revista Piauí. Dentre os projetos

² Disponível em: <https://www.transparencia.org.br>. Acesso em: 20 nov. 2023.

atuais, podemos destacar o **Achados e Pedidos**³, um repositório de pedidos feitos por meio da LAI, em parceria com a Associação Brasileira de Jornalismo Investigativo (Abraji) e financiado pela Fundação Ford.

Este projeto funciona de forma colaborativa, em que qualquer usuário pode adicionar pedidos feitos e avaliar a qualidade de respostas obtidas, e o esforço coletivo se traduz na construção de uma base de dados que permite analisar a taxa de pedidos atendidos, não atendidos ou parcialmente atendidos. A partir de junho de 2022, o Achados e Pedidos começou a utilizar uma inteligência artificial que analisa a estrutura dos textos das solicitações e das respostas a fim de acelerar a classificação, mas ainda assim é desejada a participação dos usuários devido à possibilidade de apontar erros na análise feita automaticamente. É importante ressaltar que o código da ferramenta está disponível em repositório aberto para o público.

O *site* ainda permite analisar o atendimento e o tempo de resposta dos pedidos por nível federativo e esfera de poder. Dentre seus objetivos, o Achados e Pedidos também busca facilitar o acesso a informações, poupando o trabalho de jornalistas, ao oferecer respostas dadas anteriormente via LAI, e da administração pública, ao evitar pedidos duplicados. A plataforma conta com mais de 350 mil pedidos registrados e permite a filtragem por órgão, data, tipo de resposta, esfera de poder e nível federativo.

Como a LAI é uma ferramenta importante de apuração, o Achados e Pedidos também inclui desde janeiro de 2022 um *chatbot*, o RepLAI, desenvolvido pelo Transparência Brasil que ajuda na elaboração de recursos contra respostas negativas a pedidos da LAI feitos pelos usuários. Esse *chatbot* foi elaborado com o financiamento do Fundo de Inovação para América Latina e o Caribe. Apesar de ter sido criado para lidar com os recursos, o RepLAI também oferece instruções para pedir informações e *links* diretos para os sistemas eletrônicos que recebem os pedidos da LAI.

Entre os parceiros do projeto, estão listadas as agências de checagem Lupa e Aos Fatos, a *Open Knowledge* Brasil – responsável pela Escola de Dados e, em parceria com a Abraji, pelo Prêmio Cláudio Weber Abramo de Jornalismo de Dados – e a **Fiquem Sabendo**⁴, uma agência de dados especializada na LAI. A Fiquem Sabendo declara, entre suas formas de atuação, a meta de fortalecer o jornalismo investigativo e o jornalismo local, sendo uma fonte de dados confiáveis e checáveis. A agência, além de obter e entregar dados prontos para serem utilizados

³ Disponível em: <https://achadosepedidos.org.br/>. Acesso em: 22 nov. 2023.

⁴ Disponível em: <https://fiquemsabendo.com.br>. Acesso em: 22 nov. 2023.

por outros veículos, oferece oficinas e produz conteúdos didáticos para especialistas, estudantes e para o público geral.

O *site* da Fiquem Sabendo possui uma seção dedicada à prestação de contas, disponibilizando o estatuto vigente, relatórios de atividades, balanços patrimoniais e demonstrativos financeiros desde 2019. Sem fins lucrativos, a agência publicou mais de mil bases de dados, produziu mais de sete mil reportagens e é responsável pela WikiLAI, um portal no modelo Wikipédia dedicado à Lei de Acesso à Informação. Estão listadas entre as organizações parceiras a Associação de Jornalismo Digital (Ajour) e o Observatório da Ética Jornalística (objETHOS, vinculado ao Programa de Pós-Graduação em Jornalismo da Universidade Federal de Santa Catarina).

Outro projeto de destaque da Fiquem Sabendo é a newsletter Don't LAI to me, uma comunicação que já passou das cem edições e chega quinzenalmente para os assinantes com dados inéditos e investigações que partiram de informações obtidas via LAI. Todo material da Fiquem Sabendo está sob a licença *Creative Commons* que permite republicação ou adaptação mesmo que para fins comerciais, desde que identifiquem a fonte e indiquem o *link* para a publicação original, ou seja, a agência cobra a transparência governamental, pratica e incentiva a transparência no jornalismo.

Com este breve panorama estabelecido, confirmamos a relevância da transparência como ferramenta fundamental para a manutenção e a fiscalização de processos democráticos e como, ao divulgar dados sobre a transparência governamental, a transparência se torna também uma prática nessas iniciativas que são frequentemente intermediárias entre as fontes originais e os veículos que vão reportar as informações – nos casos em que os dados são apenas publicados de forma bruta, sem uma construção jornalística ou análises. Em seguida, vamos pensar na transparência por meio de 15 critérios estabelecidos por Gehrke (2021) para, então, analisar o recorte da transparência praticada no jornalismo.

1.2 Protocolo de Transparência no Jornalismo Guiado por Dados

Conforme observado neste capítulo, a transparência ocupa um papel de destaque no jornalismo quando pensamos na credibilidade e em uma melhor relação com as audiências. Especificamente no cenário digital e no jornalismo de dados, há uma ampla gama de ferramentas e iniciativas para propagar tais práticas e democratizar o acesso à informação.

Assim, esta seção pretende observar os parâmetros do Protocolo de Transparência desenvolvido por Gehrke (2021), que permite a análise de jornais que trabalham com dados, a fim de compreender melhor as diferentes dimensões da transparência. Esta será a ferramenta utilizada no próximo capítulo para realizar a análise comparativa entre os objetos da dissertação, *The Pudding* e Gênero e Número.

A partir do panorama da pandemia da Covid-19, o chamado “apagão de dados” do Ministério da Saúde e a ampla desinformação sobre a doença, Gehrke (2021) observou de perto a relevância dos mecanismos de transparência e das políticas de dados abertos. Em sua tese, a autora ressaltou que o papel do jornalismo durante esse período foi explicar às pessoas o que estava acontecendo e como interpretar a situação, exigindo dos profissionais um certo ceticismo diante de estatísticas que variavam em contexto de produção e disponibilidade, a depender dos governos regionais.

Neste sentido, se mostrava necessário que as limitações dos registros fossem expostas, explicando métodos de trabalho, indicando fontes consultadas e qual o nível de confiabilidade de tais fontes, oferecendo, então, transparência ao público. Em sua pesquisa, Gehrke (2021) analisa a transparência por meio de dois conceitos: como operadora de uma prática ética – tendo como base as práticas operacionais da reportagem – e como recurso discursivo de legitimação – relacionando-se à credibilidade. De acordo com a autora, as categorias se conectam porque uma boa prática jornalística é necessária para construir a confiança da audiência.

Ao longo da tese, a autora defende que a transparência de produtos jornalísticos deve abarcar, além dos aspectos operacionais, os organizacionais, incluindo as formas de financiamento e políticas editoriais, acrescentando que o conhecimento sobre a origem dos recursos pode informar sobre as coberturas feitas ou ignoradas pelo jornal. Assim, ela chega ao Protocolo de Transparência no Jornalismo Guiado por Dados, que tem como função a análise de produtos já publicados.

Gehrke (2021, p. 166-167) estrutura o Protocolo em dois níveis, organizacional e operacional, que somam 15 elementos. Os parâmetros organizacionais são: “assinatura, perfil e expertise; contato; correção de erros; fontes de financiamento; manual da redação e princípios editoriais; política de diversidade; política de privacidade; e tipo de conteúdo”. Os elementos do nível operacional são: “data de publicação; documentos; fontes de dados; graus de incerteza; marcas de apuração; metodologia; e reprodutibilidade e código aberto”. Cada um dos parâmetros terá seus critérios explicitados a seguir.

A primeira categoria, **assinatura, perfil e expertise**, considera se as matérias são assinadas, se é possível acessar informações biográficas e outros trabalhos do autor e se há alguma indicação do nível de conhecimento do autor sobre o tema abordado. **Contato** consiste na avaliação dos modos disponíveis para contatar o jornal, se esses caminhos são acessíveis para o leitor, se há uma editoria de audiência e se o jornal possui ombudsman. Ainda nesse elemento, analisa-se a presença, ou não, de seção de comentários do leitor e, no caso de haver essa seção, se ela é efetivamente mediada pelo jornal ou apenas remete a uma rede social.

Correção de erros trata das políticas existentes para corrigir erros e de como acontecem as retificações no jornal: se os erros e eventuais mudanças são indicados no texto original com data e hora das alterações, se há uma política clara sobre a correção de erros e se é de fácil acesso para os leitores, se há uma seção específica do jornal para as retratações, e de que forma os leitores podem comunicar erros encontrados. **Fontes de financiamento** inclui, além da identificação de patrocinadores e fontes de financiamento do jornal, o acesso ao balanço patrimonial detalhado, seção de prestação de contas, a disponibilização da receita proveniente da audiência e, no caso de fazer parte de uma corporação ou conglomerado, a indicação de empresas do mesmo grupo.

A categoria de **manual da redação e princípios editoriais** considera se a transparência é mencionada dentre os valores do jornal, se há um manual da redação disponível para os leitores e se há uma preocupação, neste manual, em explicitar a conduta ética do jornalista e suas relações com as fontes. A **política de privacidade** avalia se o leitor é informado sobre o tratamento de seus dados pessoais ao acessar o *site* e se há algum detalhamento do tipo de informação coletada. Este parâmetro também considera se o jornal apresenta dados sobre o perfil da audiência.

Política de diversidade observa as práticas existentes ou não de diversidade na equipe, nos temas e nas fontes do jornal, se há uma editoria responsável por essa política e se o jornal possui subsídios ou ferramentas para que o leitor possa monitorar a aplicação dessa política. A categoria **tipo de conteúdo** considera a presença de marcadores que permitam a identificação de conteúdos opinativos, conteúdos pagos e conteúdo editorial, se há investimento em projetos para o letramento do leitor para que ele saiba diferenciar informação e opinião e para que ele compreenda o processo produtivo de reportagens. No caso dos conteúdos opinativos, deve-se analisar também se o texto é de um especialista na temática abordada.

No nível operacional, a categoria **data de publicação** considera a visibilidade das datas dos conteúdos publicados, bem como a data de possíveis atualizações ou correções, ressaltando

que houve modificação do conteúdo original. Caso o conteúdo tenha entrado em recirculação, também é necessário observar se o leitor é alertado sobre este fato. **Documentos** é o elemento que trata dos documentos utilizados como fonte: se estão identificados, se existe a possibilidade de consulta na íntegra e de que forma eles estão disponibilizados – junto ao corpo do texto ou *hiperlink*, por exemplo.

O parâmetro **fontes de dados** questiona a identificação das fontes utilizadas, *hiperlink* para os dados originais, o acesso às planilhas utilizadas, materiais complementares que possam oferecer ao leitor o contexto de produção dos dados e maior compreensão das variáveis. **Graus de incerteza** analisa se as reportagens indicam as limitações dos dados e métodos empregados, as possíveis lacunas e o caráter provisório das informações consultadas, apontando o que ainda precisa ser explicado ou apurado.

Marcas de apuração é a categoria em que se observa nas reportagens se está claro o trabalho de apuração e verificação do jornalista, com menções de dificuldades de acesso aos dados, o caminho percorrido para encontrá-los, as tentativas de contato com fontes e se houve necessidade de recorrer à LAI. A categoria da **metodologia** se refere à seção de nota metodológica nas reportagens, muito abordada nesta dissertação, que apresenta os procedimentos de seleção, coleta e análise dos dados, apontando caminhos para que a audiência consiga replicar a análise realizada pelo jornal.

Por último, o parâmetro de **reprodutibilidade e código aberto** analisa a divulgação de códigos-fonte, no caso do uso de linguagens de programação, e dos procedimentos realizados para seleção e análise. Se o código estiver disponível, avalia-se também se ele está compartilhado em repositório aberto e possui licença *creative commons* para reprodução. Diante desses 15 parâmetros e de seus critérios, a aplicação do Protocolo pode ser realizada de forma a pontuar elementos que tenham maior relevância para a análise e, apesar de ter sido elaborado tendo em mente o jornalismo de dados, os critérios de análise também servem para outros produtos jornalísticos (Gehrke, 2021).

2 O JORNALISMO DE DADOS E AS TRANSPARÊNCIAS

Pensar as diferentes formas de transparência especificamente no jornalismo de dados requer maior clareza sobre o que diferencia essa prática em relação ao jornalismo tido como tradicional. Para Träsel (2019), o jornalismo guiado por dados é a aplicação da computação e dos saberes das ciências sociais na coleta, processamento, interpretação e apresentação de dados. Mas se considerarmos que o fazer jornalístico parte de informações, e que não é inédita a apuração que envolve números e tabelas, nem o uso de gráficos, o jornalismo de dados é tão antigo quanto o jornalismo.

Um dos exemplos clássicos para o uso de dados, ainda de forma analógica, foi a visualização elaborada pela enfermeira inglesa Florence Nightingale⁵ no século XIX. A prática atual herda traços diretos da Reportagem com Auxílio do Computador (RAC), cuja primeira aplicação remonta ao anos 1950, e do jornalismo de precisão, na década de 1970 (Gray; Bounegru; Chambers, 2012). Mas, a partir da popularização dos computadores pessoais, tornou-se visível o processo de dataficação, ou seja, a transformação de informações – inclusive aquelas que antes não eram vistas como dados, como localização e interações sociais – em dados quantificáveis (Mayer-Schönberger; Cukier, 2013).

Como o mundo digital permite registrar, armazenar e processar quantidades massivas de informações no formato de dados, pensar, então, sobre o que é diferente no jornalismo de dados poderia ser visto como um trabalho de tradução. Durante a fase de apuração, os dados são procurados de acordo com a pauta e é realizada a coleta – ou, a partir de dados já disponíveis para o jornalista, as planilhas são “entrevistadas”, como dito por Träsel (2014a).

As entrevistas feitas no jornalismo tradicional também cabem nesta modalidade, na investigação das bases obtidas ou construídas, mas esta fase pode incluir também contatos com as fontes dos dados para maior contextualização sobre possíveis vieses e permissões de uso. A redação agora inclui o planejamento das visualizações, interativas ou não, que pretendem mostrar o resultado das análises feitas a partir das planilhas. E a publicação pode envolver, ainda, uma preocupação especial com a responsividade, uma forma de garantir o funcionamento dos gráficos para diferentes plataformas de reprodução e a adaptação das visualizações para diferentes tamanhos de tela.

⁵ Florence Nightingale reuniu dados sobre a morte de soldados ingleses durante a Guerra da Crimeia e elaborou um gráfico em 1858 para mostrar a proporção de mortes que não aconteceram durante combates, mas que poderiam ser evitadas com melhores condições de higiene no hospital, por exemplo (Belisário *et al.*, 2020).

Essa equivalência de etapas, ainda que abarque mudanças que exigem novas habilidades dos profissionais, trazem familiaridade para o que antes parecia muito diferente. Afinal, como argumenta Träsel (2014a), ferramentas digitais e bancos de dados servem como instrumentos que não substituem as habilidades dos repórteres, mas as potencializam. Ainda assim, o uso de dados é muitas vezes visto como um recorte dentro do campo, uma abordagem pontual e específica, apesar de não parecer possível separar de forma rigorosa as práticas jornalísticas em meio digital.

Mielniczuk e Träsel (2017) alertam para o fato de que as bases de dados são equivalentes a fontes humanas atualmente, e, neste cenário, é ideal que as técnicas de coleta, análise e apresentação de números estejam presentes nas diversas modalidades de reportagem, e não restritas ao que é chamado de jornalismo de dados. Os pontos que parecem divergir em relação às práticas tradicionais, como a elaboração de planilhas e as visualizações interativas, são alcançados a partir das técnicas que lidam com os avanços tecnológicos e o volume massivo de informação disponível, e a manipulação e leitura desses dados é mediada por artefatos para posterior atribuição de sentido.

Como exemplificado no capítulo anterior, a atual variedade de ferramentas e plataformas disponíveis, acessíveis para pessoas que não têm habilidades computacionais como a programação, possibilitam uma abertura de mercado para profissionais que não tiveram treinamento de dados durante a graduação ou por meios formais, dando espaço também para novas pautas – para Mayer-Schönberger e Cukier (2013), a atual oferta de quantidades massivas de dados implica em um potencial informativo que só é limitado pela criatividade.

Com as técnicas de jornalismo de dados, como Träsel (2019) destaca, é possível manejar com facilidade milhares ou milhões de registros, comparar fontes e produzir novos conhecimentos, apresentando resultados em formatos variados. Especificamente, a prática de “entrevistar planilhas” surge como caminho para descobrir pautas inéditas, também aumenta a confiabilidade. Mielniczuk e Träsel (2017) consideram o jornalismo guiado por dados como uma inovação que oferece recursos para desempenhar com maior eficiência a função social do jornalismo, e ressaltam que a educação, seja formal ou informal, é o ponto crucial para a difusão dessas técnicas para a comunidade profissional.

Da mesma forma que na prática tradicional, no jornalismo de dados o objetivo é transmitir uma informação, e, para isso, o leitor deve receber os dados de forma a compreender a história por trás dos números – isso significa que não basta oferecer planilhas e números soltos para a audiência, e sim o que foi analisado a partir dos dados e o contexto em que eles se

apresentam. Tendo em mente a proximidade entre o jornalismo tido como tradicional e o chamado jornalismo de dados, este capítulo traz dois jornais que trabalham o jornalismo de dados mas com focos diferentes.

O *The Pudding*, veículo estadunidense, destaca a dimensão visual e é adepto da interatividade em larga escala, enquanto a *Gênero e Número*, jornal brasileiro, apresenta visualizações simples priorizando pautas que têm impacto social. O que os dois jornais têm em comum é que são nativos digitais e a adoção de práticas de transparência, que, como é defendido ao longo desta dissertação, é uma ferramenta importante para o aprendizado do jornalismo de dados. Nas próximas seções, cada um deles será analisado e, por fim, será realizada uma análise comparativa que terá como base o Protocolo de Transparência de Gehrke (2021).

2.1 *The Pudding*

Fundado em janeiro de 2017, o *The Pudding* é um jornal digital que produz reportagens de dados. Com o objetivo de explicar temas complexos ou encontrar respostas para perguntas que surgem no cotidiano, a equipe do *The Pudding* traduz em visualizações elaboradas, mas acessíveis, as bases de dados exploradas. A visualização de dados, normalmente considerada apenas uma das etapas do jornalismo de dados, tem papel central no *The Pudding* – por isso, as matérias são chamadas por eles de ensaios visuais.

O nome *The Pudding* vem da frase “the proof is in the pudding” (a prova está no pudim, em tradução livre), uma corruptela do provérbio “the proof of the pudding is in the eating” (a prova do pudim está no comer), que significa que só se sabe se algo é bom ou ruim ao experimentar (NPR, 2012), ou que o valor de algo pode ser comprovado ao colocá-lo em teste. Isso se relaciona com um dos objetivos declarados pelo projeto: colocar debates cotidianos em teste e apresentar com dados a resposta encontrada.

Figura 1 – Página inicial do *The Pudding*

The Pudding

is a digital publication that ~~explains ideas debated in culture with visual essays~~ brings you longform data journalism. You might have seen our story on [mapping famous people](#) , but we've also made stuff about [the birthday paradox](#)  and [roasting your Spotify](#) . Some of my favorite projects are about [This American Life](#) , [the sounds of Mexico City](#) , and [crossword inclusivity](#) . So, welcome to The Pudding. Get cozy, you could be here a while. Explore all of our stories below.

Fonte: Página inicial do *The Pudding* em captura de tela realizada pela autora. Disponível em:

<https://pudding.cool>. Acesso em: 26 jun. 2023.

Além do papel central da visualização de dados nos ensaios visuais, a equipe do *The Pudding* traduz a conexão com o jornalismo de dados na forma como se identificam: jornalistas-engenheiros. Este conceito foi desenvolvido por Matt Daniels (2015), fundador do *The Pudding*, ao perceber que alguns de seus artigos preferidos de *The New York Times* e *Bloomberg* eram construídos a partir do uso de linguagens de código e eram, predominantemente, *software*, não prosa. Assim, ele traduziu a ideia de profissionais generalistas, capazes de trabalhar com todas as etapas de um projeto de dados, nos jornalistas-engenheiros.

A equipe do *The Pudding* busca incorporar esse princípio, compondo uma redação multidisciplinar a partir de caminhos acadêmicos e profissionais variados. Ainda que todas as trajetórias tenham convergido para o jornalismo de dados, elas têm origem nas áreas de neurociência, psicologia, filosofia, administração, engenharia de software, ciência da computação e jornalismo. Neste ponto, vale ressaltar que, apesar do *The Pudding* dispensar o diploma de jornalismo para o exercício da profissão, a intenção de trazer este objeto de pesquisa não é corroborar esta prática, e sim propor maior profundidade no ensino do jornalismo de dados na graduação ou o fácil acesso a conteúdos que possibilitem o aprendizado por parte de profissionais da área.

Quando pensamos no profissional de jornalismo de dados, considerando que, devido à presença constante do digital e das bases na prática cotidiana dos jornalistas, o conceito do jornalista-engenheiro não é uma ideia revolucionária. Ainda assim, o termo será mantido na

análise deste objeto para reforçar que todos os colaboradores da equipe têm conhecimentos jornalísticos e técnicos para desenvolver qualquer e toda etapa necessária na construção de uma reportagem de dados, com ressaltando que, no caso do *The Pudding*, o enfoque nas visualizações interativas exige maior familiaridade com linguagens de programação do que o mínimo necessário no jornalismo de dados.

Em uma equipe formada por jornalistas-engenheiros, não há separação prévia entre as funções na execução de cada matéria. No *The Pudding*, a equipe se reestrutura de acordo com os projetos em desenvolvimento, funcionando de forma mais próxima de um coletivo do que de uma hierarquia. Por isso, os jornalistas-engenheiros costumam estabelecer parceiros de projeto, uma pessoa na equipe que vai acompanhar de perto o desenvolvimento de uma matéria, oferecer retorno diário e colaborar em diversas etapas, como Caitlyn Ralph explicou durante uma palestra⁶ no *Encode 2019*. Como todos têm capacidade técnica de desenvolver qualquer etapa do trabalho com dados, o processo editorial já estabelecido evita que a ausência de um chefe seja prejudicial, e isso começa na seleção dos temas a serem explorados.

Além de uma equipe de jornalistas-engenheiros, o *The Pudding* possui algumas especificidades a serem explicitadas ao longo desta seção: a clareza sobre as formas de lucro do *site*; o compartilhamento frequente das etapas de construção das matérias; a disponibilização de tutoriais sobre as ferramentas utilizadas e sobre a produção de reportagens de dados. Considerando que o *The Pudding* não é um veículo tradicional, nem uma agência dentro de um jornal maior, é necessário entender o que o diferencia de grandes jornais. Zamith (2019), ao analisar os conteúdos produzidos por *The New York Times* e *The Washington Post* em 2017, o ano de criação do *The Pudding*, observa que não há complexidade nas visualizações, que têm baixo nível de interatividade, e que estes jornais não disponibilizam conjuntos originais de dados, dependendo fortemente de fontes oficiais – práticas opostas do que é observado no *The Pudding*.

Para observar o processo editorial do *The Pudding*, podemos partir do tutorial “Continue, pivot or put it down”⁷, escrito por Amber Thomas em 2020. Este é um dos principais recursos de transparência do jornal, visto que explica, passo a passo, as etapas de trabalho – que podem ser resumidas como as fases de ideias, exploração e apresentação –, podendo ser visto como recurso de aprendizado ao explicitar o que é necessário para construir os ensaios visuais.

⁶ Disponível em: <https://www.youtube.com/watch?v=nFImiNE4ZZg>. Acesso em 3 jun. 2023.

⁷ Disponível em: <https://pudding.cool/process/pivot-continue-down/>. Acesso em: 29 dez. 2022.

O modelo do *The Pudding* pode ser reproduzido para outras equipes de trabalho, ou, ainda, para produções individuais.

O fluxo de trabalho do *The Pudding* começa na etapa de ideias, partindo dos *backlogs*. O *backlog* é a coleção de ideias, perguntas e potenciais matérias que não foram desenvolvidas, um lugar para registrar constantemente os pensamentos e as inspirações que talvez resultem em ensaios visuais, ou questões interessantes que emergem do cotidiano. Cada um dos jornalistas-engenheiros possui um *backlog* próprio, mas existe, também, um registro coletivo que é disponibilizado⁸ para o público. Observando os processos produtivos do *The Pudding* e as matérias publicadas, nota-se que esse uso do *backlog* é possível porque o jornal lida, em grande parte, com pautas frias ou, pelo menos, que não são tão sensíveis ao tempo. Nesse contexto, é possível guardar as ideias não trabalhadas para momentos posteriores sem que se perca a relevância do tema.

Semanalmente, durante uma reunião chamada *story time* (hora da história, em tradução livre), um dos jornalistas-engenheiros é o responsável por trazer ideias do *backlog* como sugestão de próximas pautas, e, então, a equipe ajuda a selecionar qual delas será desenvolvida para se tornar um ensaio visual. Seguindo o padrão de matérias do *The Pudding*, o tema precisa ser do interesse do jornalista-engenheiro que trouxe a ideia, visto que ele passará semanas focado naquele assunto, e contribuir com o propósito do jornal de gerar conversa. Como apontado anteriormente, trabalhar pautas quentes não é a prioridade do *The Pudding*, então o aspecto temporal não parece pesar na seleção de pautas.

Para o desenvolvimento das histórias, parte-se de uma pergunta que gere interesse na equipe e, em seguida, deve-se pensar se esta pergunta pode ser respondida com dados. Caso negativo, retorna-se à fase de elaborar perguntas, mudando o ângulo ou recomeçando a partir de uma nova ideia. Essa pergunta pode ter um foco amplo ou estreito, diferenciação que foi abordada na série de tutoriais “How to make dope shit”⁹ – escrita por Ilia Blinderman e dividida em partes um¹⁰, dois¹¹ e três¹². O foco amplo significa partir de uma análise exploratória dos dados, e essa abordagem funciona quando o conjunto de dados é interessante ou a análise tem resultados impactantes.

⁸ Disponível em: https://docs.google.com/spreadsheets/d/1neuenck0xy5piNZaaxPhPL0Wsmpe_gqFjeXqV3MbZE/edit#gid=0. Acesso em: 20 jun. 2023.

⁹ Em tradução livre: “Como fazer coisas maneiras”.

¹⁰ Disponível em: <https://pudding.cool/process/how-to-make-dope-shit-part-1/>. Acesso em: 28 dez. 2022.

¹¹ Disponível em: <https://pudding.cool/process/how-to-make-dope-shit-part-2/>. Acesso em: 6 jun. 2023.

¹² Disponível em: <https://pudding.cool/process/how-to-make-dope-shit-part-3/>. Acesso em: 6 jun. 2023.

Um exemplo de foco amplo foi a matéria “The structure of stand-up comedy”¹³. Em uma transmissão no *YouTube*¹⁴, Matt Daniels e Russell Goldenberg explicam que deu-se início à investigação com uma coleta de dados, sem muita ideia de qual era o objetivo do ensaio visual, e a consequência disso é que muitos dos dados coletados foram descartados. Apesar de ser comum dar início a um projeto de visualização de dados partindo do máximo de dados que foi coletado até então, Fry (2008) considera que é mais interessante partir de perguntas. Essa ideia está alinhada com a outra abordagem, de foco estreito, que é a comumente utilizada pelos jornalistas-engenheiros: parte-se de uma pergunta central e, então, são procurados os dados disponíveis que talvez respondam à pergunta.

Depois de escolhida a pauta, a próxima etapa é de exploração do tema, em que procura-se o que já foi escrito sobre o tópico e se há dados disponíveis ou a serem coletados. O objetivo da fase é analisar se há material para produzir uma reportagem. Caso esses dados já existam, é fundamental compreender com que objetivo eles foram coletados, visto que isso vai influenciar nas interpretações. Caso o conjunto de dados ainda não esteja pronto, a criação desse conjunto pode envolver uma coleta manual ou a busca por fontes de dados brutos, como *sites* de órgãos governamentais. Nessa fase, os jornalistas-engenheiros recomendam uma análise ética da coleta e do uso dos dados.

Essa questão foi abordada por Craig, Ketterer e Yousuf (2017) na análise de casos controversos em que jornalistas publicaram dados sensíveis – nomes e endereços – ao repercutirem tiroteios em massa nos Estados Unidos, mas, de acordo com o artigo, o estudo tem amplitude maior que estes casos, visto que levanta questões éticas para jornalistas que trabalham com dados. Dentre os principais problemas estão a publicação de informações pessoais, dados que coloquem pessoas em risco ou que possam causar impacto negativo caso estejam desatualizados ou errados, bem como dados que não podem ser verificados (Craig; Ketterer; Yousuf, 2017).

Tendo isso em mente e com o conjunto de dados pronto, inicia-se a análise, que, no caso do *The Pudding*, pode incluir algumas linguagens de programação, como *Python*, *R*, *JavaScript* e *SQL*, mas, dependendo da complexidade da tarefa, é possível também utilizar ferramentas como *Excel* e *Google Sheets*, que não exigem o trabalho com códigos. É ainda durante a fase exploratória que se cria um gancho para a matéria, é determinado o recorte do tema e o

¹³ Disponível em: <https://pudding.cool/2018/02/stand-up/>. Acesso em: 23 jun. 2023.

¹⁴ Disponível em: <https://www.youtube.com/watch?v=SQ7DiSW-C4k>. Acesso em: 26 jun. 2023.

jornalista-engenheiro responsável tenta provar para a equipe que o conteúdo é interessante, mostrando uma análise inicial dos dados ou por meio de *storyboards* de um rascunho do ensaio visual, em que são analisadas especificações do projeto e se a visualização será interativa, por exemplo.

Por mais que haja uma pessoa responsável pelo desenvolvimento da matéria, todas as etapas do trabalho envolvem *feedbacks* da equipe, devido à não hierarquização dentro da redação. Ainda que cada um dos ensaios visuais seja atribuído a um jornalista-engenheiro e que seja frequente o trabalho em duplas ou trios, a política editorial é que as matérias passem pelas outras pessoas da equipe antes da publicação. Dessa forma, todos os jornalistas-engenheiros podem colaborar com sugestões ou são convidados a participar da construção da narrativa visual. Nas reuniões de acompanhamento das reportagens, são trazidas questões que podem mudar o rumo da matéria, e, só depois dos ajustes e das conclusões às quais os dados permitem chegar, tem início a fase da apresentação, em que se começa a produzir efetivamente as visualizações que antes eram rascunhos ou versões estáticas, sem interatividade.

Ainda em “How to make dope shit”, o jornalista-engenheiro Ilia Blinderman oferece dicas gerais para se pensar o *design* dos ensaios visuais: observar outras visualizações para entender como foram construídas e se inspirar; pensar na mensagem a ser comunicada e como isso é traduzido imagetivamente; não trabalhar a etapa de *design* direto em código; caso o objetivo seja a interatividade, utilizar linguagens de programação. Essas recomendações se alinham com o que é visto em Fry (2008), Cairo (2019) e em outras pesquisas sobre a visualização de dados. É nesta fase em que os ensaios visuais tomam a forma final e a narrativa é publicada.

Todas as fases de construção dos ensaios visuais são realizadas em plataformas que permitem o compartilhamento dos arquivos e a edição dentro da equipe, e todos têm acesso total ao que está em desenvolvimento. Essas práticas colaborativas se refletem também na audiência, visto que as matérias são acompanhadas por uma seção chamada metodologia que aponta as fontes dos dados e, frequentemente, os procedimentos importantes realizados na análise.

Além disso, a equipe disponibiliza as bases de dados originais no *GitHub*¹⁵ do jornal, bem como os códigos utilizados para programar o *site*. Como observado no processo editorial do *The Pudding*, as habilidades jornalísticas da equipe são acompanhadas pelo conhecimento

¹⁵ O *GitHub* é amplamente utilizado por programadores por ser uma plataforma de hospedagem que permite a colaboração e o compartilhamento de arquivos e códigos.

de linguagens de programação. Na construção da interface gráfica das páginas *web*, utiliza-se principalmente *HTML*, *CSS* e *JavaScript* – em redações tradicionais, essa parte do trabalho não está nas mãos dos jornalistas, mas, com os jornalistas-engenheiros, as funções são integradas.

Pensando especificamente na visualização de dados, a interatividade depende de bibliotecas das linguagens de programação *JavaScript* ou *Svelte*, enquanto os gráficos estáticos são criados a partir de ferramentas como *Figma*, uma plataforma de *design*, *Flourish*, que oferece modelos prontos de gráficos a serem personalizados, ou com o auxílio da linguagem de programação *R*. Ainda que todos esses recursos pareçam distantes da realidade dos jornalistas tradicionais, a equipe do *The Pudding* advoga pelo uso de ferramentas gratuitas para que isso não seja um impeditivo para quem quiser produzir reportagens inspiradas nos ensaios visuais.

Jornal nativo do ambiente digital e sem redação presencial desde sua criação, o *The Pudding* não possui propagandas de terceiros no *site*. Sem esta opção, que é uma forma corriqueira de obter lucro na *internet*, o *The Pudding* tem como principal fonte de renda o *Polygraph*, uma agência de dados própria que permite que o *The Pudding*, além de manter sem visual limpo e sem propagandas, funcione sem a necessidade de assinatura, garantindo que todo o conteúdo produzido seja de livre acesso. Isso não impede que o *The Pudding* publique matérias patrocinadas, mas estas mantêm o perfil editorial do *site*.

Um exemplo é “30 years of American anxieties”¹⁶, cujo patrocínio da *IBM* possibilitou o acesso à ferramenta *IBM Watson Studio* para a análise de 20.000 cartas de leitores enviadas à coluna *Dear Abby*, fundada em 1956. Da mesma forma, ainda que o *Polygraph* seja a versão “de aluguel” do *The Pudding*, os dois portais se mantêm alinhados com a mesma missão, contar histórias com dados e visualizações, mas com a diferença de serem direcionadas para propósitos de outras organizações, em vez de seguir os interesses da equipe (Polygraph, 2022).

Alguns projetos de destaque do *Polygraph* são: “Reconstructing seven days of protests”, em parceria com *The Washington Post* e *Google News Initiative*; uma plataforma de narrativa de dados para a *Harvard Kennedy School*; e a matéria interativa “One trillion Minecraft views” para o *YouTube*. Entre os formatos publicados com a assinatura do *Polygraph*, além de artigos e postagens em *sites* e outras plataformas, foram produzidas narrativas visuais baseadas em dados para vídeos, apresentações internas e externas e instalações artísticas.

Ainda reforçando a importância da transparência ao revelar detalhes sobre o aspecto financeiro do jornal, o *The Pudding* e o *Polygraph* declaram que os salários da equipe são

¹⁶ Disponível em: <https://pudding.cool/2018/11/dearabby/>. Acesso em: 26 abr. 2023.

baseados nas funções de jornalismo interativo e visualização de dados, reconhecendo que o conjunto de habilidades necessário na equipe se alinha com o mercado de tecnologia. Os valores não são informados para a audiência.

Além do *Polygraph*, outra fonte de renda do *The Pudding* são as doações dos leitores. O modelo de patrocínio – também utilizado pela Agência Pública e pelo *Intercept*, por exemplo – se diferencia da assinatura. Enquanto a assinatura significa a existência de barreiras de acesso para os leitores que não assinam, comum em jornais tradicionais como O Globo e Estadão, os patrocínios costumam manter o conteúdo livre e os projetos beneficiados oferecem recompensas para os doadores.

Com mais de 400 patronos¹⁷, o site oferece quatro modalidades diferentes de assinatura, com valores entre dois e 500 dólares – o plano básico já inclui conteúdos dos bastidores de produção, detalhando a construção de matérias, enquanto o plano mais caro é um modelo criado para empresas, tendo como uma das vantagens uma palestra por ano. Os lucros a partir das doações são direcionados para contratar *freelancers*, e, atualmente, as colaborações externas acontecem a cada um ou dois meses.

Não são apenas os funcionários do *The Pudding* que possuem espaço para desenvolver as narrativas visuais no *site* – as publicações somam mais de 70 colaboradores externos. Existem dois caminhos para colaborar com o *The Pudding*: acompanhando o *backlog* ou sugerindo uma ideia inédita. Para sugerir uma ideia, é necessário observar os princípios do *The Pudding*: deve ser uma pauta digna de debate, há uma verdade profunda sobre o tópico e é um assunto melhor mostrar do que contar.

Não é exigida nenhuma qualificação específica para os colaboradores *freelancers*, sendo necessário apenas mandar uma breve descrição da pauta sugerida com uma justificativa para a relevância do tema e qual será o envolvimento da equipe do *The Pudding* – se os jornalistas-engenheiros precisam ajudar na análise dos dados ou na construção das visualizações, por exemplo. A seção “Pitch a story”¹⁸ mostra o tipo de colaboração desejada e a página conta com um *link* para um formulário em que o colaborador interessado precisa dar informações como: um resumo em uma frase do objetivo da matéria; o motivo para o tema da matéria ser interessante e o que a diferencia de outras; por que ela deve ser construída de forma visual; em que ponto o colaborador está no desenvolvimento da matéria. De forma geral, os *freelancers*

¹⁷ Disponível em: <https://www.patreon.com/thepudding>. Acesso em: 26 jun. 2023.

¹⁸ Disponível em: <https://pudding.cool/pitch/>. Acesso em: 25 jun. 2023.

recebem sete mil dólares por um ensaio visual completo, mas esse valor é reduzido conforme a necessidade de assistência para a produção da reportagem (The Pudding, 2023).

Além disso, seguindo o padrão *The Pudding* de produção fora do ciclo de notícias, os colaboradores podem levar meses para concluir uma narrativa visual, visto que muitos dos *freelancers* possuem outros empregos. Assim como evidenciado no processo editorial apresentado anteriormente, a seção reforça que as histórias começam com boas perguntas, e, caso haja a necessidade de buscar inspiração, o *backlog* da equipe está disponível. Esse *backlog* é uma planilha aberta para visualização e reúne pautas que foram trazidas pela equipe mas ainda não foram desenvolvidas, seja por falta de dados suficientes ou pela necessidade de redirecionar o tema.

Em maio de 2023, o *backlog* reunia 51 ideias, sendo duas inspiradas em sugestões de leitores. A mais antiga é de março de 2017, enquanto a mais recente é de novembro de 2021. As propostas são bem diversificadas. Algumas incluem o conjunto de dados que deu origem à pauta – a ideia de observar as páginas da Wikipédia com pico de tráfego de acordo com o dia do ano, por exemplo –, enquanto outras são bem genéricas – um rastreador para acompanhar o que a *internet* está falando sobre o ator Timothée Chalamet – ou apenas com uma breve contextualização, mas sem indicação de como construir a matéria – como o questionamento se é necessária alguma habilidade para *fantasy sports*.

A seção *Our Resources*¹⁹ reúne tutoriais publicados no *site*, perguntas frequentes, *links* externos para vídeos publicados no canal do *YouTube* e participações dos jornalistas-engenheiros em eventos e entrevistas. Esses recursos abrem para o público os bastidores do processo editorial e, principalmente, das ferramentas utilizadas na construção dos ensaios visuais. Em 2023, são 14 tutoriais que abordam diferentes etapas do trabalho com dados, sendo a fase da visualização a com maior ênfase nos materiais disponíveis. Segundo Daniels (2017), essas postagens são uma forma de retribuir à comunidade, visto que eles usam ferramentas disponibilizadas gratuitamente e fizeram uso de tutoriais durante o aprendizado das técnicas de visualização.

Para compreender a que público esses tutoriais são direcionados, eles foram analisados e categorizados para fins didáticos nesta pesquisa nos níveis iniciante, intermediário e avançado. Essa classificação considera o conhecimento prévio necessário para compreender e

¹⁹ Disponível em: <https://pudding.cool/resources/>. Acesso em: 26 jun. 2023.

aplicar o conteúdo, e apenas os materiais direcionados para iniciantes serão explorados no próximo capítulo.

Com isso em mente, vamos analisar brevemente os recursos disponibilizados pelo *The Pudding*, dando destaque maior para aqueles que colaboram com o aprendizado de conhecimentos tidos como mais básicos ou gerais, sem abordar as etapas avançadas que requerem programação. O objetivo deste recorte é reunir materiais que sejam acessíveis e aplicáveis para jornalistas ou estudantes da área, oferecendo caminhos para a aproximação do jornalismo de dados.

Quadro 1 – Tutoriais publicados no *The Pudding*

Título do tutorial	Principais assuntos	Nível
HOW TO MAKE DOPE SHIT Part 1: Working with Data	Como trabalhar com dados; primeiros passos para aprender programação; recursos para aprender a programar	Iniciante
HOW TO MAKE DOPE SHIT Part 2: Design	Como fazer visualização de dados; recursos para aprender sobre tipos de visualizações; recursos para aprender a criar visualizações programando	Iniciante
HOW TO MAKE DOPE SHIT Part 3: Storytelling	Como construir uma história com dados; passos da <i>visual essay</i>	Iniciante
Continue, pivot or put it down	Explica o fluxo de trabalho do <i>The Pudding</i>	Iniciante
The Courage (and Disappointment) of Pitching a Visual Essay	Experiência da Erin Davis em publicar uma matéria no <i>The Pudding</i>	Iniciante
How Many Users Resize Their Browser?	Sobre <i>design</i> responsivo	Intermediário
The Making of the Weighted Pivot Scatter Plot	Explicando a construção de uma visualização específica	Intermediário
Responsive scrollytelling best practices	Sobre <i>design</i> responsivo	Intermediário
Making Data Viz Without SVG Using D3 & Flexbox	Como montar diferentes gráficos em D3; programação	Avançado
Easier scrollytelling with position sticky	CSS para <i>scrollytelling</i>	Avançado
An Introduction to Scrollama.js	Apresentando o <i>Scrollama</i>	Avançado

Título do tutorial	Principais assuntos	Nível
Regional smoothing using R	Explicando uma ferramenta para dados geoespaciais	Avançado
How to implement scrollytelling with six different libraries	Comparando bibliotecas para <i>scrollytelling</i>	Avançado
We created a speaker rider because journalism didn't have one	Lançamento de um projeto sobre inclusão em eventos e palestras	Não se aplica

Fonte: *The Pudding*, 2017-2021

Os cinco tutoriais categorizados como de nível iniciante já foram mencionados neste capítulo, reforçando que são de fácil compreensão. “Continue, pivot of put it down” fala sobre o processo editorial do *The Pudding*, enquanto “The Courage (and Disappointment) of Pitching a Visual Essay” representa as especificidades da participação de um *freelancer* no jornal, trazendo uma ideia rejeitada e outra que foi desenvolvida e publicada com os jornalistas-engenheiros.

A série “How to make dope shit”, dividida em três partes, foi motivada pelas perguntas de leitores do *The Pudding* sobre as habilidades necessárias para produzir as reportagens, explica as diferentes etapas do trabalho com dados no desenvolvimento dos ensaios visuais e pretende ser um guia não-técnico. Por isso, é considerada nesta pesquisa como sendo de nível iniciante, mesmo que fale bastante sobre programação. A primeira parte do tutorial, com foco nos dados, oferece *links* externos para quem deseja aprender *R*, *SQL* e *JavaScript*.

Além disso, uma seção dedicada a *Python* inclui a indicação do *First Python Notebook*²⁰. Este é um guia voltado para a análise de dados com a linguagem de programação *Python* criado por Ben Welsh, um jornalista de dados que trabalhou por 15 anos no *Los Angeles Times*, exemplificando mais um profissional da área que corrobora com a ideia da democratização do conhecimento.

Na segunda parte da série, o foco é o *design*, e a primeira recomendação é observar o trabalho de outras pessoas para saber o que tem sido produzido na área, a fim de entender o que pode ser feito e buscar inspiração. Para isso, o tutorial oferece uma lista de referências, incluindo jornais como *The Washington Post*, que possui uma página no *Twitter* para reunir os

²⁰ Disponível em: <https://palewi.re/docs/first-python-notebook/>. Acesso em: 26 jun. 2023.

gráficos publicados, e o *The Guardian*, cujas visualizações estão agrupadas em uma seção do *site*.

Depois de coletar as referências, é fundamental pensar na mensagem principal a ser transmitida, para, depois, decidir qual gráfico será mais efetivo. O tutorial sugere o *Data Viz Catalogue*²¹, um acervo de modelos de gráfico que aponta a visualização ideal de acordo com o que se quer mostrar. Tendo em mente que essa série é voltada para iniciantes no trabalho com dados, uma das principais recomendações do tutorial é não fazer o *design* diretamente em código, visto que essa não é uma prática comum mesmo entre os jornalistas-engenheiros. Pensar primeiro no modelo estático, sem interatividade, e montar o protótipo em uma ferramenta mais simples tem como objetivo garantir que a visualização será capaz de comunicar o que foi observado na análise dos dados.

Mas, para quem deseja aprender linguagens de programação voltadas para a etapa do *design*, o tutorial recomenda um livro para iniciantes, observar o código de visualizações produzidas por outros profissionais e fazer perguntas – neste caso, o meio mais comum para solucionar dúvidas de programação é o *Stack Overflow*, um *site* utilizado por programadores e desenvolvedores que funciona de acordo com os ideais da inteligência coletiva²²: um conhecimento distribuído, coordenado em tempo real, que tem como base o enriquecimento mútuo (Lévy, 2007).

Enquanto as duas primeiras partes dessa série focaram em oferecer caminhos para o aprendizado de técnicas, a terceira e última parte é focada em um conceito mais amplo, o *storytelling*. A construção de narrativas dos ensaios visuais do *The Pudding* envolve também o *design*, que, no tutorial, se divide entre três possibilidades. A primeira delas é para o caso da mensagem a ser transmitida ser simples: gráficos. Por serem fáceis de produzir, é possível utilizar vários em uma mesma matéria, e essa é a opção mais comum no jornalismo de dados, principalmente em veículos tradicionais.

Se o ensaio visual for para reprodução em redes sociais ou o objetivo for uma visualização mais agradável para smartphones, a recomendação do tutorial são os modelos *tapping* ou *stepping*, em que a história progride de acordo com os cliques do leitor. A terceira possibilidade de *design* é o *scrollytelling*, que oferece uma experiência mais fluida e mais

²¹ Disponível em: <https://datavizcatalogue.com/search.html>. Acesso em: 27 jun. 2023.

²² Assim como mencionado anteriormente sobre a cibercultura, a inteligência coletiva é uma terminologia superada por não haver mais o otimismo em relação às possibilidades a partir da chegada dos computadores pessoais e da difusão da *internet*, mas grupos como programadores e jornalistas de dados ainda reproduzem tais ideias em suas práticas como grupos.

divertida para o leitor, mas que depende de linguagens de programação e maior atenção a detalhes. Esta é a opção mais utilizada no *The Pudding*, definida no tutorial “How to implement scrollytelling with six different libraries” como um modo de revelar ou mudar o conteúdo da narrativa visual conforme o usuário rola a página.

O *scrollytelling* foi popularizado pela grande reportagem multimídia “Snow Fall: The Avalanche at Tunnel Creek”²³, de *The New York Times*, publicada cinco anos antes do lançamento do *The Pudding*. Esse estilo tem como diferencial a possibilidade de uma leitura imersiva, suavizando as transições imagéticas e textuais, que passam a depender apenas da rolagem de tela. Aliando gráficos interessantes, interatividade e *scrollytelling*, o conteúdo apresentado torna-se mais convidativo e acessível, mesmo quando a pauta exige profundidade.

Devido ao foco nas visualizações de dados, os elementos textuais no *The Pudding* ocupam papel secundário, funcionando como complemento e contextualização. No ensaio visual “Let's learn about waveforms”²⁴, por exemplo, o *freelancer* Josh Comeau construiu um guia interativo com *scrollytelling* para explicar diferentes formatos de ondas e como funcionam os fundamentos físicos do som. Esse conteúdo, partindo de conceitos da Física, parece direcionado a um nicho, mas, na verdade, é voltado para a audiência geral, ou seja, pode ser consumido sem a necessidade de conhecimento prévio.

Os principais motivos para utilizar o *scrollytelling* são explicitados no tutorial “Responsive scrollytelling best practices”: essa abordagem evita que o usuário precise clicar para avançar a matéria, permitindo, em vez disso, que a narrativa se apresente a partir da rolagem de tela, que é o comportamento padrão do usuário. Outro ponto positivo é que o modelo não esconde conteúdo, ou seja, não depende de movimentos do cursor para apresentar anotações ou variações menores na visualização.

O jornalista-engenheiro Russel Samora, meses depois de escrever o tutorial “How to implement scrollytelling with six different libraries”, elaborou “An introduction to Scrollama.js”, um guia introdutório para a biblioteca de *JavaScript* que ele criou, *Scrollama*, e que já foi utilizada por veículos mais conhecidos que o *The Pudding*, como *Washington Post*, *Wall Street Journal* e *ProPublica*. O objetivo do *Scrollama* é facilitar a implementação do *scrollytelling* nos ensaios visuais e melhorar a experiência de usuário. Ainda tendo em mente a

²³ Disponível em: <https://www.nytimes.com/projects/2012/snow-fall/index.html#/?part=tunnel-creek>. Acesso em: 19 jun. 2023.

²⁴ Disponível em: <https://pudding.cool/2018/02/waveforms/>. Acesso em: 28 jun. 2023.

transparência como valor e como prática, além dos tutoriais, a biblioteca está disponível para uso no *GitHub*²⁵.

Independentemente do modelo escolhido para o *design*, é fundamental ter em mente a responsividade, uma questão essencial para planejar como será a parte visual da narrativa. Pensar na responsividade significa considerar como o ensaio visual vai se adaptar ao ser reproduzido em dispositivos diversos. De acordo com o tutorial “Responsive scrollytelling best practices”, não importa se a visualização começa a ser estruturada para a versão *mobile* ou a *desktop*, desde que as duas estejam no planejamento. Além disso, a versão *mobile* diz respeito não só ao tamanho do navegador, mas à possibilidade de *internet* ruim ou gráficos que dependem do movimento do cursor, por exemplo.

A responsividade das visualizações também é discutida em “How many users resize their browser?”. Neste tutorial, classificado como intermediário nesta pesquisa devido ao trecho que fala sobre programação, Russell Samora apresenta a informação de que 2 a 3% dos leitores redimensionam o tamanho do navegador. Pensar em responsividade evita que a audiência enfrente problemas ao consumir reportagens que incluem visualização de dados, melhorando a experiência do usuário.

Os tutoriais restantes tratam de ferramentas específicas, como é o caso de “The Making of the Weighted Pivot Scatter Plot”, que trata da produção de um gráfico derivado do de dispersão, e requerem o uso de linguagens de programação. Por isso, não serão abordados nesta pesquisa, apesar de serem escritos de forma simples, sendo a única dificuldade de entendimento e aplicabilidade os conhecimentos específicos. Destaca-se, por último, que a postagem “We created a speaker rider because journalism didn’t have one” não se trata de um tutorial de jornalismo de dados, mas um guia, criado em parceria com *The OpenNews*, que procura incentivar a diversidade e a inclusão em palestras.

Partindo para uma breve análise dos recursos anexados ao *site*, mas que não são publicados no *The Pudding*, é possível observar as categorias *Live coding videos*²⁶, *Talks & presentations*, *Podcasts* e *Pudding Cup*. Além de tentar capturar a atenção da audiência mesmo em temas complexos por meio de narrativas interativas, a equipe de jornalistas-engenheiros tem se empenhado em desenvolver ferramentas e criar tutoriais para amplificar o alcance das linguagens de programação.

²⁵ Disponível em: <https://github.com/russellsamora/scrollama>. Acesso em: 6 jun. 2023.

²⁶ Decidimos não traduzir as seções dos tutoriais do Pudding, assim como das matérias, usando os nomes originais.

Isso pode ser observado nas séries de vídeo “The Svelte Way”, com três vídeos que explicam o uso de técnicas específicas de *Svelte*, “Live coding our site in SvelteKit”, com 25 vídeos, em que os jornalistas-engenheiros mostram detalhes sobre a construção do *site*, “Living coding a story from start to finish”, com 50 vídeos, e “Live coding sessions”, com cinco vídeos. Essas quatro séries são direcionadas para quem possui conhecimentos avançados na área, visto que tratam especificamente do trabalho com linguagens de programação. A exceção é o vídeo “The Pudding Process: The Structure of Stand-Up Comedy”, que mostra o passo a passo da construção de uma matéria, como foi mencionado anteriormente no capítulo.

Assim como na observação das matérias publicadas, os tutoriais disponíveis no *site* e os vídeos mencionados mostram o papel central das visualizações para o *The Pudding*. Considerando que os ensaios visuais exigem o conhecimento de linguagens de programação para sua produção, a transparência é fundamental para a divulgação desses conhecimentos no meio jornalístico, trazendo novas possibilidades para quem deseja construir matérias visualmente elaboradas.

A importância das visualizações para o jornalismo de dados também se reflete na *Pudding Cup*, uma premiação que acontece desde o primeiro ano do *The Pudding* em que são eleitas as melhores histórias visuais e guiadas por dados. Em 2017, a competição era, na verdade, uma lista das matérias preferidas dos jornalistas-engenheiros, de acordo com o que consumiram durante o ano. A partir do ano seguinte, porém, a equipe decidiu transformar o evento em uma forma de premiar pessoas que trabalham com dados, abrindo inscrições para que os candidatos viessem como indicações da audiência e oferecendo um prêmio de 3.141,59 dólares.

Como o total é dividido entre os ganhadores, o valor do prêmio é uma brincadeira com o número pi, que tem a mesma pronúncia que *pie*, porque cada um dos vencedores recebe uma fatia da “torta”. Outra mudança que impactou a *Pudding Cup* é que, desde 2019, são considerados apenas projetos não-comerciais, ou seja, que não receberam remuneração para serem produzidos e publicados, e foram criados por diversão, para projetos pessoais ou com fins educacionais.

Seguindo para os outros recursos disponíveis, é importante considerar que os *podcasts* são materiais em língua inglesa e, ao contrário dos outros recursos, não possuem elementos visuais e nem a possibilidade de tradução automática. Portanto, não são de fácil acesso para pessoas que não possuem proximidade com o idioma e não serão explorados nesta pesquisa.

Serão analisados os recursos da categoria *Talks & presentations* que se mostram relevantes para o aprendizado do jornalismo de dados. Novamente, os conteúdos são divididos entre os níveis iniciante, intermediário e avançado, seguindo os mesmos critérios explicitados anteriormente na classificação dos tutoriais – ou seja, de acordo com a necessidade de conhecimentos prévios para o aproveitamento do conteúdo.

Quadro 2 – *Talks & presentations*

Título ²⁷	Onde aconteceu	Principais assuntos	Nível
What is data storytelling?	Knight Center for Journalism in the Americas	O que é narrativa de dados	Iniciante
Things we've learned from telling the “fun” data stories	VIS in Practice 2019	Diferenciais do <i>The Pudding</i>	Iniciante
Develop your data-driven storytelling chops with four simple principles	Google Cloud Next	Princípios para narrativas com dados	Iniciante
Shaping ideas into data stories	Data Stories Speaker Series	O que significa ser um jornalista-engenheiro; transição de carreira; como ter ideias de pautas para matérias de dados	Iniciante
From story-time to send off	Encode 2019	Processo editorial do <i>The Pudding</i> ; o que são ensaios visuais	Iniciante
Animating data visualization in Svelte	Svelte Summit	Como criar visualizações animadas com a linguagem de programação <i>Svelte</i>	Avançado

Fonte: *The Pudding*, seção *Talks & presentations*, 2018-2022

Como pode ser observado no quadro acima, as palestras são, de forma geral, para um público iniciante. Com exceção da “Animating data visualization in Svelte”, que foi realizada

²⁷ Tradução livre dos títulos: O que é narrativa de dados?; Coisas que aprendemos contando histórias “divertidas” de dados; Desenvolva suas habilidades de narrativas dirigidas por dados com quatro princípios simples; Transformando ideias em histórias de dados; Do *storytime* ao lançamento; Animando visualizações de dados em *Svelte*.

numa conferência voltada para a linguagem de programação *Svelte*, as outras cinco apresentações são de fácil entendimento mesmo para quem é leigo e, por diferentes caminhos e temáticas, cumprem a função de transmitir conhecimentos iniciais sobre visualização de dados e sobre o *The Pudding* em si.

Em “What is data storytelling?”²⁸, aula do último módulo no curso “Data journalism and visualization with free tools” do Centro *Knight* de Jornalismo nas Américas, Jan Diehm explica que narrativa de dados se difere do uso de dados na ciência, quando há um distanciamento do público geral e não acadêmico, e de arte com dados, que, apesar de visualmente atraente, não costuma transmitir mensagens com facilidade. A narrativa de dados estaria localizada, então, em algum ponto entre esses dois pólos do espectro, combinando aspectos da função de dados acadêmicos com as características observadas na aplicação artística. Nesse caso, o objetivo é alcançar uma audiência a mais ampla possível e transmitir uma mensagem clara.

A palestra “Things we've learned from telling the ‘fun’ data stories” resume o que é o *The Pudding* e diferencia os tipos de ensaios visuais que a equipe publica: os temas tidos como sérios e aqueles que são vistos como divertidos. Enquanto os temas sérios incluem saúde, emprego, eleições, imigração e direitos reprodutivos, os temas divertidos, pelos quais o *The Pudding* é mais conhecido, abordam assuntos da cultura *pop*, como música e esportes. A palestra também explica que essa amplitude de assuntos só é possível devido à independência editorial do *The Pudding*, que, por sua vez, só existe por conta do modelo financeiro baseado no *Polygraph*, como explicado anteriormente.

Outra apresentação de nível básico, compreensível para iniciantes na área, é “Develop your data-driven storytelling chops with four simple principles”²⁹. Nesta palestra, o jornalista-engenheiro Ilia Blinderman explora algumas matérias do *The Pudding* e fala sobre os ensaios visuais sob a perspectiva de quatro princípios para a construção de narrativas de dados. Estes princípios são o foco da conferência, e, por esse motivo, são explicitados pelo mediador do evento antes da participação de Blinderman, que funciona como um breve estudo de caso.

Os quatro princípios podem ser aplicados em narrativas como os ensaios visuais, reportagens de dados que dão destaque para a visualização, ou mesmo para a apresentação de dados em reuniões empresariais, como é exemplificado pelo mediador. A primeira das

²⁸ Disponível em: https://www.youtube.com/watch?v=XTvX_wjQwRI. Acesso em: 27 jun. 2023.

²⁹ Disponível em: <https://www.youtube.com/watch?v=yOp3Y7nf6Hc>. Acesso em: 28 jun. 2023.

recomendações é simplesmente contar uma história, pensando se é uma narrativa exploratória, que incentiva o leitor a chegar às suas próprias conclusões, ou explanatória, em que o leitor é guiado para observar os resultados da análise feita pelo autor.

O segundo princípio é deixar os dados guiarem a parte visual. Isso é fundamental para transmitir informações de forma clara, adequando os gráficos às funções da análise. A interatividade aparece como terceiro princípio, visto que é uma forma de engajar a audiência e permitir explorações paralelas, ainda que seja uma matéria explanatória. O último princípio é ter cuidado com o *design*. Enquanto no *The Pudding* as visualizações são construídas de forma personalizada e cada criação parece única, a recomendação é utilizar modelos prontos no caso de não haver habilidade de programação ou artística. O importante é manter a consistência entre os elementos, como cor, fontes e tamanhos utilizados, e manter a audiência focada, evitando componentes que distraem o leitor.

Partindo de conceitos amplos sobre as narrativas de dados e focando em aspectos mais específicos, “Shaping ideas into data stories”³⁰ aborda o que é ser jornalista-engenheiro e como diferentes caminhos profissionais podem levar a essa área – a palestrante Amber Thomas partiu da Biologia Marinha e, mesmo assim, encontrou pontos de convergência entre a carreira anterior e o trabalho no jornalismo de dados. Além disso, a apresentação explora caminhos para encontrar pautas interessantes, partindo do princípio de que qualquer pessoa é criativa e que as ideias surgem do cotidiano.

Entre as principais sugestões, estão: seguir ideias que têm início com uma dúvida ou curiosidade; e selecionar temas pelos quais se é apaixonado, como é a prática frequente do *The Pudding*. A fase de *brainstorming* deve incluir pesquisas básicas para saber se a ideia é original e se há dados disponíveis para desenvolver o tema, ou como será montada a base de dados, quando essas bases não são encontradas já prontas. Thomas sugere alguns lugares para encontrar esses dados, como *sites* governamentais, Wikipédia e o *Kaggle*, um portal sobre ciência de dados que reúne conjuntos de dados disponibilizados por usuários, cursos de linguagens de programação e aprendizado de máquina e competições na área. O *Kaggle* é mais um dos sintomas das práticas colaborativas observadas na ética *hacker* e, conseqüentemente, no imaginário dos jornalistas de dados.

No caso de não encontrar dados que se encaixem na pauta, há, ainda, a possibilidade de construir uma base de dados que sirva ao tema. Observando as matérias do *The Pudding*, dentre

³⁰ Disponível em: <https://www.youtube.com/watch?v=J3Awc4HxDK0>. Acesso em: 27 jun. 2023.

as principais fontes de dados utilizadas estão: órgãos governamentais; pesquisas de outras agências de notícias ou pesquisas acadêmicas; APIs³¹ (Interfaces de Programação de Aplicação, ou *Application Programming Interfaces*, no original) de organizações como *Wikipedia*, *Spotify*, *Reddit*, *IMDb* e *New York Times*; listas da *Billboard*; e coleta própria, seja por meio de observação ou raspagem de dados de fontes variadas.

A palestra reforça também a necessidade de pensar eticamente se os dados devem ser coletados ou utilizados, e que o uso de visualização de dados é interessante para explicar temas que são delicados ou de difícil compreensão. Utilizar recursos visuais para facilitar o entendimento de um assunto complexo foi o caso em “The birthday paradox experiment”³². A matéria é um experimento interativo atualizado a cada leitor que acessa a página e informa sua data de aniversário, com o objetivo de explicar o paradoxo do aniversário – para que pelo menos duas pessoas dentro de um grupo tenham a mesma data de aniversário, basta reunir 23 pessoas.

Por último, em “From story-time to send off”, a jornalista-engenheira Caitlyn Ralph aborda o processo editorial do *The Pudding* e o conceito de ensaio visual, explicando que a necessidade de criar um vocabulário específico é justificada porque as novas possibilidades na área não se encaixam nas categorias já estabelecidas. Além disso, ela reforça a importância da diversidade profissional e acadêmica para a construção da equipe de jornalistas-engenheiros, considerando que, apesar das diferentes especialidades, é possível formar um grupo de generalistas.

2.2 Gênero e Número

A Gênero e Número³³ (GN) está em atividade desde 2016 e utiliza dados para produzir reportagens, conteúdo audiovisual, pesquisas e relatórios com o foco em questões de gênero e raça. Apesar de sediada no Rio de Janeiro, a redação produz remotamente e é membro da Associação de Jornalismo Digital (Ajo). A GN busca investigar desigualdades no Brasil e, a partir de suas produções, colaborar com debates em prol dos direitos humanos e da equidade. Além de declarar a transparência como um dos valores do jornal, a GN adota a licença *creative*

³¹ As APIs, nesses casos, são interfaces que permitem ao usuário acessar de forma simplificada informações que, pelos métodos tradicionais, exigiriam um conhecimento sobre a implementação das funções ou dos dados desejados. APIs também podem ser utilizadas para facilitar a conexão entre diferentes aplicações.

³² Disponível em: <https://pudding.cool/2018/04/birthday-paradox/>. Acesso em: 28 jun. 2023.

³³ Disponível em: <https://www.generonumero.media>. Acesso em: 8 nov. 2023.

commons que permite o compartilhamento dos conteúdos, desde que sejam dados os créditos à fonte.

Figura 2 – Página inicial da Gênero e Número



Fonte: Página inicial da Gênero e Número em captura de tela realizada pela autora. Disponível em:

<https://www.generonumero.media>. Acesso em: 8 nov. 2023.

Com foco na diversidade, a equipe do jornal conta com 11 mulheres e um homem. A separação dos cargos é declarada no *site*: Natália Leão é diretora de pesquisa e projeto; Marília Ferrari é diretora de design e inovação; Vitória Régia da Silva é gerente de jornalismo; Maria Martha Bruno é gerente de audiência e distribuição; Adriana Costantin é gerente financeira; Aline Gatto Boueri é editora; Victória Sacagami é designer de informação; Schirlei Alves é repórter; Michelle Cortes é analista de captação e desenvolvimento institucional; Diego Nunes da Rocha é analista de dados; Miriã Damasceno é estagiária de comunicação; e Sabrina Mendes é estagiária de design. Todos os cargos cujas atividades são jornalísticas são de profissionais formados na área, seja na graduação de Jornalismo ou de Comunicação Social.

A GN, a partir de 2022, passou a ser também uma Associação, com o objetivo de “entregar ao mundo uma organização perene, com maior potencial de sucessão e longevidade, governança mais compartilhada e tomadas de decisão coerentes com seus objetivos”, de acordo com o *site*, incluindo novos profissionais na equipe e acrescentando atribuições a pessoas que já faziam parte da equipe anterior. Observando, então, os cargos da Associação Gênero e Número, temos o Conselho Fiscal, com três novos funcionários; 21 Associadas, mesclando novas e antigas colaboradoras; e a Diretoria Executiva, formada exclusivamente por

profissionais da equipe do jornal: Natália Leão é a presidente da Associação; Vitória Régia da Silva é a vice-presidente; Adriana Costantin é tesoureira; Maria Martha Bruno é secretária de articulação; e Marilia Ferrari é secretária de dados abertos e gestão da informação.

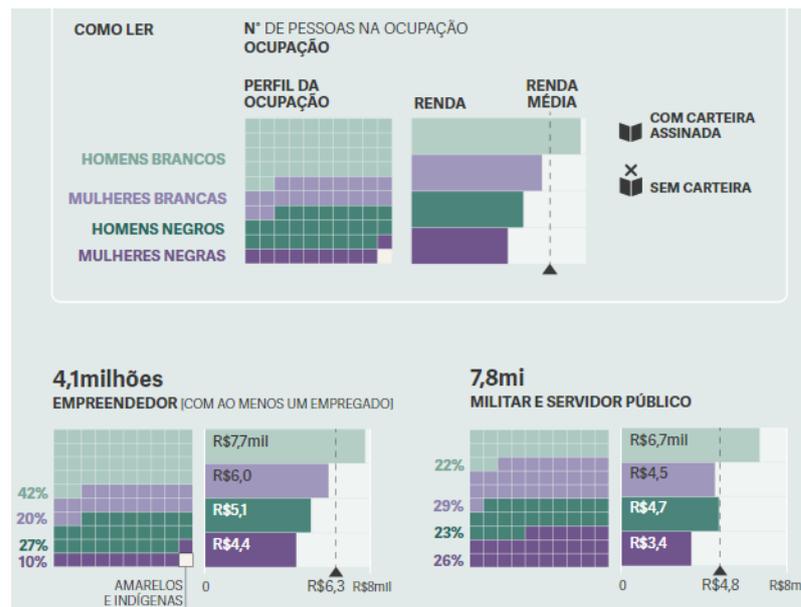
Além das equipes do GN e da Associação Gênero e Número, mais de 15 colaboradores já participaram de reportagens e outras publicações com o selo GN – todos são jornalistas e/ou profissionais envolvidos com as pautas trabalhadas, como a pesquisadora Ayra Dias do Movimento Mulheres Negras Decidem e diversas pesquisadoras do Grupo de Estudos Multidisciplinares da Ação Afirmativa. Entre os autores listados pelo jornal, também é possível encontrar outros jornais, *sites* ou agências, como a Revista Azmina e a Agência Diadorim, que fizeram projetos em parceria com a GN.

O diferencial da GN é trabalhar especificamente as pautas de gênero, raça, grupos minorizados e temáticas ambientais no Brasil a partir da análise de dados, e essas produções variam em formato. Além das entrevistas e dos artigos, que, de forma geral, fogem do escopo do trabalho com dados, as publicações se separam entre reportagens e vídeos ou documentários.

As reportagens de dados seguem um formato mais próximo do tradicional, dando destaque para os elementos textuais e distanciando-se da ideia dos ensaios visuais, apresentando gráficos simples e estáticos, mas guiando a narrativa a partir das análises feitas. Sempre são apontadas as fonte dos dados, e as bases são disponibilizadas ao fim das matérias, reunidas também na seção Dados Abertos³⁴ do *site*. Em alguns casos, uma seção chamada de metodologia é incluída para explicar a obtenção dos dados e as análises feitas, bem como possíveis falhas na apuração.

³⁴ Disponível em: <https://www.generonumero.media/dados-abertos/>. Acesso em: 9 nov. 2023.

Figura 3 – Exemplo de gráfico da Gênero e Número



Fonte: Captura de tela da matéria “Homens brancos no topo” realizada pela autora. Disponível em: <https://www.generonumero.media/reportagens/homens-brancos-topo/>. Acesso em: 28 jun. 2023.

As editorias trabalhadas no jornal incluem política, trabalho e direitos reprodutivos – pauta que levou a Gênero e Número a desenvolver o documentário Verde-Esperanza em parceria com a Filmes da Fonte em 2022, observando o aborto legal na América Latina. Este produto está diretamente relacionado à campanha de apoio³⁵ da GN, que, até novembro de 2023, conta com 42 assinantes, chegando a pouco mais de mil reais mensais. São oferecidas cinco opções de apoio – de 12, 25, 50, 70 e 100 reais – com o objetivo de lançar nacionalmente o documentário, com exhibições gratuitas e convites especiais para os apoiadores.

Para cada nível de patrocínio, as recompensas são ajustadas: todos garantem o nome registrado no *site* do filme, mas, a partir de 50 reais, o apoiador recebe um cartaz do filme em casa, acrescido de outros brindes da divulgação do filme para aqueles que doam pelo menos 70 reais mensais, enquanto a faixa mais alta da campanha garante, também, duas obras de arte que valorizam o protagonismo negro – “Travessias”, de Luna Bastos, e “Somos muitas”, de Heloísa Hariadne.

A campanha também possui uma categoria especial que não é relacionada ao documentário, direcionada para estudantes e com o valor de doação começando em cinco reais, que é o mais baixo permitido pela plataforma. Essa forma de apoio garante uma vaga em um aulão de jornalismo de dados oferecido por profissionais da equipe GN a cada dois meses, de

³⁵ Disponível em: <https://www.catarse.me/generonumero>. Acesso em: 9 nov. 2023.

acordo com a demanda de patrocínios. Esse aulão inclui todas as etapas do trabalho com dados, partindo da elaboração da pauta, passando pela coleta e análise dos dados, até a criação das visualizações.

Fora a variedade nos tipos de conteúdos produzidos, a GN possui sete projetos listados no *site*, todos relacionados às temáticas que pautam o jornal. “Racismo à Brasileira”³⁶ reúne relatos de pessoas negras sobre casos de racismo que sofreram, trazendo a situação vivida e o impacto que isso trouxe, bem como *webstories* que contextualizam as relações raciais no Brasil. O projeto ainda está em desenvolvimento, sendo disponibilizado um formulário para que mais pessoas contem suas histórias.

“Câncer de Mama Hoje”³⁷ foi publicado durante o Outubro Rosa de 2022 e é uma parceria com o Instituto Avon, apresentando dados do Sistema Único de Saúde (SUS) sobre o câncer de mama no país ao longo da pandemia de Covid-19. O objetivo é alertar para os perigos do diagnóstico tardio e a necessidade de maiores investimentos dos governos para o tratamento público da doença. O projeto conta com reportagens, relatórios com a análise dos dados e *link* para as planilhas utilizadas na apuração.

Partindo para um recorte de gênero no contexto profissional, “Open Box da Ciência”³⁸ apresenta um levantamento de dados coletados da plataforma Lattes sobre pesquisadoras brasileiras que se destacaram por produtividade acadêmica e/ou premiações em suas áreas. Publicado em 2020, o projeto inclui visualizações interativas, reportagens, a compilação de dados que foram utilizados na apuração, uma breve biografia de cada uma das pesquisadoras e uma extensa seção de metodologia. Também é disponibilizada uma calculadora que mede, a partir dos dados informados pelo usuário, em que lugar da “Escala de Produção Científica” o leitor está.

Ainda no contexto profissional, “Mulheres no Jornalismo”³⁹ é uma pesquisa realizada em parceria com a Associação Brasileira de Jornalismo Investigativo (Abraji) e com apoio do Google News Lab, tendo como foco os desafios enfrentados por mulheres jornalistas no exercício da profissão, como assédio e violências sofridos durante a rotina de trabalho. Após coletar as respostas de 477 mulheres de 271 veículos diferentes, a pesquisa foi compilada em um relatório.

³⁶ Disponível em: www.racismo.generonumero.media/. Acesso em: 9 nov. 2023.

³⁷ Disponível em: <https://institutoavon.org.br/outubrorosa/cancerdemamahoje/>. Acesso em: 9 nov. 2023.

³⁸ Disponível em: <https://www.openciencia.com.br>. Acesso em: 9 nov. 2023.

³⁹ Disponível em: <https://www.mulheresnojornalismo.org.br>. Acesso em: 9 nov. 2023.

Dando enfoque às relações entre religião e o cenário político brasileiro, dois projetos apresentam dados sobre o tema: “Religião e Poder”⁴⁰ e “Reino Sagrado da Desinformação”⁴¹. “Religião e poder” é uma plataforma que reúne dados abertos, pesquisas, artigos e reportagens sobre a religião na política institucional e a atuação de agentes políticos com identidade religiosa nas três esferas do poder público, e foi produzido em parceria com o Instituto de Estudos da Religião, que tem alimentado a plataforma com dados de pesquisas em andamento.

O segundo projeto, “Reino Sagrado da Desinformação”, apresenta em gráficos e reportagens o Brasil de 2019 a partir da análise das relações entre mídia, política e religiões no país, e, ainda que coordenado pela GN, contou, também, com a participação de profissionais externos, como pesquisadores e cientistas de dados. Foi construído um banco de dados a fim de construir um panorama nacional e regional sobre as chamadas “ideologias de gênero” e a desinformação.

Também considerando manifestações políticas, “Violência Contra LGBTQ+”⁴² é um projeto construído a partir de entrevistas realizadas em janeiro de 2019 com 400 respondentes das capitais Rio de Janeiro, Salvador e São Paulo, que resultou em uma pesquisa sobre a violência contra pessoas LGBTQIAPN+ no contexto eleitoral e pós-eleitoral. Contando com o apoio da Fundação Ford, os resultados são disponibilizados em forma de relatório, em PDF, mas os dados utilizados também estão liberados para *download*.

2.3 Uma análise comparativa

Após observar de forma isolada os dois jornais, é importante realizar uma análise comparativa para recortar, dentre as diversas formas de transparência utilizadas por cada um, aquelas que têm maior relevância para a pesquisa. Essa comparação também tem o objetivo de mostrar que, mesmo em jornais bem diferentes e apesar da enorme diferença na quantidade de informações oferecidas por cada um deles, elementos parecidos ou complementares podem ser aproveitados no contexto do aprendizado do jornalismo de dados.

Com essa finalidade, serão considerados os critérios do Protocolo de Transparência de Gehrke (2021), abordado em detalhe no capítulo anterior, sob a lente das categorias didáticas

⁴⁰ Disponível em: <https://religioepoder.org.br>. Acesso em: 9 nov. 2023.

⁴¹ Disponível em: <https://www.reinodadesinformacao.com.br/>. Acesso em: 9 nov. 2023.

⁴² Disponível em: www.violencialgbt.com.br. Acesso em: 9 nov. 2023.

de transparência sugeridas anteriormente: institucional, editorial-metodológica e tecnológica. Apesar da nossa categoria de interesse ser a editorial-metodológica, a separação entre estes estratos não é hermética e é importante levar em conta o contexto de produção. Assim, começamos a análise pela camada institucional, que trata dos valores organizacionais do jornal. Os parâmetros do Protocolo que se encaixam nesta categoria são: contato; fontes de financiamento; manual de redação e princípios editoriais; política de privacidade; política de diversidade.

Quadro 3 – Parâmetros na transparência institucional

Parâmetros	<i>The Pudding</i>	Gênero e Número
Contato	Há <i>e-mails</i> disponíveis para contato no <i>site</i> . Não há <i>ombudsman</i> nem área para comentários nas matérias.	Há uma página “Fale Conosco” dentro do <i>site</i> , mas está fora do ar. Não há <i>ombudsman</i> nem área para comentários nas matérias.
Fontes de financiamento	É clara a relação com o <i>Polygraph</i> , projeto que viabiliza financeiramente o <i>The Pudding</i> , e o direcionamento de pelo menos parte da verba arrecadada no programa de apoiadores – a contratação de <i>freelancers</i> .	Com exceção do programa de apoiadores, que é direcionado principalmente para o lançamento do documentário Verde-Esperanza, não há informações sobre o financiamento da GN.
Manual de redação e princípios editoriais	A seção que fala sobre o <i>The Pudding</i> explicita os princípios do jornal e o foco em ter uma equipe formada por jornalistas-engenheiros.	A transparência está entre os valores da GN, que declara em detalhe a missão, a visão e o propósito do jornal.
Política de privacidade	Possui uma página dedicada à política de privacidade, explicando os dados coletados dos leitores e como eles são utilizados.	Não declara a política de privacidade do <i>site</i> .
Política de diversidade	Não declara política de diversidade, mas apresenta reportagens com recortes raciais, de classe e gênero.	Tem a diversidade como um dos valores do jornal, possui equipe majoritariamente feminina e trabalha pautas que buscam revelar a desigualdade no país.

Fonte: Elaborado pela autora a partir dos parâmetros do Protocolo de Transparência (Gehrke, 2021) e da análise dos jornais.

Na camada editorial-metodológica, em que são abordados os métodos e procedimentos das reportagens, estão os seguintes parâmetros do Protocolo de Transparência no Jornalismo Guiado por Dados: assinatura, perfil e expertise; correção de erros; tipo de conteúdo; data de publicação; documentos; fontes de dados; graus de incerteza; marcas de apuração; metodologia. Esta é a categoria de maior interesse para a pesquisa, além de ser a forma à qual normalmente se refere ao debater a transparência no jornalismo, como visto anteriormente.

Quadro 4 – Parâmetros na transparência editorial-metodológica

Parâmetros	<i>The Pudding</i>	Gênero e Número
Assinatura, perfil e expertise	Os conteúdos são assinados e os autores possuem página com informações biográficas e profissionais.	Os conteúdos são assinados e os autores possuem página com informações biográficas e profissionais.
Correção de erros	O jornal não possui seção dedicada às correções e as matérias não indicam onde os leitores podem reportar erros.	O jornal não possui seção dedicada às correções e as matérias não indicam onde os leitores podem reportar erros.
Tipo de conteúdo	Os conteúdos patrocinados ou realizados em parcerias com outras empresas são sinalizados.	Os conteúdos patrocinados ou realizados em parcerias com outras empresas são sinalizados. Também há uma sinalização para artigos de opinião, com indicação de autoria e <i>link</i> para uma breve biografia/especialidade dos autores.
Data de publicação	Algumas matérias só possuem data de publicação na página principal e no <i>link</i> , mas não no corpo do texto. No caso de matérias que foram atualizadas, a data em que as mudanças foram feitas é indicada.	Todas as matérias possuem a data de publicação no corpo do texto.
Documentos	Não costumam usar documentos como fonte, porque trabalham mais com bases originais.	Principalmente quando abordam documentos legais, o <i>link</i> é disponibilizado nas matérias.
Fontes de dados	As fontes sempre são identificadas. As bases de dados originais são disponibilizadas em formatos diversos para análise.	As fontes sempre são identificadas. As bases de dados já trabalhadas e filtradas são disponibilizadas em muitos casos, o que nem sempre acontece com as originais.

Parâmetros	<i>The Pudding</i>	Gênero e Número
Graus de incerteza	As escolhas feitas ao longo da análise de dados é frequentemente indicada na metodologia das matérias.	Há o apontamento de lacunas principalmente em dados coletados por órgãos oficiais, alertando o leitor sobre possíveis subnotificações e categorias que não são bem definidas na coleta dos dados.
Marcas de apuração	Além das informações sobre o trabalho com dados, não é comum a presença de marcas de apuração.	Sempre é indicado o uso da LAI, quando aplicável.
Metodologia	As matérias possuem seções de metodologia e o <i>site</i> reúne tutoriais sobre os processos do trabalho com dados.	Nem todas as matérias possuem uma seção de metodologia.

Fonte: Elaborado pela autora a partir dos parâmetros do Protocolo de Transparência (Gehrke, 2021) e da análise dos jornais.

Por último, a camada da transparência tecnológica engloba apenas um dos parâmetros do Protocolo (Gehrke, 2021), reprodutibilidade e código aberto, por ser o único que explicitamente trata das ferramentas digitais que fazem parte do trabalho com dados. Só o *The Pudding* atende a este critério, ao disponibilizar códigos utilizados para a análise de dados, a construção das visualizações, bibliotecas em linguagens de programação que são utilizadas no *site* e tutoriais de como reproduzir alguns recursos explorados pela equipe de jornalistas-engenheiros.

Com esta análise comparativa, é possível notar como diferentes jornais podem manifestar as camadas de transparência em diferentes graus, e como a transparência em si é um conceito prismático, que engloba muitas práticas. Da mesma forma, observa-se a amplitude de resultados alcançáveis dentro do jornalismo de dados, variando em complexidade e dificuldade de reprodutibilidade e, conseqüentemente, aprendizado. Longe de alcançar uma padronização, o jornalismo de dados se beneficia da transparência ao oferecer uma janela que permite observar os bastidores de processos ainda distantes da formação dos profissionais brasileiros.

Os parâmetros do Protocolo de Transparência no Jornalismo Guiado por Dados (Gehrke, 2021) que fazem parte da transparência editorial-metodológica – assinatura, perfil e expertise; correção de erros; tipo de conteúdo; data de publicação; documentos; fontes de dados; graus de incerteza; marcas de apuração; e metodologia – serão abordados novamente no

próximo capítulo, em que os materiais do *The Pudding* e da Gênero e Número serão insumo para pensar o aprendizado do jornalismo de dados a partir da transparência.

3 O APRENDIZADO DO JORNALISMO DE DADOS A PARTIR DA TRANSPARÊNCIA EDITORIAL-METODOLÓGICA

Durante a graduação, tive meu primeiro contato com o jornalismo de dados ao cursar a disciplina Jornalismo Digital, no quinto período, quando foram apresentados diferentes formatos de reportagens, e, na ocasião, a professora comentou com a turma sobre um curso gratuito do Centro *Knight* de Jornalismo nas Américas, da Universidade do Texas em Austin. Durante o curso, percebi os paralelos do jornalismo de dados, que, até então, parecia muito distante da realidade que me havia sido apresentada em dois anos de graduação, com o jornalismo “tradicional”.

No semestre seguinte, pude, finalmente, cursar Jornalismo de Dados, que era uma disciplina oferecida como eletiva. A primeira avaliação era um estudo de caso de qualquer veículo que utilizasse jornalismo de dados, e a segunda avaliação era construir uma reportagem baseada em dados. Apesar das aulas no curso e na disciplina, partir de exemplos reais, explorados em sala de aula durante a primeira avaliação, foi o que deu materialidade ao que estava idealizado, possibilitando ver o jornalismo de dados na prática e aplicado a temáticas variadas.

Da mesma forma, quando fui monitora de duas turmas da disciplina no ano seguinte, pude observar nos alunos que muitos tinham resistência em trabalhar com planilhas e números, mas que havia um aumento de interesse conforme eles eram expostos a pautas com as quais tinham mais afinidade. Mostrando o uso de dados em temáticas interessantes para eles, era mais fácil que assimilassem as técnicas observadas como análogas ao processo de apuração tradicional, e as possíveis dificuldades não pareciam ser vistas como barreiras insuperáveis.

Isso teve impacto na variedade de temas no trabalho final, que incluíam reportagens sobre a relação entre a Saga Crepúsculo e a pirataria no Brasil⁴³, a falta de representatividade em novelas brasileiras⁴⁴ e a luta pelo direito à dignidade menstrual⁴⁵. A variedade de temas significa o uso de diferentes fontes de dados e variedade nos personagens e especialistas entrevistados, enriquecendo o conjunto produzido pelas turmas.

⁴³ Disponível em: <https://uvadados.wordpress.com/2021/06/16/entre-vampiros-e-piratas/>. Acesso em: 27 nov. 2023.

⁴⁴ Disponível em: <https://uvadados.wordpress.com/2021/06/21/novelas-brasileiras-ainda-falham-na-representatividade/>. Acesso em 27 nov. 2023.

⁴⁵ Disponível em: <https://uvadados.wordpress.com/2021/06/22/direito-negado-dignidade-menstrual-nao-e-para-todas/>. Acesso em: 27 nov. 2023.

Essa experiência pessoal, em que observar o trabalho com dados foi parte fundamental no aprendizado das técnicas ao apontar diferentes caminhos da prática, se conecta com o que tem sido pesquisado sobre o ensino do jornalismo de dados na graduação, em que há dificuldade de implantar disciplinas que realmente preparem os alunos para trabalhar na área. O resultado disso são profissionais autodidatas ou que precisam procurar cursos especializados. Essa experiência também se conecta com a ideia de que a transparência jornalística pode permitir a reprodutibilidade, que colabora com a democratização do conhecimento e facilita o aprendizado do jornalismo de dados.

Para debater sobre o ensino de jornalismo de dados no Brasil, Ventura (2021) observou as Diretrizes Curriculares Nacionais (DCN) para o curso de Jornalismo, instituídas em 2013 e com prazo de dois anos para adequação. Na tese, a autora destaca a carência de disciplinas voltadas especificamente para o ensino do jornalismo de dados ou de áreas afins que envolvam estatística e matemática. A escassez de disciplinas específicas na graduação tem sido suprida por iniciativas externas e independentes, como os *Massive Open Online Course (MOOC)* ofertados pela Abraji e pelo Centro *Knight* para o Jornalismo nas Américas e como visto em vários dos portais voltados para a transparência mencionados no primeiro capítulo. Para mapear o campo, Ventura (2021) analisou grades curriculares, as dificuldades de professores com os temas mencionados e a percepção de profissionais que atuam no jornalismo de dados.

Além de Ventura (2021), Mielniczuk e Träsel (2017) também puderam observar em suas pesquisas que as competências relacionadas a jornalismo de dados acabam tangenciando disciplinas que tratam de jornalismo investigativo, jornalismo digital ou jornalismo científico, o que aponta para aspectos atribuídos ao trabalho com dados: um processo de apuração, a interseção entre práticas jornalísticas e ferramentas computacionais, e a ideia da reprodutibilidade, respectivamente.

É interessante notar que, dentre as dificuldades observadas nos alunos da graduação, segundo a pesquisa de Ventura (2021), 29% dos professores relataram que falta conhecimento prévio, mas isso não era restrito a questões tecnológicas e áreas como a matemática – muitos deles apontaram a falta de competências jornalísticas como apuração, elaboração de pauta, processo de investigação e conhecimento político e social. Essas respostas corroboram as ideias de que o surgimento e a implementação de novas tecnologias no processo jornalístico não alteram conceitos como interesse público, fato e acontecimento (Karam, 2014), e de que o uso de dados parece ser um dos modos de ajuste do jornalismo à cibercultura e à economia do conhecimento (Träsel, 2014a).

O jornalista de dados continua sendo um profissional de imprensa e necessita refletir sobre a forma como a mídia opera (Araújo, 2016), e esse desdobramento não representa uma transformação completa do jornalismo tido como tradicional, e sim uma adesão de ferramentas e rotinas produtivas. As condições que tornam o trabalho com dados parte do cotidiano, justificando a necessidade de aprendermos as técnicas do jornalismo de dados, são as mesmas que transformaram nossa cognição e, conseqüentemente, nossas formas de aprendizado: a revolução digital, a partir do quarto final do século XX, opera modulações nos saberes e nas subjetividades, dando novo fôlego aos estudos sobre a cognição, que, pela tradição da filosofia ocidental, é tarefa prioritária ou exclusiva dos processos mentais (Régis, 2022).

Neste sentido, Alonso (2018) advoga pela ideia do *homo sapiens*, termo que representa as gerações digitais, os nascidos após 1990, que estão em contato permanente com a computação em rede e a cultura cibernética. Mas, para além de pensar em uma questão geracional, o *homo sapiens* metaforiza as estratégias para lidar com a abundância de dados disponíveis e as mudanças sociais relacionadas à globalização e ao mundo cibernético (Alonso, 2018). O autor também exemplifica o movimento *hacker* como modo de ação do *homo sapiens*, que, como vimos anteriormente, se alinha às ideias de colaboração, compartilhamento e autodidatismo muito populares no imaginário e nas práticas dos jornalistas de dados.

De acordo com Régis (2022), as práticas com mídias digitais favorecem o desenvolvimento de competências relacionadas ao letramento clássico – como a interpretação de textos, a lógica, a atenção focada e a capacidade de resolução de problemas –, mas, também, de competências que são potencializadas por recursos multimidiáticos e hipertextuais – como a colaboração para a produção de conteúdo, a exploração de ambientes midiáticos, a atenção dividida, a autonomia, o aprendizado por tentativa e erro, a descentralização e a criatividade. Todos elementos se mostram valiosos no aprendizado do jornalismo de dados, que alinha práticas tradicionais do âmbito comunicacional e técnicas específicas do meio computacional.

Além disso, é importante pensar que o conhecimento está sempre em contexto de mudança (Alonso, 2018). Mielniczuk e Träsel (2017) afirmam que, devido à constante evolução tecnológica, é mais importante ensinar processos cognitivos para lidar com as etapas de apuração, redação e edição de notícias a partir de bases de dados do que se ater a ferramentas específicas, garantindo maior flexibilidade aos profissionais. Da mesma forma, Gehrke (2021) entende que os jornalistas de dados não necessariamente devem aprender uma linguagem de programação, apesar do conhecimento técnico resultar em facilidades e autonomia, mas que são fundamentais o pensamento crítico, a segurança para acessar fontes documentais utilizadas

no trabalho com dados e o conhecimento para identificar limites e possibilidades desse tipo de jornalismo.

Também alinhado com essa ideia, Salaverría (2011) reforça que a formação dos alunos de jornalismo não deve ser focada em uma qualificação instrumental avançada, mas no desenvolvimento da habilidade de inovar e criar produtos editoriais de qualidade, de forma a aproveitar o potencial de comunicação das ferramentas disponíveis. De acordo com o autor, ainda que seja importante oferecer treinamentos de plataformas e dispositivos específicos, essa é a parte menos importante do aprendizado, sendo fundamental mostrar que os recursos digitais são inerentes ao trabalho jornalístico.

Também Berret e Phillips (2016) advogam por competências abrangentes na hora de estudar o jornalismo de dados, em vez de focar em ferramentas ou linguagens de programação específicas. Segundo Fry (2008), a habilidade mais importante e menos técnica para entender dados é fazer boas perguntas, e o que vai determinar o sucesso de uma visualização é a habilidade de fazer a audiência captar a mensagem que está sendo transmitida e se entusiasmar com o que foi encontrado nos dados. Isso significa que as habilidades jornalísticas ainda impactam mais do que as habilidades em ferramentas específicas para profissionais que trabalham com dados.

Como visto em Gray, Bounegru e Chambers (2012), o jornalismo de dados é parte de uma longa tradição, mas que responde a novas circunstâncias e condições. E justamente por representar um conjunto de práticas relativamente recentes, ainda não há uma formalização do ensino de jornalismo de dados. Mielniczuk e Träsel (2017) sugerem a criação de laboratórios de ensino e extensão como modelos mais adequados de ensino para processos de apuração e edição que exigem o uso intensivo de técnicas computacionais, mas que, idealmente, as práticas de coleta, análise e apresentação de números deveriam ser parte do cotidiano em todas as modalidades de reportagem, visto que as bases de dados têm a relevância de fontes humanas atualmente.

Em Ventura (2021), 68% dos professores acreditam que o jornalismo de dados deve ser uma disciplina obrigatória na graduação, apontando como justificativas a relevância do tema, a especificidade da área, a necessidade da prática e a demanda do mercado de trabalho. Os argumentos dos professores que optaram pelo jornalismo de dados como optativa, correspondendo a 18% dos professores entrevistados, são direcionados ao interesse (ou a falta dele) dos alunos, trazendo também a ideia de que as técnicas específicas não são necessárias

para a formação geral de um jornalista, mas apenas para os que vão atuar em redação, além de serem voltados para exatas e tecnologia da informação.

A falta de interesse dos alunos, que pode ser um desafio em qualquer área de conhecimento, se manifesta no jornalismo de dados principalmente por conta da proximidade com os números, mas isso pode ser mitigado com o uso de metodologias ativas, promovendo o contato com jornalistas que atuam na área e permitindo que os alunos trabalhem com temas em que se sintam à vontade (Ventura, 2021). Dessa forma, utilizar recursos de transparência como ferramentas de aprendizado pode impactar positivamente o contato de estudantes e profissionais que vão estudar o jornalismo de dados.

Considerando o atual cenário do ensino do jornalismo de dados no Brasil e das mudanças na cognição e nas práticas profissionais, o próximo subcapítulo é dedicado a explorar os recursos de transparência editorial-metodológica do *The Pudding* e da Gênero e Número a partir da análise comparativa realizada no capítulo anterior, tendo em mente as diferentes etapas do trabalho com dados. Assim como o estudo da ética, normalmente associada à reflexão teórica, precisa mobilizar situações práticas com o objetivo de trazer conceitos abstratos para a práxis cotidiana (Fittipaldi, 2019), o estudo do jornalismo de dados se torna mais frutífero ao serem observados produtos que concretizem as diferentes técnicas de apuração e visualização.

Essa tentativa de sistematizar a transparência como recurso de aprendizado tem em vista tanto o profissional autodidata, que é a tendência no campo, quanto o aluno de graduação. Por essa razão, a pesquisa foca na transparência editorial-metodológica, que trata principalmente dos processos de apuração, coleta e análise dos dados, como abordado anteriormente. Mesmo assim, a análise de reportagens de dados a partir da transparência editorial-metodológica é uma etapa secundária, depois de ter sido feita a análise do jornal em si de acordo com os 15 parâmetros do Protocolo, comprovando que o veículo jornalístico selecionado é, sim, transparente.

3.1 A transparência das etapas do trabalho com dados

Como estabelecido no primeiro capítulo da dissertação, a transparência como prática jornalística significa revelar os processos produtivos e estruturas organizacionais. Por ser um termo muito amplo, esta pesquisa adotou três categorias não-canônicas e didáticas para

possibilitar uma análise voltada para o aprendizado do jornalismo de dados: a transparência institucional, a editorial-metodológica e a tecnológica.

Sem esquecer que essas divisões não são herméticas, ressalta-se que o recorte desta seção é a transparência editorial-metodológica. Sendo aquela que mais se conecta com o senso comum do que é transparência e que se apresenta por meio da abertura de procedimentos de reportagem e de etapas de apuração, é o tipo de transparência que se mostra mais útil para a formação de outros profissionais, passando, de forma geral, pelo detalhamento do trajeto percorrido até a publicação.

O *The Pudding* e a *Gênero e Número*, analisados no capítulo anterior de acordo com o Protocolo de Transparência no Jornalismo Guiado por Dados (Gehrke, 2021), são dois exemplos muito diferentes de jornalismo de dados e de diferentes níveis de transparência. Ainda assim, partindo da hipótese de que a transparência colabora com o aprendizado do jornalismo de dados, vamos cruzar as etapas do trabalho com dados com os recursos disponibilizados pelos jornais digitais, focando nos parâmetros da transparência editorial-metodológica.

Como foi observado por meio da análise comparativa, dizer se um veículo é transparente ou não é, na verdade, considerar pontos de análise de acordo com o objetivo da transparência. Pensando no aprendizado do jornalismo de dados, é necessário, depois de uma investigação geral do jornal, uma investigação específica de qual produto pode ser utilizado com esta finalidade. Um recurso de aprendizado, nesse contexto da transparência-editorial metodológica, é, principalmente, aquele que revela suas **fontes de dados** e a **metodologia**. Como observado por Gehrke (2021) na elaboração do Protocolo, os parâmetros podem possuir pesos diferentes diante de recortes diferentes, e esses dois elementos se destacam nesta pesquisa.

Voltando à análise comparativa, é possível observar que a *Gênero e Número* é menos transparente que o *The Pudding*, tanto em número de parâmetros cumpridos como em regularidade em cumprir parâmetros. Assim, para encontrar reportagens que atendam ao objetivo da pesquisa, o caminho mais simples é ir diretamente na seção Dados Abertos do *site*, que reúne as bases consultadas ou construídas pela equipe, separadas de acordo com as reportagens em que foram analisadas. Ou seja, esse é o recorte das reportagens que possuem não só a indicação da fonte de dados, mas também o *link* direto para essas bases, que podem ser baixadas e reaproveitadas.

“Fases avançadas do câncer de mama consomem 80% dos recursos do SUS para tratamento da doença”⁴⁶ está na página de Dados Abertos e possui uma seção dedicada à metodologia, então se qualifica para ser analisada nos parâmetros da transparência editorial-metodológica. A matéria aborda os diagnósticos e tratamentos de câncer de mama durante o primeiro ano da pandemia e possíveis impactos disso no Sistema Único de Saúde (SUS).

Quadro 5 – Parâmetros na transparência editorial-metodológica na matéria da Gênero e Número

Parâmetros	“Fases avançadas do câncer de mama consomem 80% dos recursos do SUS para tratamento da doença” (Gênero e Número)
Assinatura, perfil e expertise	A matéria é assinada por Aline Gatto Boueri, Maria Martha Bruno e Natália Leão e possui nota de rodapé indicando a função de cada uma – repórter colaboradora, diretora de conteúdo e coordenadora de dados, respectivamente.
Correção de erros	Não há indicação de que houve correção de erros na matéria.
Tipo de conteúdo	Matéria realizada em parceria com o Instituto Avon, como parte do projeto Câncer de Mama Hoje.
Data de publicação	20 de outubro de 2021.
Documentos	Entre os documentos analisados, estão uma nota técnica assinada por médicos, pesquisadores e pelo Conselho Nacional de Secretários de Saúde, duas leis e um relatório do Instituto Brasileiro de Geografia e Estatística (IBGE). Todos possuem <i>link</i> direto para o material original.
Fontes de dados	Sistema de Informações Ambulatoriais (SIA-SUS), Instituto Brasileiro de Geografia e Estatística (IBGE), Ministério da Saúde, Secretarias Estaduais de Saúde. Estão disponíveis duas bases com os dados da reportagem.
Graus de incerteza	A matéria aponta a falta de dados oficiais no Brasil sobre o câncer de mama em pessoas trans e não-binárias. O recorte temporal dos dados coletados e as falhas nos registros de dados públicos são destacados na seção de metodologia.
Marcas de apuração	O uso da Lei de Acesso à Informação é apontado na nota metodológica, assim como alguns dados que foram descartados na apuração e dificuldades na comunicação com algumas Secretarias Estaduais de Saúde.
Metodologia	A reportagem possui uma seção dedicada à metodologia que indica fontes dos dados, os recortes temporais da coleta, informações sobre os pedidos feitos via LAI e escolhas na análise

⁴⁶ Disponível em: <https://www.generonumero.media/reportagens/fases-cancer-mama-sus/>. Acesso em: 29 nov. 2023.

Parâmetros	“Fases avançadas do câncer de mama consomem 80% dos recursos do SUS para tratamento da doença” (Gênero e Número)
	dos dados.

Fonte: Elaborado pela autora a partir dos parâmetros do Protocolo de Transparência (Gehrke, 2021) e da análise da matéria “Fases avançadas do câncer de mama consomem 80% dos recursos do SUS para tratamento da doença”.

O *The Pudding*, como estabelecido anteriormente, oferece mais recursos de transparência do que a Gênero e Número, disponibilizando tutoriais que colaboram com o aprendizado de linguagens de programação, recursos de visualização interativa e outras ferramentas utilizadas no trabalho com dados. No capítulo anterior da dissertação, estes tutoriais foram analisados e categorizados a fim de colaborar com a discussão sobre a transparência no jornalismo de dados, mas, considerando que a ideia deste capítulo é trabalhar especificamente a transparência editorial-metodológica dentro do jornalismo de dados, será analisada uma matéria publicada no *site*.

Como o padrão do *The Pudding* é incluir a seção de metodologia em todas as publicações, foi selecionada a matéria “Women’s pockets are inferior”⁴⁷ para a checagem dos parâmetros do Protocolo de Transparência (Gehrke, 2021). Essa reportagem busca comparar o tamanho dos bolsos em calças jeans femininas e masculinas, e, para isso, as jornalistas-engenheiras mediram bolsos de 80 calças: para cada uma das 20 marcas analisadas, elas selecionaram dois tipos de modelagem de mesmo tamanho de cintura e registraram o tamanho de bolso da versão feminina e da masculina.

Quadro 6 – Parâmetros na transparência editorial-metodológica na matéria do *The Pudding*

Parâmetros	“Women’s pockets are inferior” (<i>The Pudding</i>)
Assinatura, perfil e expertise	Jan Diehm e Amber Thomas assinam a matéria, e o <i>link</i> nos nomes encaminha para uma página que inclui uma breve informação biográfica e a lista de matérias que cada uma delas participou no <i>The Pudding</i> .
Correção de erros	Não há indicação de que houve correção de erros na matéria.
Tipo de conteúdo	Não há indicação específica, mas é uma reportagem de dados.
Data de publicação	Agosto de 2018.

⁴⁷ Disponível em: <https://pudding.cool/2018/08/pockets/>. Acesso em: 29 nov. 2023.

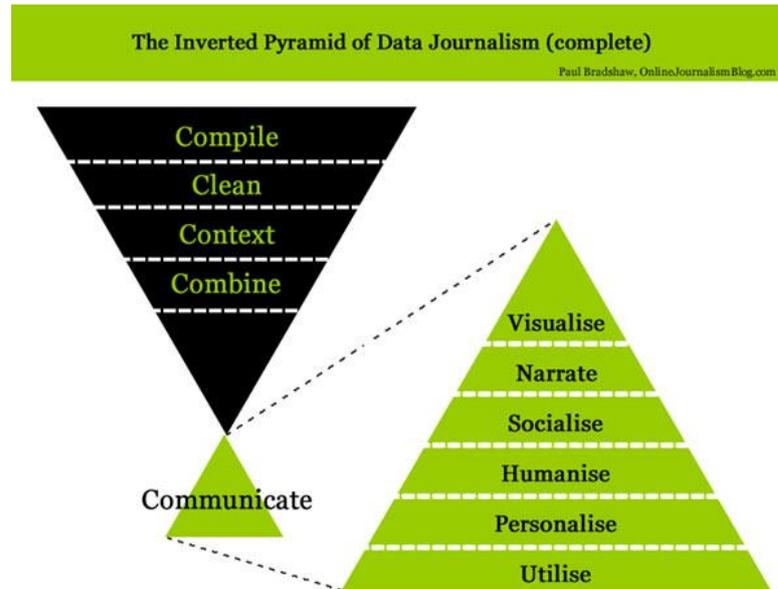
Parâmetros	“Women’s pockets are inferior” (<i>The Pudding</i>)
Documentos	Um relatório de antropometria do <i>National Center for Health Statistics</i> é indicado na seção de metodologia.
Fontes de dados	Os dados foram coletados manualmente para a matéria e estão disponíveis para <i>download</i> .
Graus de incerteza	Não há menção de graus de incerteza dos dados trabalhados.
Marcas de apuração	O texto menciona diretamente a coleta dos dados.
Metodologia	A matéria inclui uma seção de metodologia que explica o processo de coleta de dados e cuidados que foram tomados para garantir uma base de comparação dos bolsos, porque a matéria analisa modelos diferentes e marcas diferentes.

Fonte: Elaborado pela autora a partir dos parâmetros do Protocolo de Transparência (Gehrke, 2021) e da análise da matéria “Women’s pockets are inferior”.

Após registrar as práticas de transparência da Gênero e Número e do *The Pudding* e de analisar especificamente uma reportagem de dados de cada um dos jornais, tendo como recorte os parâmetros de **fontes de dados** e **metodologia**, vamos observar as etapas do trabalho com dados e desconstruir as duas matérias mencionadas. Dessa forma, é possível mobilizar, por meio da exploração de ambientes midiáticos, a autonomia, a criatividade e o aprendizado por tentativa e erro, competências observadas nos estudos sobre cognição diante das mídias digitais (Régis, 2022).

Conforme defendido ao longo da dissertação, o jornalismo de dados, por ser uma aplicação de técnicas computacionais no jornalismo, possui fases de apuração e publicação similares ao jornalismo tido como tradicional. Bradshaw (2011) sistematizou as etapas de trabalho com dados em uma pirâmide invertida, simbolizando a grande quantidade de informação que vai se afunilando no processo produtivo até ser comunicada para o público. Esta estruturação é amplamente aceita nos campos profissional e teórico do jornalismo de dados, e consiste nas fases de compilação, limpeza, contextualização, combinação e comunicação (Bradshaw, 2011). A comunicação também é desdobrada pelo autor em visualização, narração, socialização, humanização, personalização e utilização, como pode ser observado na imagem abaixo.

Figura 4 – A pirâmide invertida do jornalismo de dados



Fonte: Elaborada por Paul Bradshaw (2011).

Outra sistematização das fases do jornalismo de dados é o chamado *data pipeline*, uma metodologia desenvolvida pela Escola de Dados, iniciativa da *Open Knowledge Brasil*. Esse fluxo de trabalho pode ser feito intercalando as fases, e as etapas já finalizadas podem ser retomadas de acordo com as necessidades de apuração, mas a esquematização linear, para fins didáticos, é: pergunte, encontre, obtenha, verifique, limpe, analise e apresente (Belisário *et al.*, 2020).

Como as duas estruturas são análogas e populares na área, vamos identificar nesta seção as fases do trabalho com dados por meio de etapas numeradas, a fim de simplificar a análise das matérias da *Gênero e Número* e do *The Pudding* como recurso de aprendizado. O quadro abaixo resume cada uma dessas etapas e explicita a equivalência de fases em Bradshaw (2011) e Belisário *et al.* (2020).

Quadro 7 – Comparando as etapas do jornalismo de dados

Fases do trabalho com dados	A pirâmide invertida de Bradshaw (2011)	O <i>data pipeline</i> de Belisário <i>et al.</i> (2020)
Fase 1	compilação	pergunte; encontre; obtenha
Fase 2	limpeza; contextualização	verifique; limpe
Fase 3	combinação	analise
Fase 4	comunicação	apresente

Fonte: Elaborado pela autora a partir das estruturas de Bradshaw (2011) e Belisário *et al.* (2020).

Agora, tendo em mente essas equivalências, voltamos às reportagens analisadas por meio do Protocolo de Transparência (Gehrke, 2021) para observar como cada uma das fases do trabalho com dados se manifesta na prática. A produção de uma matéria de dados pode começar com um conjunto de dados a ser explorado na tentativa de encontrar uma pauta, mas, mesmo que a quantidade de informação seja grande, a chave para construir uma reportagem de dados é pensar nas perguntas a serem feitas (Fry, 2008), ou, como visto em Träsel (2014a), entrevistar a planilha. Nos casos observados, em vez de começar com uma base de dados, a **fase 1** nas duas matérias analisadas tem início com perguntas e, então, são compilados dados na tentativa de encontrar respostas.

Ao procurar dados a serem analisados, como foi o caso da Gênero e Número, é importante ressaltar a diferença entre dados disponíveis e acessíveis, como destacado por Gehrke (2021): alguns dados podem estar disponíveis, como tabelas em relatórios no formato PDF, mas necessitam ser convertidos para formatos como CSV ou XLS para que sejam abertos em leitores de planilhas. Além disso, os dados devem ser obtidos de fontes confiáveis e liberados para uso, mas, no caso da extração de dados em um *site* ou repositório que não tenha sido criado com esse objetivo, deve-se considerar os direitos de uso dos dados (Fry, 2008).

Na Gênero e Número, a investigação busca observar como a pandemia da Covid-19 impactou o diagnóstico e o tratamento do câncer de mama no SUS. Para isso, era necessário analisar dados de antes da pandemia para comparar os dois períodos. Como visto no primeiro capítulo, a transparência governamental é frequentemente insumo de pautas e, nesse caso, a matéria é totalmente construída em dados públicos. O levantamento inicial dos dados foi feito pelo DataSUS, utilizando a ferramenta TabNet, uma interface que permite investigar bases de dados do SUS de forma rápida por meio dos filtros disponibilizados e, posteriormente, exportar os dados em planilhas.

Este é um exemplo de pauta inédita que surgiu com a análise de dados disponíveis por meio da transparência ativa, ou seja, são informações disponíveis para qualquer pessoa que acesse a plataforma. A transparência passiva também foi utilizada na apuração, quando foram obtidos dados do Ministério da Saúde e de Secretarias Estaduais de Saúde por meio da LAI, com o objetivo de analisar o orçamento destinado ao diagnóstico e tratamento do câncer de mama.

Para montar uma base de dados original, como foi o caso do *The Pudding*, é possível utilizar procedimentos variados. Se a quantidade de dados for pequena e/ou se a coleta é feita

de forma lenta e manual, o método mais básico é abrir uma planilha e registrar os dados de forma organizada, em que cada coluna trata de uma variável e cada linha corresponde a uma entrada diferente. Para informações em maior volume e disponíveis em *sites* que permitam a extração de informações, é válido procurar a disponibilidade de *APIs* que facilitem a coleta e que possuam informações básicas de uso, ou procurar aplicativos que auxiliem na raspagem dos dados (Gray; Bounegru; Chambers, 2012).

No caso do *The Pudding*, a pergunta que deu origem à investigação dependia da coleta própria dos dados, porque não era possível encontrar a medida padronizada de bolsos de calças jeans na *internet*. Na falta de bases de dados disponíveis, montar sua própria base se conecta com a ideia de datafocar fenômenos, ou seja, colocar as informações em um formato quantificado de modo a possibilitar tabulação e análise (Mayer-Schönberger; Cukier, 2013), e essa prática é muito comum no *The Pudding*.

Partindo de uma ideia relativamente simples de comprovar ou refutar, era necessário coletar medidas de bolsos em calças jeans de marcas e modelagens diferentes, em uma quantidade que permitisse observar padrões. A matéria inclui diagramas que mostram como as medidas foram feitas, a fim de comprovar a padronização dessas medidas, visto que foram realizadas em lugares diferentes e por pessoas diferentes.

Após a **fase 1**, em que são reunidos os dados disponíveis ou montadas as bases necessárias para a investigação, chegamos na **fase 2**, de checagem, contextualização e limpeza. É esta a etapa em que são verificados os dados em outras fontes que possam corroborar ou negar as informações obtidas originalmente, seja cruzando bases diferentes ou entrevistando pessoas e órgãos oficiais que tratem do tema. Nesta fase, são analisadas as formas de coleta dos dados e os recortes feitos, ou seja, o contexto de produção dos dados, e são consideradas as possíveis limitações, que podem ser trabalhadas na apuração ou pelo menos informadas ao público.

Além disso, são considerados dicionários de dados, metadados, anexos metodológicos ou apêndices estatísticos – todos esses termos representam informações complementares aos dados que explicam detalhes sobre a coleta, o que significam os campos da planilha e como foram preenchidos (Belisário *et al.*, 2020). Também é nesta fase em que os dados são organizados e filtrados.

No caso da matéria da Gênero e Número, houve a tentativa de conferir os dados obtidos via DataSUS por meio da assessoria de imprensa das Secretarias Estaduais, mas, entre os estados analisados, só São Paulo, Pará e Rio Grande do Sul responderam. Durante a organização dos dados, as jornalistas decidiram separar informações de tratamento e diagnóstico, separando,

também, os casos de acordo com o estágio da doença. Observando mais informações sobre os dados obtidos, notou-se que havia a possibilidade de atraso em até três meses no registro dos procedimentos e que os dados referem-se à data de faturamento do serviço, e não da realização. Outro ponto importante é que não seria possível identificar o número de pacientes por meio dos dados obtidos, visto que os dados só apontavam o número de procedimentos aprovados para pagamento pelas Secretarias e cada pessoa poderia realizar mais de um procedimento.

Durante a checagem e a limpeza, dados que fogem do padrão são verificados e, caso estejam corretos, são destacados na apuração ou descartados, caso não sejam relevantes para o contexto. Isso aconteceu com as informações obtidas via LAI, que foram descartadas por tratarem do orçamento destinado, pelo Ministério da Saúde e das Secretarias Estaduais, ao diagnóstico e ao tratamento oncológico de forma geral, ou seja, não havia, nesses dados, o recorte do câncer de mama. Outros dados descartados foram os do Sistema de Informação do Câncer (SisCan), porque não havia diferenciação entre as mamografias de rastreamento ou diagnóstico, e dados do Sistema de Informações Hospitalares (SIH-SUS), porque não era possível determinar exatamente o impacto da Covid-19 nas internações e na mortalidade registrada.

Os dados que efetivamente foram usados na matéria estão disponíveis em uma pasta no *Google Drive* em dois conjuntos de planilhas, sendo possível o *download* e a cópia. Há anotações nas planilhas sobre o recorte temporal do ano em que os dados foram coletados – alguns dados vão até abril de 2021, enquanto outros chegam até junho. Uma planilha reúne os dados sobre tratamento e diagnóstico entre 2010 e 2021 e a outra apresenta o agrupamento por estágios entre 2019 e 2021, mas as duas consistem em uma aba com dados nacionais e abas específicas para os estados analisados.

A **fase 2** na matéria do *The Pudding* foi significativamente mais simples, porque os dados foram coletados manualmente e individualmente, ou seja, só havia margem para erro humano e não erro técnico. Assim, não houve a necessidade de limpeza de planilhas para o caso de dados considerados *outliers*, entradas duplicadas ou falta de dados. Os dados foram organizados desde a coleta, com o planejamento prévio de como seriam feitas as medições dos bolsos e o estabelecimento das variáveis que seriam analisadas, então também não houve descartes de dados registrados na matéria.

A preocupação com possíveis limitações dos dados também foi mitigada na **fase 1**, estabelecendo a padronização do tamanho de cintura das calças a serem analisadas de forma que fizesse sentido comparar modelagens femininas e masculinas. A planilha com os dados

utilizados está disponível no *GitHub*, acompanhada por metadados dos arquivos que explicam o que cada tipo de dado significa e o formato dos dados.

Passando para a **fase 3**, de combinação e análise, quando os dados organizados e limpos podem ser associados a outras informações pertinentes, são realizadas operações matemáticas – como médias, porcentagens e proporções – e são observadas correlações, causalidades ou padrões. Essa fase é apontada por Belisário *et al.* (2020) como a que necessita de maior rigor metodológico, que, em conjunto com práticas de transparência editorial-metodológica, torna possível a verificação, o apontamento de eventuais erros de análise e vieses.

Na matéria da Gênero e Número, destaca-se que os casos de câncer em estágio três e quatro, mais avançados, dependem de tratamentos mais caros, e, com o impacto da pandemia de Covid-19 na demora de diagnóstico da doença, os valores direcionados pelo Ministério da Saúde e pelas Secretarias Estaduais para o SUS corroboram a ideia de que o SUS, pensado com o objetivo de prevenção e detecção de doenças, se torna insustentável em períodos prolongados de austeridade fiscal – que era a realidade na época de produção da reportagem –, porque necessita de maiores investimentos públicos. O aumento de incidência nos diagnósticos de câncer de mama no Brasil e no mundo também é comparado com os dados do IBGE que se referem ao aumento da expectativa de vida, uma correlação fundamental para compreender o cenário, considerando que a doença atinge principalmente pessoas com mais de 60 anos.

Já na apuração do *The Pudding*, o tamanho dos bolsos é analisado principalmente de acordo com o tamanho médio feminino e masculino, e esse dado é comparado com o tamanho de objetos que normalmente podem ser guardados em bolsos: três celulares de marcas diferentes, caneta e um modelo de carteira específico para o bolso frontal da calça. Também é considerado o tamanho médio de mão feminina e masculina a partir de dados antropométricos dos Estados Unidos. O objetivo dessa comparação é comprovar a relevância do tema, tentando afastar a pauta de vieses de confirmação, porque o tema surgiu de uma experiência individual. Pelo mesmo motivo, é abordado também o processo histórico do surgimento dos bolsos embutidos em roupas, porque o padrão europeu a partir do século XVIII já não permitia que mulheres tivessem bolsos na roupa a fim de evitar volumes indesejados na silhueta.

Na última etapa, a **fase 4**, trata-se da apresentação dos dados. Pensando nas visualizações, frequentes no jornalismo de dados, é importante ter em mente que o objetivo é aumentar a compreensão dos leitores, com uma comunicação efetiva (Belisário *et al.*, 2020). Assim como em reportagens “tradicionais”, a reportagem de dados deve informar, e apenas apresentar os dados obtidos ou visualizações sem contextualização não cumprem essa função.

O uso de técnicas elaboradas ao lidar com dados não dispensa a construção narrativa da matéria e uma apuração cuidadosa.

Ao construir gráficos e visualizações de dados, é essencial refletir sobre quais elementos são necessários e quais são acessórios, e, em termos da construção da narrativa e clareza na leitura, deve-se pensar na coesão entre os elementos visuais, para que o leitor não tenha dificuldade em aprender a lê-los. Um exemplo dessa coesão é manter a mesma cor ao tratar de uma mesma variável em diferentes visualizações, facilitando a identificação mesmo em análises diferentes dentro da mesma reportagem.

Como observado anteriormente na seção sobre o *The Pudding*, o foco do jornal são as visualizações, que costumam ser interativas, e essa matéria não foge do padrão. Há desenhos que mostram o tamanho médio dos bolsos femininos e masculinos com o sombreado de todos os que foram medidos, ou seja, a variedade de tamanhos é explicitada visualmente ao longo de toda matéria, apesar do destaque para os tamanhos médios, que facilita a comparação. Uma parte da matéria utiliza o recurso de *scrollytelling*, em que, por meio da rolagem da tela, são exibidos diferentes elementos visuais.

A reportagem inclui uma seção que exibe todos os 80 bolsos medidos e permite filtrar de acordo com modelagem, marca e faixa de preço, separando os modelos femininos e os masculinos. Em outro ponto, é possível comparar o maior, o menor e o tamanho médio em cada modelagem, e há outra parte interativa que ilustra quais dos objetos analisados cabem ou não dentro dos bolsos. Todas essas visualizações foram feitas e publicadas com o uso de linguagens de programação e, por isso, não são simples para reproduzir, mas os códigos utilizados estão disponíveis no *GitHub* do *The Pudding*.

A Gênero e Número, que produz elementos visuais mais simples, como gráficos de barras, e raramente utiliza a interatividade, apresenta nesta matéria gráficos cuja base está em formato de imagem estática, mas com anotações e valores adicionados na página com o uso de linguagens de programação. Ainda assim, são visualizações passíveis de reprodução em ferramentas como o *Flourish* e o *Infogram*, sites gratuitos que permitem a criação de gráficos e são de fácil uso. Como os dados não foram produzidos pelo jornal, todas as visualizações indicam a fonte.

O jornalismo de dados não se restringe a visualizações, apesar do papel central que elas ocupam no imaginário sobre a área. Há algumas situações em que elas não são recomendadas, de acordo com Gray, Bounegru e Chambers (2012): quando a história é melhor contada com textos e/ou recursos multimídia; quando há poucos dados ou pouca variação entre os dados;

quando o elemento espacial não é significativo ou convincente; quando uma tabela pode ilustrar melhor do que um gráfico. Isso pode ser observado na reportagem da Gênero e Número, em que muitos dos dados são informados ao longo do texto da reportagem quando não há uma informação a ser comparada ou como anotações, a parte textual que acompanha gráficos e que permite maior contextualização da informação apresentada (Cairo, 2019).

Na **fase 4**, conforme observamos nos exemplos analisados, frequentemente são utilizadas linguagens de programação, seja para construir visualizações interativas ou possibilitar que os gráficos criados se adaptem às diferentes telas ao serem redimensionados – para que a audiência possa consumir os produtos em computadores e celulares, por exemplo. Mas, como visto na seção anterior, pesquisadores e profissionais da área recomendam priorizar um conhecimento mais abrangente e deixam claro que é possível fazer jornalismo de dados sem programar.

Esse conhecimento abrangente, quando consideramos a visualização de dados, significa entender a mensagem que diferentes tipos de gráficos transmitem, ou a noção de codificação, que Cairo (2019) define como a propriedade que varia de acordo com o que se quer representar. Os elementos da codificação não devem ser escolhidos de forma arbitrária ou apenas de acordo com o gosto pessoal do jornalista – no caso dos gráficos de barras, a codificação diz respeito à altura; em mapas, a cor ou a sombra (Cairo, 2019). Afinal, cada conjunto de dados tem necessidades específicas para ser exibido, e essas necessidades têm tanta importância quanto os dados em si (Fry, 2008).

É importante lembrar que a visualização não é apenas um elemento do produto final no jornalismo de dados, mas também faz parte da apuração ao ajudar a identificar temas e padrões, levantar perguntas para o resto da reportagem, identificar valores atípicos – os *outliers* – na amostra ou apontar falhas (Gray; Bounegru; Chambers, 2012). Como dito anteriormente, a linearidade das fases do trabalho com dados nem sempre é observada na prática, porque as etapas podem seguir outra ordem ou serem repetidas, de acordo com as necessidades da apuração.

CONCLUSÃO

Em um cenário de crise de credibilidade nas instituições, a transparência ocupa o espaço de aliada no jornalismo. Um termo que engloba muitas práticas e escopos de aplicação, a transparência é fonte de pautas e colabora com a manutenção da democracia, quando pensamos na transparência governamental ou de empresas; ferramenta de colaboração entre profissionais da área quando há o compartilhamento informações obtidas, como visto no primeiro capítulo; e, na hipótese central desta pesquisa, um recurso de aprendizado, principalmente no jornalismo de dados, que herdou traços da ética *hacker* e dos ideais da cibercultura.

A transparência jornalística ainda está mais atrelada ao imaginário dos jornalistas do que às práticas, e as tentativas de adoção da transparência como valor profissional é mais um esforço individual do que uma movimentação institucional nas redações. Mesmo assim, tem sido construída no Brasil, desde a implementação da Lei de Acesso à Informação, uma rede de iniciativas e de jornalistas interessados em propagar a transparência e democratizar a informação. Algumas dessas iniciativas, como a Fiquem Sabendo e a Transparência Brasil, chegam a praticar uma espécie de “metatransparência”, visto que são focados em divulgar dados brutos e análises advindas da transparência governamental e, ao mesmo tempo, são transparentes com suas fontes de financiamento e conexões institucionais.

Como o objetivo desta dissertação era pensar na transparência como um caminho para aprender jornalismo de dados, foi necessário especificar o termo, separando em três categorias sistematizadas na pesquisa para facilitar o recorte: a transparência institucional, a editorial-metodológica e a tecnológica. Considerando que o foco é o aprendizado da produção de reportagens de dados, a categoria de interesse é a transparência editorial-metodológica, que trata dos procedimentos de apuração e apresentação dos dados.

Devido à abrangência do termo e da amplitude de significados ao dizer, por exemplo, que um jornal é transparente, a pesquisa utilizou o Protocolo de Transparência no Jornalismo Guiado por Dados, desenvolvido por Gehrke (2021), para analisar o *The Pudding* e a Gênero e Número, dois veículos nativos do meio digital e que são voltados para o jornalismo de dados, área de interesse desta dissertação. Pensar o aprendizado de jornalismo de dados é fundamental em um cenário em que lidar com dados se torna uma prática corriqueira nas redações, e não mais a exceção, e que, cada vez mais, a quantidade massiva de dados disponíveis é uma fonte inesgotável de pautas.

Focar especificamente no aprendizado por meio de recursos de transparência é um caminho possível para lidar com as dificuldades em implementar o jornalismo de dados na graduação – ainda não há padronização nas Diretrizes Curriculares sobre o tema, falta capacitação específica para os professores das disciplinas oferecidas e falta interesse por parte dos alunos por conta da distância entre jornalismo e áreas como a matemática e a estatística, que fazem parte do mesmo universo computacional em que habita o jornalismo de dados.

Os recursos de transparência, ao serem uma ferramenta de aprendizado, aproximam estudantes e profissionais de temas e veículos que os interessem, aliviam as visões tecnofóbicas e tecnofílicas que cercam o jornalismo de dados, e têm o potencial de impactar a formação de profissionais que, idealmente, vão se sentir também estimulados a praticarem a transparência. Além disso, aprender ao observar a prática e ter a possibilidade de enxergar as fases do trabalho com dados por meio da exploração de mídias digitais se alinha com competências observadas nos estudos sobre as mudanças cognitivas a partir da popularização da *internet*, como o autodidatismo, que é muito comum entre pessoas que trabalham com o jornalismo de dados.

A fim de refletir sobre como a transparência pode ser transformada em recurso de aprendizado de jornalismo de dados, os parâmetros estabelecidos pelo Protocolo (Gehrke, 2021) foram cruzados com as categorias criadas na pesquisa e foram recortados os que melhor se alinhavam com a transparência editorial-metodológica. Depois, os objetos foram analisados individualmente e foi realizada uma análise comparativa, para estabelecer se ambos eram transparentes, ou seja, se serviam ao propósito da pesquisa.

Dentre os parâmetros do Protocolo, os de maior peso, de acordo com o tema da dissertação, eram **fontes de dados** e **metodologia**, e esses dois foram a linha de corte na hora de selecionar uma reportagem de cada um dos jornais analisados. Afinal, para aprender jornalismo de dados é necessário compreender as diferentes etapas do trabalho com dados e seria essencial observar essas etapas em reportagens publicadas.

Dois reportagens que cumpriam a exigência mínima foram selecionadas e analisadas nos parâmetros que haviam sido conectados com a transparência editorial-metodológica. Tendo sido aprovadas nessa etapa, o passo seguinte foi observar as etapas do trabalho com dados em cada uma delas. Como foi possível notar desde as análises feitas no segundo capítulo, o *The Pudding* e a *Gênero e Número* são jornais muito diferentes e, conseqüentemente, os exemplos de reportagens observadas reproduzem essa diferença. Mas as variações que aparecem em cada uma das etapas é sintomática da gama de técnicas disponíveis para a construção de reportagens com dados.

A pesquisa exemplifica, assim, como os recursos de transparência editorial-metodológica publicados em reportagens de dados podem ser sistematizados para o aprendizado, após a confirmação de que o veículo e a reportagem escolhidos são, de fato, transparentes. Longe de esgotar o tema, esta dissertação pretende colaborar com as discussões sobre a transparência como valor jornalístico e o aprendizado do jornalismo de dados no Brasil.

Uma limitação da pesquisa e, também, um possível desdobramento da dissertação, é a aplicação da sistematização dos recursos de transparência editorial-metodológica em turmas da graduação. Seria interessante comparar as expectativas dos alunos no início da disciplina e a percepção que têm do jornalismo de dados depois do contato aprofundado com os recursos de transparência, possibilitando o refinamento da sistematização dos recursos e até mesmo a inclusão de novas etapas nesse processo, se necessário.

No caso de disciplinas aprofundadas ou de cursos de maior duração, o mesmo procedimento utilizado com a transparência editorial-metodológica pode ser reproduzido para a transparência tecnológica, a fim de que os alunos aprendam a trabalhar com linguagens de programação para construir visualizações elaboradas, analisar grandes volumes de dados e extrair informações por meio da raspagem de dados, por exemplo.

Pensar no aprendizado do jornalismo de dados por meio de recursos de transparência, ao mesmo tempo em que a transparência é considerada uma forma de combater a crise de credibilidade, aponta para a formação de profissionais que se conectam com a transparência como valor e, idealmente, esse impacto poderá ser observado nas redações a longo prazo, trazendo a transparência do imaginário e dos esforços individuais para uma prática organizacional. Esse movimento beneficia o jornalismo ao aproximar as audiências e, como visto na dissertação, na formação de futuros profissionais.

REFERÊNCIAS

- ALONSO, A. A mente e os artifícios do *Homo Zapiens*. **TRANZ**, Rio de Janeiro, v. 13, p. 1-25, dez. 2018. Disponível em: http://www.tranz.org.br/13_edicao/TranZ18-Aristides-ArTecnologia.pdf. Acesso em: 19 ago. 2022.
- ARAÚJO, L. V. Apontamentos iniciais sobre a prática do jornalismo de dados à luz da ética profissional. **Âncora**, João Pessoa, v.3, n. 2, p. 179-193, jul./dez. 2016. Disponível em: <https://periodicos.ufpb.br/index.php/ancora/article/view/30843>. Acesso em: 19 nov. 2023.
- BAZZO, J. **O Jornalismo de Dados e sua incorporação no Bacharelado em Jornalismo**. 2019. 204 f. Dissertação (Mestrado em Comunicação) – Faculdade de Informação e Comunicação, Universidade Federal de Goiás, Goiânia, 2019.
- BAZZO, J.; MARTINS, D. L.; BARBOSA, F. A. C. O surgimento da pesquisa em Jornalismo de Dados no Brasil. **Intexto**, Porto Alegre, n. 50, p. 280-302, set./dez. 2020. Disponível em: <https://seer.ufrgs.br/intexto/article/view/89536>. Acesso em: 28 jun. 2023.
- BELISÁRIO, A. *et al.* **Fluxo do trabalho com dados: do zero à prática**. São Paulo: Open Knowledge Brasil, 2020.
- BERRET, C.; Phillips, C. **Teaching data and computacional journalism**. Nova York: Columbia Journalism School, 2016.
- BRADSHAW, P. The inverted pyramid of data journalism. **Online Journalism Blog**, 2011. <https://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>. Acesso em: 20 nov. 2023.
- BRENOL, M. V. **Transparência digital e jornalismo: modalidades comunicativas com uso de dados públicos**. 2019. 237 f. Tese (Doutorado em Comunicação) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2019.
- CAIRES, D. R. **Desenhando a infografia em base de dados: uma análise sobre profundidade do conteúdo e formas inteligíveis da visualização da informação**. 2019. 201 f. Dissertação (Mestrado em Comunicação e Cultura Contemporâneas) – Faculdade de Comunicação, Universidade Federal da Bahia, Salvador, 2019.
- CAIRO, A. Por que a visualização de dados falha se não pensamos sobre seu propósito. **Lumina**, Juiz de Fora, v. 13, n. 2, p. 117-125, mai./ago. 2019. Disponível em: <https://periodicos.ufjf.br/index.php/lumina/article/view/27762/19022>. Acesso em: 5 jun. 2023
- CHRISTOFOLETTI, R. (ed.). **Transparência no jornalismo: o que é e como se faz?** Florianópolis: Universidade Federal de Santa Catarina, 2022. Disponível em: <https://repositorio.ufsc.br/handle/123456789/236880>. Acesso em: 20 jun. 2023.
- CHRISTOFOLETTI, R.; BECKER, D.; PAUL, D. Transparência e ética jornalística: análise de códigos deontológicos nos maiores mercados de notícia latino-americanos. **Fronteiras:**

estudos midiáticos, São Leopoldo, v. 23, n. 3, p. 91-103, set./dez. 2021. Disponível em: <http://revistas.unisinos.br/index.php/fronteiras/article/view/23023/60748901>. Acesso em: 26 jul. 2022.

CRAIG, D.; KETTERER, S.; YOUSUF, M. To Post or Not to Post: Online Discussion of Gun Permit Mapping and the Development of Ethical Standards in Data Journalism. **Journalism & Mass Communication Quarterly**, Columbia, v. 94, n.1, p. 168-188, jan. 2017. Disponível em: <https://www.researchgate.net/publication/312187007>. Acesso em: 5 jun. 2023.

DANIELS, M. The Journalist-Engineer. **Medium**, 2015. Disponível em: https://medium.com/@matthew_daniels/the-journalist-engineer-c9c1a72b993f. Acesso em: 29 abr. 2023.

DANIELS, M. The State of The Pudding, 2018. **Medium**, 2017. Disponível em: https://medium.com/@matthew_daniels/the-state-of-the-pudding-2018-9661ab4d299c. Acesso em: 26 abr. 2023.

FITTIPALDI, M. (org.) **Ética no jornalismo 4.0**. Rio de Janeiro: iVentura, 2019.

FREITAS, F. R. R. **A visualização guiada por dados na TV o infográfico como efeito de realidade e elemento de articulação da narrativa telejornalística**. 2018. 155 f. Dissertação (Mestrado em Comunicação) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

FRY, B. **Visualizing Data**. Sebastopol: O'Reilly Media, 2008.

GEHRKE, M. **Os elementos de transparência no jornalismo guiado por dados**. 2021. 231 f. Tese (Doutorado em Comunicação) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2021.

GEHRKE, M. **O uso de fontes documentais no jornalismo guiado por dados**. 2018. 129 f. Dissertação (Mestrado em Comunicação e Informação) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

GEHRKE, M. Transparency as a key element of data journalism: perceptions of Brazilian professionals. In: **COMPUTATION + JOURNALISM SYMPOSIUM 2020**, 2020, Boston. **Anais eletrônicos** [...] Boston: Northeastern University, 2020. Disponível em: https://cpb-us-w2.wpmucdn.com/sites.northeastern.edu/dist/0/367/files/2020/02/CJ_2020_paper_8.pdf. Acesso em: 26 jul. 2022.

GONÇALVES, F. E. **A Lei de Acesso à Informação como ferramenta de apuração: uma análise do uso da LAI por jornalistas**. 2019. 177 f. Dissertação (Mestrado em Comunicação) – Faculdade de Comunicação, Universidade de Brasília, Brasília, 2019.

GRAY, J.; BOUNEGRU, L.; CHAMBERS, L. (org.) **Manual do Jornalismo de Dados**. Sebastopol: O'Reilly Media, 2012.

HUF, N. **Data comics**: Comics for data visualization in journalism. 2021. 130 f. Dissertação (Mestrado em Comunicação e Cultura Contemporâneas) – Faculdade de Comunicação, Universidade Federal da Bahia, Salvador, 2021.

KARAM, F. J. C. **Jornalismo, ética e liberdade**. São Paulo: Summus Editorial, 2014.

LAVOR, T. M. O. **Apropriação de dados e jornalismo**: um estudo de caso dos núcleos de dados no Ceará. 2021. 93 f. Dissertação (Mestrado em Comunicação) – Programa de Pós-Graduação em Comunicação, Universidade Federal do Ceará, Fortaleza, 2021.

LOPES, O. C. **Estratégias de familiarização e autorreflexão em visualizações de dados interativas**. 2022. 175 f. Dissertação (Mestrado em Jornalismo) – Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, Florianópolis, 2022.

LÉVY, P. **A inteligência coletiva**: por uma antropologia do ciberespaço. São Paulo: Edições Loyola, 2007.

LIMA, P. M. **O micromundo dos jornalistas de dados no Brasil**: carreira profissional e construção de identidade. 2021. 426 f. Tese (Doutorado em Jornalismo) – Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, Florianópolis, 2021.

MANCINI, L.; VASCONCELLOS, F. Jornalismo de Dados: conceito e categorias. **Fronteiras**: estudos midiáticos, São Leopoldo, v. 18, n. 1, p. 69-82, jan./abr. 2016. Disponível em: <https://revistas.unisinos.br/index.php/fronteiras/article/view/fem.2016.181.07/5300>. Acesso em: 11 jun. 2023.

MASTRELLA, B. **Os discursos de legitimação do Jornalismo Guiado por Dados**. 2019. 234 f. Dissertação (Mestrado em Comunicação) – Faculdade de Comunicação, Universidade de Brasília, Brasília, 2019.

MAYER-SCHÖNBERGER, V.; CUKIER, K. **Big Data**: Como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana. Rio de Janeiro: Elsevier, 2013.

MIELNICZUK, L.; TRÄSEL, M. Jornalismo guiado por dados como inovação profissional e seus desafios para a educação. **Contemporanea**, Salvador, v. 15, n. 2, p. 609-629, maio/ago. 2017. Disponível em: <https://periodicos.ufba.br/index.php/contemporaneaposcom/article/view/21510/15119>. Acesso em: 19 ago. 2022.

MORETZSOHN, S. “Profissionalismo” e “objetividade”: o jornalismo na contramão da política. **Biblioteca On-line de Ciências da Comunicação**, 2001. Disponível em: <http://www.bocc.ubi.pt/pag/moretzsohn-sylvia-profissionalismo-jornalismo.pdf>. Acesso em: 19 set. 2022.

NPR. **The Origin Of ‘Proof Is In The Pudding’**, 2012. Disponível em: <https://www.npr.org/2012/08/24/159975466/corrections-and-comments-to-stories>. Acesso em: 20 jun. 2023.

POLYGRAPH. **Polygraph**, [s.d.]. Disponível em: <https://polygraph.cool>. Acesso em: 26 dez. 2022.

RÉGIS, F. **Cognição e afeto na comunicação**: conectando corpo, mente, meio e tecnologia. Porto Alegre: Sulina, 2022.

RÉGIS, F. **Nós, ciborgues**: tecnologias de informação e subjetividade homem-máquina. Curitiba: PUCPRESS, 2012.

REUTERS Institute for the Study of Journalism. **Digital News Report 2023**. Disponível em: <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023>. Acesso em: 20 jun. 2023.

RIBEIRO, B. P. **As compreensões dos jornalistas de dados sobre transparência nos processos produtivos**. 2022. 131 f. Dissertação (Mestrado em Comunicação) – Centro de Ciências Sociais e Humanas, Universidade Federal de Santa Maria, Santa Maria, 2022.

RODRIGUES, C. M. **Um estudo de caso sobre jornalismo de dados abertos**: método, evidência e transparência nas rotinas produtivas no Brasil e na Inglaterra. 2021. 603 f. Tese (Doutorado em Comunicação) – Departamento de Comunicação Social, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2021.

RODRIGUES, K. C. **A construção do acontecimento no jornalismo guiado por dados**: estudo de veículos de comunicação latino-americanos. 2021. 257 f. Tese (Doutorado em Comunicação) – Faculdade de Arquitetura, Artes, Comunicação e Design, Universidade Estadual Paulista, Bauru, 2021.

SALAVERRÍA, R. O ciberjornalismo encontra a universidade: ideias para melhorar o ensino e a pesquisa. **Brazilian journalism research**, Brasília, v. 7, n. 2, p. 141–157, 2011. Disponível em: <https://bjr.sbpjor.org.br/bjr/article/view/343>. Acesso em: 17 ago. 2022.

SANTOS, M. F. L. Tem #DDJBR aqui? Mapeando a presença do jornalismo de dados no Brasil. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 42., 2019, Belém. **Anais eletrônicos** [...] Belém: Intercom, 2019. Disponível em: <https://www.portalintercom.org.br/anais/nacional2019/resumos/R14-1396-1.pdf>. Acesso em: 10 jul. 2023.

SILVA, E. C. S. **Universo de dados e o desafio aos jornalistas em meio à pandemia de Covid-19**: contribuições do consórcio de veículos de imprensa no Brasil. 146 f. Dissertação (Mestrado em Comunicação) – Centro de Ciências da Educação, Universidade Federal do Piauí, Teresina, 2022.

SCHUDSON, M. **Discovering the news**: A social history of American newspapers. Nova York: Basic Books, 1978.

THE PUDDING. **Pitch a story** [s.d.]. Disponível em: <https://pudding.cool/pitch/>. Acesso em: 5 jun. 2023.

TRÄSEL, M. **Entrevistando planilhas**: Estudo das crenças e do ethos de um grupo de profissionais de jornalismo guiado por dados no Brasil. 2014. 314 f. Tese (Doutorado em

Comunicação Social) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2014a.

TRÄSEL, M. Hacks and hackers: the ethos and beliefs of a group of Data-Driven Journalism professionals in Brazil. **FAMECOS**, Porto Alegre, v. 25, n. 1, p. 1-14, jan./abr. 2018.

Disponível em:

<https://revistaseletronicas.pucrs.br/ojs/index.php/revistafamecos/article/view/27589>. Acesso em: 9 jul. 2023.

TRÄSEL, M. Jornalismo guiado por dados: aproximações entre a identidade jornalística e a cultura hacker. **Estudos em Jornalismo e Mídia**, Santa Catarina, v. 11, n. 1, p. 291-304, jan./jun. 2014b. Disponível em:

Disponível em:

<https://periodicos.ufsc.br/index.php/jornalismo/article/view/1984-6924.2014v11n1p291>.

Acesso em: 20 jun. 2023.

TRÄSEL, M. Panorama do ensino de jornalismo guiado por dados no Brasil. *In:*

ENCONTRO NACIONAL DE PESQUISADORES EM JORNALISMO, 17., 2019, Goiânia.

Anais eletrônicos [...] Goiânia: SBPJor, 2019. Disponível em:

<https://sbpjour.org.br/congresso/index.php/sbpjour/sbpjour2019/paper/viewFile/2180/1238>.

Acesso em: 19 ago. 2022.

VASCONCELLOS, F.; MANCINI, L. As assimetrias regionais da prática do Jornalismo de Dados no Brasil. *In:* SEMINÁRIO DE PESQUISA EM JORNALISMO INVESTIGATIVO, 3., 2016, São Paulo. **Anais...** São Paulo: ABRAJI, 2016.

VENTURA, M. P. **O ensino de Jornalismo de Dados: desafios e possibilidades.** 2021. 249 f. Tese (Doutorado em Jornalismo) – Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, Florianópolis, 2021.

VENTURA, M. P.; PAULINO, R. D. C. R. Jornalismo de Dados na América Latina: o perfil do ensino. *In:* MEDITSCH, E.; KRONBAUER, J.; BEZERRA, J. F. (org.) **Pedagogia do Jornalismo: desafios, experiências e inovações.** Florianópolis: Editora Insular, 2020. p. 247-261.

ZAMITH, R. Transparency, Interactivity, Diversity, and Information Provenance in Everyday Data Journalism. **Digital Journalism**, [s.l.], v. 7, n. 4, p. 1-20, 2019. Disponível em:

<https://www.tandfonline.com/doi/full/10.1080/21670811.2018.1554409>. Acesso em: 6 jun. 2023.