



Podem até cair do céu, mas só se tornam dados aqui na terra: datificação na sociedade e na ciência

They may even fall from sky, but they only become data here on earth: datafication in society and science

Elizabeth Maria Freire de JESUS

Universidade Federal do Rio de Janeiro

Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais

beth@nce.ufrj.br

Abstract. *Associated with the intensification of digitalization and big data, the term “datafication” emerges amidst a diversity of perspectives, interests, expectations, understandings, controversies, opportunities and risks placed around the processes of translating subjects, things and practices into data. Based on bibliographical research, the present work aims to contribute with some reflections on the epistemological implications, taking into account the context of intensive datafication in society and science. In this direction, we will present the main concepts and characteristics relevant to the research context.*

Keywords: *Datafication. Big Data. Data-Driven Science. Information Regime.*

Resumo. Associada à intensificação da digitalização e ao big data, o termo “datificação” emerge em meio a uma diversidade de perspectivas, interesses, expectativas, entendimentos, controvérsias, oportunidades e riscos colocados em torno dos processos de tradução dos sujeitos, das coisas e práticas em dados. A partir de um levantamento bibliográfico, o presente trabalho tem como objetivo contribuir com algumas reflexões sobre as implicações epistemológicas, tendo-se em conta o contexto de intensiva datificação na sociedade e na ciência. Nessa direção, apresentaremos os principais conceitos e características pertinentes ao contexto de investigação.

Palavras-chave: Datificação. Big Data. Ciência Orientada por Dados. Regime de Informação.



Recebido: 09/09/2024 Aceito: 05/10/2024 Publicado: 10/12/2024

DOI:10.51919/revista_sh.v1i0.446

1. Introdução

O protagonismo dos dados não é uma nova característica das sociedades contemporâneas, como também a datificação não é necessariamente resultante de poderosos processadores e algoritmos (Cieslik; Margócsy, 2022). O que há de novo são a escala surpreendente no volume, na velocidade e variedade em que dados são produzidos e consumidos para diversos e diferentes propósitos e o poder dos algoritmos, dos métodos e modelos estatísticos e preditivos de inteligência artificial (IA) que (re)configuram ou moldam as relações sociais e tecnopolíticas nos mais variados domínios da sociedade e na ciência.

A agregação, a análise e correlações estatísticas de dados na/da sociedade e na/da ciência não é algo novo, no entanto, a datificação, com a tradução do mundo natural e social em dados, avança implacável. Os centros de cálculo, “locais onde cálculos literais, e não simplesmente metafóricos, são exequíveis graças ao formato matemático” (Latour, 2012, p. 256), a necessidade de governar as populações a partir da identificação de regularidades (tamanho, número de mortes, nascimentos etc.) (Foucault, 2015) e as tecnologias para assegurar o exercício da vigilância, do controle e do funcionamento do poder sobre multiplicidades populacionais em espaços específicos, como hospitais, conventos, escolas e fábricas (Foucault, 2014), se atualizam no que podemos considerar como o novo regime de informação vigente, caracterizado pela massiva extração/captura de dados e metadados que são processados em grandes infraestruturas tecnológicas por complexos algoritmos e inteligência artificial, majoritariamente realizados por grandes e poderosas empresas de tecnologia (*big techs*), sobredeterminando, em grande medida, novas formas de controle e vigilância digital (*dataveillance*), agora, mais sutis e invisíveis, possibilitadas a partir dos rastros ou pegadas digitais que os indivíduos vão deixando ao longo de suas ações informacionais no mundo digital, cada vez mais imbricado e indissociável do mundo físico.

A cada ano cresce exponencialmente o número de dados gerados em nível global. De acordo com o relatório da CISCO¹, mais de 70% da população mundial estaria conectada (conexão móvel) até 2023, tendo-se em conta, é claro, as assimetrias de ordens diversas existentes entre diferentes países e sociedades. O último infográfico “Data Never Sleeps”², produzido pela empresa Domo referente ao ano de 2022, mostra que a cada minuto do dia no mundo inteiro, 5,9 milhões de pesquisas são feitas no Google, 66 mil fotos são postadas no Instagram, mais de 347 mil tuítes são

¹ CISCO. Cisco Annual Internet Report, 2018–2023. Disponível em: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>. Acesso em: 20 out. 2023.

² Domo. Data Never Sleeps. Disponível em: <https://www.domo.com/data-never-sleeps#>. Acesso em 20 out. 2023.

publicados no Twitter, 231,4 milhões de e-mails são enviados, 104 mil horas de reuniões no Zoom, 500 horas de vídeo sobem no YouTube, 1,7 milhão de conteúdos são compartilhados no Facebook etc.

Pautando-se no discurso do “autoconhecimento por meio dos números”, de que aquilo que não pode ser quantificado, não pode ser melhorado, em 2007, os editores da revista Wired, Gary Wolf e Kevin Kelly, deslançaram o movimento “*Self-Quantified Movement*” com o objetivo de arregimentar interessados em desenvolver e utilizar métodos e tecnologias para a metrificação do cotidiano das pessoas, de suas atividades físicas.

A metáfora “os dados são o novo petróleo” do século XXI frequentemente reproduzida em diversas publicações e espaços, principalmente, mas não exclusivamente, vinculantes ao domínio empresarial e governamental, remete à ideia de que os dados, tais como o petróleo são valiosos recursos que se encontram na natureza, sendo a sua extração e processamento contínuo em grande escala legítimos em prol do desenvolvimento da economia e da sociedade contemporâneas (Couldry; Yu, 2018). A fragilidade dessa pretensa naturalização se constata logo de início, uma vez que dados são artefatos sociotécnicos, cuja produção e captura dependem da conversão de fatos ou eventos da realidade em elementos passíveis de serem classificados, agrupados, analisados e usados para fins diversos, nem sempre legítimos ou moralmente justificáveis.

Vivemos, já a algum tempo, em um mundo digitalmente globalizado onde praticamente todas as ações e práticas informacionais, cada vez mais se realizam mediadas ou atravessadas por dispositivos tecnológicos digitais. Somos *prosumers* – produtores e consumidores de dados e informações, temos que lidar com o excesso de dados e informações de boa e má qualidade, com fontes credíveis e não, com inúmeras estratégias de desinformação. A sociedade datificada contemporânea, experencia atualizadas formas de sociabilidade, de poder, de saber, sobre si e sobre o outro, de produzir e comunicar conhecimento, científico ou não, promovidas e sustentadas por novas racionalidades e discursos criados (forjados) e autorreferenciados no âmbito do regime de informação vigente.

A explosão na produção e uso de *big data* têm impactos epistemológicos (Borgman, 2011; Sayão; Sales, 2019). A partir de um levantamento bibliográfico, o presente trabalho tem como objetivo apontar algumas implicações epistemológicas, salientando a importância do desenvolvimento de competências instrumentais e, sobretudo, críticas, tendo-se em conta o contexto de intensiva datificação na sociedade e na ciência. Nessa direção, serão apresentados conceitos e características pertinentes ao contexto de investigação.

2. Dados, *big data*, datificação e o novo regime de informação

Associada à intensificação da digitalização e ao *big data*, o termo “datificação” emerge em meio a uma diversidade de perspectivas, interesses, expectativas, entendimentos, controvérsias, oportunidades e riscos colocados em torno dos processos de tradução dos sujeitos, das coisas e práticas em dados. Embora sejam distintos, os termos digitalização, *big data* e datificação estão fortemente inter-relacionados e orbitam em torno de um elemento comum: dados (dados digitais).

A palavra dados, derivada da palavra latina *data* que é o plural de *datum*, tem inúmeros e diferentes entendimentos que podem variar significativamente.

Dados podem ser entendidos como sendo “matéria-prima produzida pela abstração do mundo em categorias, medidas e outras formas de representação – números, caracteres, símbolos, imagens, sons, ondas eletromagnéticas, bits – que constituem os blocos de construção a partir dos quais a informação e o conhecimento são criados” (Kitchin, 2014, p.2).

De uma forma ampla, dados podem ser considerados como

uma representação reinterpretável de informações de maneira formalizada, adequada para comunicação, interpretação ou processamento. Exemplos de dados incluem uma sequência de bits, uma tabela de números, os caracteres de uma página, a gravação de sons feitos por uma pessoa falando ou um espécime de rocha lunar. (CCSDS, 2012,10)

Outra definição bastante citada é a que considera dado (digital) como “qualquer informação que possa ser armazenada em formato digital, incluindo texto, números, imagens, vídeos ou filmes, áudio, software, algoritmos, equações, animações, modelos, simulações, etc. Esses dados podem ser gerados por vários meios, incluindo observação, computação ou experimento” (NATIONAL SCIENCE BOARD, 2005, p.9).

Cabe ressaltar a natureza relacional e situacional dos dados, sendo estes sujeitos a muitas interpretações. Nessa direção, Borgman (2011, p.5) pontua que “os dados podem existir apenas aos olhos de quem vê: o reconhecimento de que uma observação, artefato ou registro constitui dados é em si um ato acadêmico”.

Para Nielsen e Hjørland,

Dados são sempre registrados com base em alguns interesses, perspectivas, tecnologias e práticas situadas que determinam o seu significado e utilidade em diferentes contextos. (Nielsen; Hjørland, 2014, p.225)

Com essas definições de “dados”, é possível perceber que são muitas e variadas as formas que os dados podem assumir dentro e fora do domínio da ciência, que as fontes de dados também podem variar amplamente e que os significados e o valor atribuído aos dados guardam forte relação com os contextos de produção, de uso e reuso dos dados, interesses e práticas.

Ao longo da segunda metade do século XX, mais e mais fenômenos no mundo tornaram-se passíveis de serem capturados ou traduzidos como dados, devido a radical expansão e integração de computação, redes, dispositivos digitais e armazenamento e processamento de dados que, em

seu conjunto, forneceu uma plataforma robusta para a explosão do *big data*, além de ser o meio pelo qual o *big data* é gerado, processado, compartilhado e analisado.

As populares e multidimensionais definições de *big data* são baseadas em suas características, sinteticamente apresentadas na literatura em torno de vários V's. Aqueles que melhor justificam o prefixo “big” associado à palavra em inglês “data” são o Volume, a Velocidade de criação/acumulação e de processamento de dados em tempo real ou quase, e a Variedade, referindo-se à heterogeneidade de tipos de dados, estruturados e não-estruturados.

Boyd e Crawford (2012, p.663), definem *big data* como “um fenômeno cultural, tecnológico e acadêmico”, consubstanciado pelo poder de computação e a precisão algorítmica para reunir, analisar, vincular e comparar grandes conjuntos de dados; pela análise e identificação de padrões utilizados como base para reivindicações discursivas e práticas diversas.

À medida que uma miríade de dispositivos tecnológicos digitais pervasivos e ubíquos se multiplicam entre tarefas e locais e interagem entre si, mais e mais dados digitais são produzidos/coletados/extraídos. Cada vez mais o mundo é convertido em dados.

A datificação é muito mais que digitalização – processo de conversão de algo em dígitos binários por meio de dispositivos apropriados, como um escâner, tornando-o acessível por meio de sistema computacional e/ou dispositivo eletrônico. Segundo Van Dijck (2014, p.198), a datificação envolve a tradução, em princípio, de quaisquer ações ou fenômenos em dados digitais, sendo considerado como “um novo paradigma na ciência e na sociedade”; como um novo modo de conhecer e interpretar o mundo.

A datificação e a mineração de dados estão, portanto, na base de uma racionalidade que busca um novo paradigma social-científico, alicerçado em uma crença gradualmente generalizada na objetividade da quantificação e dos métodos quantificáveis como sendo a principal chave para entender a sociabilidade e o comportamento social. “A premissa de que grandes conjuntos de dados oferecem uma forma superior de inteligência e conhecimento que podem gerar insights que antes eram impossíveis, com a aura de verdade, objetividade e precisão” (Boyd; Crawford, 2012, p.2), tem sido compartilhada por diversos e diferentes atores, desempenhando “um papel profundamente ideológico na interseção entre sociabilidade, pesquisa e comércio” (Van Dijck, 2017, p.45), e, portanto, problemático sob diversos aspectos.

Mejias e Couldry oferecem interessante consideração que vão ao encontro do objetivo deste artigo, que busca articular datificação e regime de informação,

A produção de dados não pode ser separada de dois elementos essenciais: a infraestrutura externa através da qual são recolhidos, processados e armazenados, e os processos de geração de valor, que incluem a monetização, mas também meios de controle estatal, produção cultural, capacitação cívica, etc. Esta infraestrutura e esses processos são multicamadas e globais, incluindo mecanismos de disseminação, acesso, armazenamento, análise e vigilância que pertencem ou são controlados principalmente por empresas e estados. (Mejias; Couldry, 2019, p. 4)

2.1. Regime de Informação

Pensando acerca dos fluxos de informação que nos cercam, sejam culturais, acadêmico-científicos, financeiros, comerciais etc., e inspirado na Teoria Ator-Rede, Frohmann apresenta o conceito de regime de informação como sendo, “qualquer sistema ou rede mais ou menos estável em que a informação flui através de canais determinados de produtores específicos, via estruturas organizacionais específicas, para consumidores ou usuários específicos” (Frohmann, 1995, s/p).

Como um plexo de relações e agências, “um regime de informação está exposto a certas possibilidades e condições culturais, políticas e econômicas, que nele se expressam e nele se constituem (González de Gomez, 2012, p. 43). O regime de informação,

Remete à uma equação entre potencialidades e forças de diferentes origens e direções que, combinadas, definem os modos de distribuição - entre diferentes atores sociais, atividades e regiões - da potência construtiva e do poder seletivo que se manifestam na geração, circulação, acesso e uso de informação. (González de Gómez; Chicanel, 2008, p.5)

Como uma ferramenta conceitual-metodológica, o conceito de regime de informação permite situar, diagnosticar e analisar as relações de diversos e heterogêneos atores, práticas, discursos e recursos que nele, no regime de informação, se expressam e nele se constituem, e, sobretudo, analisar as relação informação-poder.

A datificação envolve, portanto, um plexo de atores heterogêneos interessados na coleta sistemática, processamento algorítmico, compartilhamento, comercialização/venda de dados e metadados capturados, em grande medida, de forma sub-reptícia, sem a ciência ou consentimento do usuário/cliente de redes sociais digitais, plataformas de internet entre outras tecnologias, sendo muito deles considerados por Zuboff (2020) como “superávit comportamental” – excedentes que são transformados em produtos de predição comercializados num novo tipo de mercadoria que Zuboff chama de mercado de comportamentos futuros.

A aceitação acrítica da datificação pautada em premissas ideológicas calcada na neutralidade e racionalidade técnica da performatividade algorítmica, requer e tem recebido a atenção, em especial, quando se trata das implicações epistemológicas no campo das ciências, cada vez mais influenciadas pelos avanços do *big data* e pela utilização de infraestruturas computacionais e de sofisticados métodos ou ferramentas de análise e visualização de grandes volumes de dados em suas cadeias metodológicas.

3. Algumas implicações epistemológicas

Vivemos em mundo “rico” em dados – ainda que estes, em sua grande maioria, não sejam ricos em termos de metadados (dados sobre dados) – onde são grandes as expectativas e apostas no

acesso, no compartilhamento, uso e reuso de dados; princípios que estão na base do movimento Ciência Aberta.

O crescimento da produção e disponibilidade de *big data* e de infraestruturas de dados, juntamente com ferramentas analíticas e de visualizações têm transformado os modos de produzir conhecimento não nos contextos da chamada “*big science*”, como também têm alcançando cada vez mais a cauda longa da ciência – aquela formada coletivamente por inúmeros projetos de pesquisa desenvolvidos em pequenos laboratórios.

A ciência orientada por dados, conhecida como o quarto paradigma da ciência, tem sido considerada como a mais adequada para explorar, extrair valor e dar sentido a conjuntos massivos e interconectados de dados, fomentar a pesquisa interdisciplinar e levar a modelos e teorias mais holísticos e extensos em sistemas complexos (ROYAL SOCIETY, 2012). O conhecimento datificado nas ciências naturais, da vida, físicas, engenharias e nas ciências humanas e sociais, marca uma virada epistemológica. Esse cenário contemporâneo, sem dúvida, desafia epistemologias estabelecidas em diferentes áreas e disciplinas científicas, como por exemplo, no que diz respeito aos processos de avaliação, validação e replicação da pesquisa com implicações para a confiabilidade e integridade da pesquisa científica.

Uma situação recente envolvendo mudanças no processo de avaliação considerando a datificação na ciência, foi o da revista científica *The Lancet* que anunciou mudanças no processo de revisão de artigos para evitar a divulgação de *papers* baseados na interpretação de grandes conjuntos de dados e que não possam ser auditados. A revista passou a exigir que mais de um autor de um artigo científico tenha acessado e checado os dados que embasam o estudo, como também que os nomes dos autores que fizeram a verificação deverão ser informados aos editores, quais dados serão compartilhados e os critérios para acessá-los, incluindo também o convite de especialistas em estatística e ciência de dados para opinar, principalmente em relação à integridade e ao uso ético das informações. Tais mudanças tiveram como pano de fundo a publicação e posterior retratação de um artigo que teve grande repercussão no meio científico e na sociedade, relacionando a hidroxiquina e a infecção do novo coronavírus, durante a pandemia de Covid-19.³

Outro aspecto crucial apontado por Lazer e colaboradores (2014) é o fato de que a maior parte do *big data* [...] não é resultado de instrumentos projetados para produzir dados válidos e confiáveis passíveis de análise científica. Vale lembrar, que a datificação tem como um dos maiores propulsores da captura e fornecimento de dados, em especial, considerados dados de pesquisa nas ciências humanas e sociais, as plataformas digitais e as redes sociais digitais com seus agenciamentos algorítmicos opacos. Reforça-se aqui, a premência do desenvolvimento de habilidades, competências instrumentais e, sobretudo, críticas para lidar e atuar de forma

³ Disponível em: <https://revistapesquisa.fapesp.br/baseado-em-informacoes-nao-confiaveis-2/>

responsável e ética nos processos de conhecer e produzir conhecimento, em especial, no domínio da ciência.

Considerando que qualquer dado ou conjunto de dados pode ser transformado e reutilizado diversas vezes desde a sua origem até seu uso em uma publicação científica e além desta, a questão da proveniência dos dados ou a cadeia de custódia dos dados se torna imensamente difícil com especial implicações para a confiabilidade da pesquisa. A gestão de dados de pesquisa e a importância do “enriquecimento” dos dados com metadados são fundamentais para a confiabilidade dos dados e da pesquisa com vistas a reutilização em novos contextos de uso e (re)interpretação.

Outra preocupação diz respeito à aplicação de métodos analíticos computacionais a grandes volumes de dados sem uma compreensão completa das suas características (por exemplo, a natureza da amostra representada pelos dados). A ausência de uma compreensão completa dos dados ou até mesmo de uma deficiente competência para lidar com dados pode levar à aplicação inadvertidamente ou de forma descuidada de ferramentas metodológicas e analíticas, apontando a necessidade do desenvolvimento de competências específicas para lidar com grandes volumes dados.

4. Conclusão

A datificação como um novo paradigma na produção de conhecimento, científico ou não, tem se tornado um princípio central e norteador não apenas entre os adeptos das tecnologias e da computação, empresas e governos, mas também entre pesquisadores e acadêmicos que a veem como uma revolucionária oportunidade de pesquisa para a compreensão e explicação do mundo social e natural. Essa virada epistemológica requer uma perspectiva crítica atenta frente à complexa rede de atores heterogêneos que envolve não só, mas principalmente grandes corporações oligopólicas, como as plataformas digitais com suas poderosas infraestruturas computacionais e processamento algorítmico altamente performativos que têm a coleta massiva de dados como parte central de seus modelos de negócio. Dados, esses, que podem vir a ser dados de pesquisa.

O desenvolvimento de competências instrumental e, sobretudo, crítica dos pesquisadores de forma a problematizar os contextos de produção dos dados, incluindo a credibilidade da fonte de dados e informações, questões éticas nos processos de coleta/extração envolvendo, por exemplo, sigilo e proteção dos dados, a proveniência dos dados entre outros, estão dentre alguns dos muitos desafios para a comunidade científica em um contexto contemporâneo de intensa datificação na ciência e na sociedade. Desafios que precisam ser enfrentados institucionalmente com um guarda-chuva de políticas e práticas, como por exemplo, políticas e iniciativas envolvendo a gestão de dados de pesquisa, incluindo a educação e o treinamento de pesquisadores nesta temática, a necessidade de elaboração do plano de gestão de dados de pesquisa (PGD) etc. Este último, o PGD, tem cada vez mais se tornado uma exigência por parte

das agências de financiamento de pesquisa internacionais e nacionais, considerado como uma das boas práticas de pesquisa visando a ética na ciência e integridade na pesquisa.

Referências

Borgman, Christine L., Research Data: Who Will Share What, with Whom, When, and Why? **RatSWD Working Paper**, n. 161, 2010. Disponível em: <https://ssrn.com/abstract=1714427> or <http://dx.doi.org/10.2139/ssrn.1714427>.

Borgman, Christine L. The conundrum of sharing research data. **Journal of the American Society for Information Science**, v. 63, n.6, 2011.

Boyd, D.; Crawford, K. Critical questions for Big Data: provocations for a cultural, technological, and scholarly phenomenon. **Information, Communication & Society, Abingdon**, v. 15, n.5, 2012. Disponível em: <https://www.bellisario.psu.edu/assets/uploads/criticalquestions.pdf>. Acesso em: 20 out. 2023.

Boyd, Danah; Crawford, Kate, Six provocations for big data. In: **Symposium on the Dynamics of the Internet and Society**, September 2011. Disponível em: <https://ssrn.com/abstract=1926431>. Acesso em: 20 out. 2023.

CIESLIK, Katarzyna; MARGÓCSY, Dániel. Datafication, Power and Control in Development: A Historical Perspective on the Perils and Longevity of Data. **Progress in Development Studies**, v. 22, n.4, p. 352–373, 2022. Disponível em: <https://journals.sagepub.com/doi/10.1177/14649934221076580>. Acesso em: 27 out. 2023.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS). **Reference model for an open archival information system (OAIS)**. Washington, DC: National Aeronautics and Space Administration, p. 1-9, 2012. Disponível em: <http://public.ccsds.org/publications/archive/650x0b1.pdf>. Acesso em: 20 out. 2023.

Couldry, Nick; Yu, Jun. Deconstructing datafication's brave new world. **New Media & Society**, v.20, n.12, 2018. Disponível em: https://www.academia.edu/37469119/Deconstructing_datafications_brave_new_world. Acesso em: 27 out. 2023.

Bruno, Fernanda. Máquinas de ver, modos de ser: visibilidade e subjetividade nas novas tecnologias de informação e de comunicação. **Revista FAMECOS**, Porto Alegre, n. 24, 2004. Disponível em: https://www.academia.edu/86405175/M%C3%A1quinas_de_ver_modos_de_ser_visibilidade_e_subjetividade_nas_novas_tecnologias_de_informa%C3%A7%C3%A3o_e_de_comunica%C3%A7%C3%A3o. Acesso em: 25 out. 2023.

Foucault, Michel. **Vigiar e punir: nascimento da prisão**. 42 ed. Petrópolis: Vozes, 2014.

Foucault, Michel. **Microfísica do Poder**. 2. ed. Rio de Janeiro: Paz e Terra, 2015.

Frohmann, Bernd. Taking information policy beyond information science: applying the actor network theory. In: **Annual Conference Proceedings**, v. 23, 1995, Canada: Canadian Association for Information, 1995.

González De Gómez, Maria Nélide. **Regime de informação: construção de um conceito**. **Informação & Sociedade: Estudos**, João Pessoa, v.22, n.3, p.43-60, set./dez. 2012.

Gómez, M. N. G.; Chicanel, M. A mudança de regimes de informação e as variações tecnológicas. **Publicação do Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação**, v. 9, 2008. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/180486>. Acesso em: 30 out. 2023.

Kitchin, Rob. Big Data, new epistemologies and paradigm shifts. **Big Data & Society**, v. 1, n. 12, 2014. Disponível em: <https://journals.sagepub.com/doi/full/10.1177/2053951714528481>. Acesso em: 20 out. 2023.

Latour, Bruno. **Reagregando o social**. 2. ed. Salvador: Edufba, 2012.

Lemos, André. Mídias locativas e vigilância. Sujeito inseguro, bolhas digitais, paredes virtuais e territórios informacionais. In: BRUNO, Fernanda; KANASHIRO, Marta; FIRMINO, Rodrigo (orgs.). **Vigilância e Visibilidade: espaço, tecnologia e identificação**. Porto Alegre: Sulina, 2010.

Mejias, Ulises A.; Couldry, Nick. Datafication. **Internet Policy Review**, v.8, n. 4, 2019.

NATIONAL SCIENCE BOARD (NSB). **Long-lived data collections: enabling research and education in the 21st century**. Arlington, VA: National Science Foundation, 2005. Disponível em: <https://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>. Acesso em: 20 out. 2023.

Nielsen, Hans Jørn; Hjørland, Birger. Curating research data: the potential roles of libraries and information professionals. **Journal of Documentation**, v. 70, n. 2, p. 221-240, 2014. Disponível em:

https://www.researchgate.net/publication/263682364_Curating_research_data_The_potential_roles_of_libraries_and_information_professionals. Acesso em: 24 out. 2023.

Sales, Luana Farias; Sayão, Luis Fernando. Uma proposta de taxonomia para dados de pesquisa. **Revista Conhecimento em Ação**, v. 4, n. 1, jan/jun. 2019c. Disponível em: <https://revistas.ufrj.br/index.php/rca/article/view/26337>. Acesso em: 18 out. 2023.

THE ROYAL SOCIETY. Science as an open enterprise. London: The Royal Society Science Policy Centre, 2012. Disponível em: <https://royalsociety.org/~media/policy/projects/sape/2012-06-20-saoe.pdf>. Acesso em: 20 out. 2023.

Van Dijck, José. Datafication, dataism and dataveillance: big data between Scientific paradigm and ideology. **Surveillance & Society**, v.12, n.2, 197–208.

Van Dijck, José. Confiamos nos dados? As implicações da datificação para o monitoramento social. **Matrizes**, v. 11, n. 1, jan./fev. 2017, p. 39-59, 2017.

Zuboff, Shoshana. A era do capitalismo de vigilância: a luta por um futuro humano na nova fronteira do poder. 1 ed. Rio de Janeiro: Intrínseca, 2020.