



O EFEITO MOSAICO NA IDENTIFICAÇÃO DE USUÁRIOS: IMPLICAÇÕES PARA A PRIVACIDADE

The mosaic effect on user identification: Implications for privacy

Amanda Garcia Gomes

Universidade Estadual Paulista (UNESP), Programa de Pós-graduação em Ciência da Informação, Marília, SP, Brasil
garcia.gomes@unesp.br
<https://orcid.org/0000-0003-3110-4040>

Ricardo César Gonçalves Sant'Ana

Universidade Estadual Paulista (UNESP), Programa de Pós-graduação em Ciência da Informação, Marília, SP, Brasil
ricardo.santana@unesp.br
<https://orcid.org/0000-0003-1387-4519>

Fernando de Assis Rodrigues

Universidade Federal do Pará (UFPA), Instituto de Ciências Sociais Aplicadas, Belém, PA, Brasil
deassis@ufpa.br
<https://orcid.org/0000-0001-9634-1202>

A lista completa com informações dos autores está no final do artigo

RESUMO

Objetivo: Caracterizar o efeito mosaico na coleta de dados pessoais em serviços *online*, sob a perspectiva da Ciência da Informação, apresentando implicações para a privacidade dos usuários.

Método: Pesquisa exploratória de abordagem qualitativa, natureza descritiva. Utilizou-se o método Análise de Conteúdo e a técnica de análise categorial na qual analisou-se as mensagens contidas nas Políticas de Privacidade de dois serviços *online* mais utilizados no ano de 2023 no Brasil, TikTok e Spotify.

Resultados: Mostraram que existem três fontes de coleta de dados pessoais: fornecidos pelo usuário ao detentor, coletados automaticamente pelo detentor e dados pessoais recebidos pelo detentor via Agentes externos. Os dados dessas fontes quando reorganizados podem revelar informações pessoais não previstas inicialmente na fase de coleta, caracterizando o efeito mosaico. Entre as implicações para a privacidade, o efeito mosaico possibilita que a reconstituição desses dados seja combinada indicando, por exemplo, a frequência e a permanência do usuário em um determinado local, caracterizando uma ação de vigilância.

Conclusões: A necessidade de investigação do efeito mosaico é uma das consequências ocasionadas pelo aumento da complexidade e do volume de dados disponíveis nos serviços *online* que recombina dados para gerar novas interpretações com potencial de identificação de usuários e, por consequência, implicações para a sua privacidade.

PALAVRAS-CHAVE: Efeito mosaico. Privacidade. Dados pessoais. Serviços *Online*.

ABSTRACT

Objective: To characterize the mosaic effect in the collection of personal data in online services, from the perspective of Information Science, presenting implications for user privacy.

Method: Exploratory research with a qualitative approach, descriptive in nature. The Content Analysis method and the categorical analysis technique were used, analyzing the messages contained in the Privacy Policies of two of the most used online services in Brazil in 2023, TikTok and Spotify.

Results: The study revealed three sources of personal data collection: data provided by the user to the data holder, data automatically collected by the data holder, and personal data received by the data holder from external agents. When reorganized, data from these sources can reveal personal information not initially foreseen during the collection phase, characterizing the mosaic effect. Among the privacy implications, the mosaic effect enables the recombination of these data, potentially indicating, for example, the frequency and duration of the user's presence at a specific location, thereby constituting a surveillance action.

Conclusions: The need to investigate the mosaic effect is one of the consequences arising from the increased complexity and volume of data available in online services, which recombine data to generate new interpretations with the potential to identify users, and consequently, implications for their privacy.

KEYWORDS: Mosaic Effect. Privacy. Personal Data. Online Services.

1 INTRODUÇÃO

A Ciência da Informação é caracterizada por ser interdisciplinar e seu espectro de atuação pode atender a demandas relativas à pesquisa de princípios e de métodos relacionados à informação. Esta ciência tem abordado os problemas informacionais que modificam a relevância da informação em diferentes contextos socioculturais e, entre as possibilidades investigativas, estão as questões relacionadas com o uso de tecnologias de acesso a dados (González de Gómez, 2000).

O imperativo tecnológico – o uso das supracitadas tecnologias – vem apresentando tanto o desenvolvimento e a aplicação de uma crescente gama de produtos e serviços de informação em atividades humanas, como também a transformação para uma sociedade da informação, aumentando a relevância dos estudos desta ciência em buscar soluções para os problemas informacionais (Saracevic, 1996).

Neste processo, a informação se torna um recurso central e as conexões globais são impulsionadas pela revolução das Tecnologias da Informação, delineando o que se chama de Sociedade em Rede (Castells, 2000). Fenômenos já consolidados, tais como o *Big Data*, caracterizam o atual cenário de crescente disponibilidade de dados em grande variedade e volume de formatos e alta velocidade de disseminação e de acesso, ampliando o potencial de uso como recurso estratégico (Coneglian; Santarem Segundo; Sant'Ana, 2017).

Este cenário é composto por uma série de elementos que envolvem humanos e máquinas. Os dados são gerados a partir de atividades humanas e processadas por agentes maquinicos. No tratamento dos dados, os processos de interação com a máquina são apresentados caminhos específicos em suas interfaces seja de dispositivos ou de sensores que ampliam a capacidade de coleta maquinica, tais dados sofrerão novas transformações para que possam ser registrados em suportes digitais, cada qual com suas características - ópticas, magnéticas ou elétricas (Sant'Ana, 2021).

Uma vez coletados, esses dados vão passar por sucessivas transformações sobre as quais sujeitos são totalmente inscientes (Affonso; Sant'Ana, 2018). Por exemplo, um sujeito que utilize um *smartphone* deixa uma trilha de dados por onde quer que passe, no qual o dispositivo tecnológico gera dados de georeferenciamento que podem ser coletados pela operadora de telefonia móvel e, eventualmente, pelas aplicações instaladas, sendo possível assim traçar um mapa de deslocamento do sujeito (Dias; Vieira, 2013).

O exemplo apresentado configura uma situação de coleta de dados que pode ser realizada por um agente que atua principalmente na definição de quais dados serão



coletados, isto é, a operadora de telefonia móvel ou as aplicações instaladas definem quais dados serão obtidos do sujeito. Entre os diferentes dados coletados encontram-se os dados pessoais. A obtenção desses dados sem consentimento do usuário ou o tratamento sem orientações de legislações de proteção pode pôr em risco a privacidade dos referenciados, sujeitos que de alguma forma são identificáveis nos dados, e que são usuários ou sujeitos passíveis de identificação dentro de um sistema de informação (Rodrigues, 2024).

Os ambientes digitais oferecem infraestrutura para a realização de serviços *online* que utilizam dados do usuário, isto é, a coleta de dados pessoais pode ocorrer nesses ambientes com a permissão do usuário ou de forma insciente. O modo e a descrição de como ocorre essa obtenção deveria estar descrito em suas políticas de privacidade.

Na coleta de dados é possível realizar procedimentos fragmentados, ou seja, pequenas porções de conjuntos de dados de um determinado sujeito. Resultados que decorrem da reconstituição de dados fragmentados recompondo um novo todo, o efeito mosaico, têm potencial de revelar informações pessoais mesmo quando resultante de diferentes fontes e até mesmo de dados, a princípio inócuos em relação à privacidade quando tratados individualmente. É a reconstituição de uma nova visão sobre o sujeito a partir de fragmentos de dados coletados em diferentes momentos, fontes ou formas (Simionato, Sant'Ana e Santos, 2013).

As políticas de privacidade são documentos disponibilizados pelos detentores de serviços *online* e obrigatórios pela legislação brasileira de dados, cujo conteúdo versa sobre informações da coleta de dados pessoais e as concessões concedidas ao detentor para o tratamento dos dados. Destaca-se que a utilização dos serviços *online* pode ocorrer de forma gratuita, sem o pagamento de taxas ou mensalidades, desde que seja concedido a permissão de coleta de dados pelo usuário.

As políticas de privacidade de serviços *online* possuem alta relevância investigativa, pois descrevem uma parcela dos dados obtidos, entretanto não apresentam informações suficientes de transparência sobre o que é realizado com os dados e, sobretudo, como são descartados. As políticas descrevem as situações que podem ser obtidas os dados, inclusive quando não se utiliza os serviços *online*, e portanto insere-se como foco do problema desta pesquisa.

Assim, apresenta-se como objetivo desta pesquisa: caracterizar o efeito mosaico na coleta de dados pessoais em serviços *online*, sob a perspectiva da Ciência da Informação, apresentando implicações para a privacidade dos usuários. A justificativa está centralizada na necessidade de reflexões sobre o alcance e o controle dos detentores sobre o aumento



do repertório de dados coletados sobre os usuários.

2 EFEITO MOSAICO

O efeito mosaico está alinhado no campo dos estudos teórico-práticos da Ciência da Informação, sobretudo à linha de dados, informação e tecnologia. Portanto, esta pesquisa buscou investigar o efeito mosaico a partir do referencial teórico presente na Ciência da Informação baseada no conceito de insciência do usuário (Affonso; Sant'Ana, 2018), Ciclo de Vida dos Dados na Ciência da Informação (Sant'Ana, 2016), processo de análise de dados em contexto de *Big Data* (Taurion, 2013), integração de dados e granularidade (Bione, 2020; Pozen, 2005; Sant'Ana, 2016; Santos; Sant'Ana, 2013; Taurion, 2013).

As preocupações sobre a necessidade de entender os dados em contexto de *Big Data* não são recentes, assim como o uso por organizações. Dias e Vieira (2013) já apresentaram questões éticas e legais emergentes que permeiam a privacidade e uso de dados por organizações. A partir das discussões dos autores, o uso de dados por organizações possibilita a criação de perfis que facilitem a aquisição de novos produtos e serviços para os usuários. Esses perfis transcendem a percepção do que os usuários acham sobre o que as organizações conhecem a seu respeito e como os dados para a construção do perfil foram obtidos, uma vez que conhecidos os hábitos dos usuários podem servir como vantagem para as organizações (Dias; Vieira, 2013).

Na criação desses perfis se percebe o conceito de insciência (Affonso; Sant'Ana, 2018), pois os usuários que utilizam serviços são alvos de processos de coleta de dados que desconhecem. Adiciona-se a coleta uma complexidade encapsulada nos processos de tratamento dos dados, que podem ocasionar problemas relacionados à privacidade devido à insciência dos usuários sobre o que é coletado, como será utilizado e quais informações podem ser obtidas a partir de análises (Affonso; Sant'Ana, 2018).

Tratar o tema sobre o volume de dados disponibilizados aos serviços *online* é um desafio frente às questões relacionadas ao acesso a dados, principalmente, aqueles que são considerados pessoais. As consequências desse desafio são as discussões sobre uma legislação acerca da utilidade dos dados e aspectos de privacidade decorrentes do crescente volume, e o seu potencial utilização para atender as demandas por produtos e serviços.

Para iniciar a discussão relacionada ao acesso a dados, adota-se o entendimento de dado como uma unidade de conteúdo de baixa carga semântica, necessariamente



relacionada a determinado contexto e composta pela tríade entidade, atributo e valor - <e,a,v> (Santos; Sant'Ana, 2013).

A entidade, quando contextualizada, representa um objeto do mundo real por meio da identificação dos seus atributos, vinculando-os ao terceiro elemento (valor) que registra o conteúdo. Por exemplo, a entidade *Pessoa*, possui um atributo *Nome*, de valor *Albert Einstein*. Para outra entidade *Pessoa*, o atributo *Nome* é *Douglas Engelbart*, e assim por diante. Neste sentido, a composição da tríade <e,a,v> também se aplica a dados pessoais.

A legislação brasileira define como dados pessoais quaisquer informações relacionadas à pessoa natural identificada ou identificável e controlador a pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de dados (Brasil, 2018). Nesta pesquisa, o controlador foi denominado como detentor, ou seja, aquele que detém o serviço *online* e, por consequência, os dados armazenados pelo serviço.

Utilizar serviços *online* possibilita ao detentor atribuir a cada sujeito um número identificador (ID) que permite, por exemplo, identificar seus atributos e valores. No exemplo supracitado, é adicionado um atributo ID para cada Pessoa, geralmente formado por um número inteiro ou o resultado de uma função *hash*¹ – o que permite identificar cada pessoa por um ID, evitando, por exemplo, nomes homônimos. Diferente de outros suportes, o uso de serviços *online* gera grandes quantidades de dados que podem ser analisados, caracterizando-os como parte do fenômeno *Big Data*.

Na atribuição do ID, o fator integração é uma preocupação da coleta por meio da identificação e validação dos atributos que serão responsáveis pela identificação unívoca de cada registro para que a integração possa ser garantida (Sant'Ana, 2016). Este fator é elemento chave para a recomposição dos dados de modo a reconstituir um mosaico dos dados.

Para Taurion (2013), há a existência de quatro atividades dentro do processo de análise de dados em *Big Data*: coleta de dados, processo de limpeza, integração ou agregação e atividade analítica. A coleta dos dados ocorre de diferentes fontes, como em sistemas transacionais, câmeras de vigilância, entre outros. O processo de limpeza, formatação e validação dos dados são realizados para que sejam armazenados e estruturados dados que não sejam incompletos ou inconsistentes. A integração dos dados

¹Uma função de *hash* criptográfico transforma qualquer bloco de dados em uma série de caracteres de comprimento fixo, são largamente utilizadas para buscar elementos em bases de dados, verificar a integridade de arquivos baixados.

possui alta relevância para caracterização do efeito mosaico, pois ainda que os dados sejam anonimizados é possível o cruzamento de dados de diversas fontes revelar informações pessoais. Por fim, ocorre a atividade analítica, quando se cumpre a análise e interpretação dos resultados obtidos.

Segundo Bione (2020), a antítese do conceito de dado pessoal seria um dado anônimo, ou seja, aquele que é incapaz de revelar a identidade de um sujeito. Entretanto, a integração de diversos dados pode revelar a imagem do quebra-cabeça, a qual era até então desfigurada (anônimo), caracterizando o chamado efeito mosaico.

Processos de coleta de dados são realizados por meio de dispositivos que podem atuar sem a participação ativa do usuário, inicialmente considerados como não identificadores, e que, portanto, ficam livres de controle ou de desconfianças mais diretas. No entanto, quando esses dados são integrados com outros dados, para subsequente tratamento, eles podem gerar o que se denomina de efeito mosaico e levar à identificação e, potencialmente, à violação de privacidade (Sant'Ana, 2020).

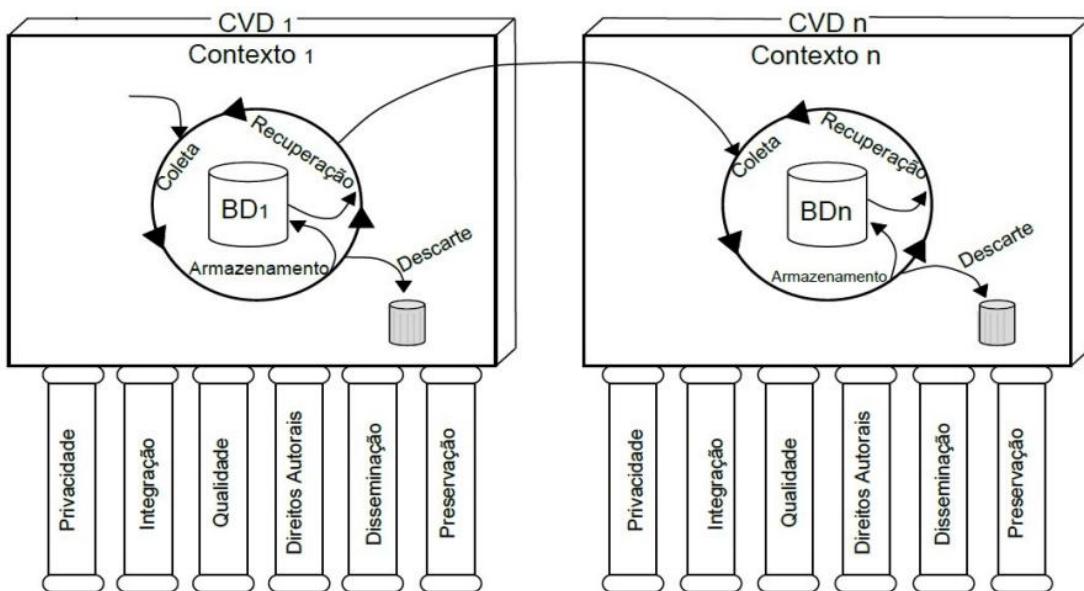
Portanto, o efeito mosaico corresponde à integração de dados originados de fontes diversas capazes de transformar esses dados em informações pessoais, permitindo a identificação de usuários de serviços *online* em determinados contextos que podem caracterizar uma potencial violação de privacidade (Pozen, 2005).

No caso dos dados pessoais, a interpretação do fluxo pelo agente por meio de uma estrutura formal, no caso, aplicando o Ciclo de Vida dos Dados (CVD) na Ciência da Informação permite a identificação de fases que indicam momentos, características e requisitos para o tratamento dos dados (Sant'Ana, 2016).

O CVD é composto por quatro fases: Coleta, que são estratégias para localizar e avaliar os dados, definindo as ferramentas necessárias; Armazenamento, aonde há a definição de quais dados serão disponibilizados, assim como a estrutura física e lógica que será utilizada para o registro; Recuperação, no qual se viabiliza que os dados sejam encontrados, acessados e passíveis de interpretação por humanos ou por máquinas, e; Descarte, aonde ocorre a limpeza ou desativação da base, ou apenas a atividade de apagar o registro inteiro ou um atributo específico. Também existem fatores transversais a todas as fases: Privacidade, Integração, Qualidade, Direitos Autorais, Disseminação e Preservação (ver Figura 1) (Sant'Ana, 2016).



Figura 1: Estrutura do Ciclo de Vida dos Dados na Ciência da Informação.



Fonte: Sant'Ana (2016).

Observa-se que a existência de um contexto terá um CVD₁ com as quatro fases sustentadas pelos seis fatores transversais. O CVD₁ poderá disponibilizar dados por meio da sua fase de Recuperação, que pode ser acessado pela fase de Coleta de um outro ciclo, em distintos contextos (CVD_n). Ou seja, o CVD_n não necessariamente é um processo da instituição detentora dos dados, fazendo com que este dado esteja vinculado legislações distintas do que fora acordado no CVD₁, inclusive, em tese, podendo ser vinculado até a legislações de outros países, no caso de uma transferência internacional dos dados – este último fenômeno sendo abordado por pesquisas sobre soberania de dados.

Desse modo, a instrumentalização por meio do CVD para perceber o efeito mosaico está na possibilidade de observar a identificação dos momentos que os dados pessoais são tratados em um contexto. O compartilhamento de dados que ocorre na interação das fases de Recuperação e Coleta entre diferentes ciclos aumentam a possibilidade de identificação do usuário, ainda que esses dados tenham sido anonimizados. Nesse caso, podem ocorrer novas interpretações a partir dos dados coletados não previstos inicialmente e isto pode caracterizar um potencial de violação de privacidade.

A relevância de investigar e de discutir o efeito mosaico está no potencial de identificação de usuários por meio da coleta de dados que inicialmente são considerados como não identificáveis, mas ao tratá-los apresentam implicações para a privacidade. Esse efeito também está relacionado ao aspecto da granularidade do conjunto de dados.

A granularidade de um conjunto de dados está vinculada ao número de atributos que o compõem e a diversidade de seus conteúdos [...] Quanto maiores as possibilidades de

obter subconjuntos a partir de um conjunto de dados, maiores serão as potencialidades de tratamento e de elaboração de resultados [...] e, quanto maiores as possibilidades, maior será o detalhamento disponível, o que determina a definição da granularidade como fina (Santos; Sant'Ana, 2013, p. 204).

Por exemplo, se um conjunto de dados estruturado contiver apenas a média do tempo de uso de um aplicativo por região do país, a granularidade será mais grossa que outro conjunto de dados que contenha as médias do tempo de uso por estado brasileiro. E a granularidade pode ser ainda mais fina em um conjunto, se estiverem disponíveis as médias por cidade, por bairro, até chegar a um único sujeito (Santos; Sant'Ana, 2013).

Adiciona-se a importância investigativa do efeito mosaico quando os conjuntos de dados são usados para tomada de decisão em organizações com fins lucrativos. Nesse contexto, dados são descritos como fundamentais para as organizações, bem como para identificar as macrotendências em um número de diferentes setores, incluindo educação, saúde, transporte e segurança (OECD, 2013).

De acordo com Silveira, Avelino e Souza (2016) sob a perspectiva estritamente econômica, o uso de dados pessoais poderia reduzir as assimetrias da informação e contribuir para aumentar a eficiência das transações econômicas nas redes digitais. Entretanto, os autores destacam que a análise crítica e as implicações sociais são pouco consideradas nas práticas discursivas das organizações. A quantidade de dados pessoais captados e armazenados para o uso pelo capital aumenta quanto mais cresce o uso das redes de serviços, informações e entretenimento.

3 PROCEDIMENTOS METODOLÓGICOS

Esta pesquisa se caracteriza com objetivo exploratório, de abordagem qualitativa e natureza descritiva. Utilizou-se o método a Análise de Conteúdo e a técnica de análise categorial (Bardin, 2011). Analisou-se as mensagens contidas nas políticas de privacidade de dois serviços *online* mais utilizados no ano de 2023 no Brasil (Migliani, 2024). A organização da análise se realizou em três fases: pré-análise, exploração do material e tratamento dos resultados (Bardin, 2011).

A fase de pré-análise correspondeu à seleção dos documentos, preparação do material (*corpus*) e a realização da leitura flutuante (Bardin, 2011). Os documentos selecionados são as políticas de privacidade dos serviços *online*, cujo critério de seleção foi o serviço mais utilizado em duas categorias: Serviços de Redes Sociais *Online* e



Streaming de música, respectivamente, o TikTok e o Spotify. A preparação do material ocorreu com o acesso às políticas em formato *Portable Document Format* (PDF) retirado diretamente dos *websites* dos serviços referente a versão do ano de 2024 e, por fim, a realização da leitura flutuante para extrair as unidades de análise (Bardin, 2011).

A fase de exploração do material ocorreu com a aplicação da técnica de análise categorial por meio da codificação temática que consiste em descobrir os núcleos de sentido que compõem os parágrafos da política de privacidade (Bardin, 2011), optando por definir seis categorias antecipadamente.

As categorias de dados foram elaboradas a partir dos indícios que contextualizaram os dados passíveis de serem coletados, observando as características comuns para agrupá-los. Desse modo, elegeu-se seis categorias (*Conta, Pagamento, Conteúdo, Uso do serviço online, Técnicos e Parceiros, anunciantes, outros serviços*).

As categorias possibilitaram organizar os dados descritos na política de privacidade. Por exemplo, as categorias de dados *Conta* e *Pagamento* foram organizadas da seguinte forma: a primeira reúne dados pessoais obrigatórios e necessários para utilizar os serviços, como o nome e um e-mail; a segunda categoria contém dados relacionados a forma de pagamento e são opcionais, porém serão obrigatórios se houver o desejo do usuário de realizar uma compra, nesse caso será requisitado dados sobre cartão de crédito, débito, etc.

O passo posterior foi observar as situações que esses dados podem ser coletados pelo detentor descritas na política de privacidade. Quando o usuário desencadeia a ação de envio dos dados é denominado *Fornecido pelo usuário*. Em contrapartida, quando os dados são obtidos sem a interferência do usuário isso quer dizer que foram *Coletados automaticamente pelo detentor*, por fim, quando os dados são enviados por terceiros chama-se essa ação de *Recebidos pelo detentor via Agentes externos*. Essas situações foram chamadas nesta pesquisa de *Fonte*, porque foi observado na política de privacidade como o ponto de partida para que os dados sejam coletados.

O Quadro 1 apresenta as três fontes de dados onde o detentor pode obter dados sobre o usuário, seguido do agrupamento de seis categorias e, por fim, a descrição deles.

Quadro 1: Categorias definidas antecipadamente para identificação do processo de coleta de dados pelos serviços

Fonte	Categoria de dados	Descrição
Fornecido pelo usuário ao detentor	Conta	São dados pessoais obrigatórios e necessários para criar um cadastro no serviço para que seja atribuído um identificador único.



	Pagamento ¹	São dados pessoais relacionados a forma de pagamento em cartões de crédito, débito, boleto ou transferências bancárias, incluindo endereço de faturamento.
	Conteúdo	São dados relacionados a mídias audiovisuais (fotografia, vídeo ou áudio) produzidas pelo usuário.
Coletados automaticamente pelo detentor	Uso do serviço <i>online</i>	São dados pessoais coletados e tratados quando o usuário usa o serviço.
Recebidos pelo detentor via Agentes externos	Técnicos	São dados relacionados a identificadores e dispositivos de <i>hardware/software</i> utilizados pelo usuário para acessar ou usar os serviços.

¹ São dados opcionais, pois só serão fornecidos caso o usuário realize uma compra.

Fonte: Elaborado pelos autores (2024).

Por último, a fase de tratamento dos resultados corresponde à análise do efeito mosaico em cada serviço *online* apresentando as implicações para a identificação de usuários. Inicialmente, foi delineado os principais resultados encontrados a partir das análises dos serviços *online*, como, por exemplo, a identificação dos dados descritos nas políticas de privacidade e os Agentes externos, a explicitação de como ocorre o efeito mosaico por meio da estrutura do Ciclo de Vida dos Dados. Em seguida, cada serviço *online* foi analisado individualmente para observar o efeito mosaico e as implicações para a privacidade a partir dos dados coletados pelo detentor e, por fim, o papel desempenhado pelos Agentes externos.

4 RESULTADOS E DISCUSSÃO

Como resultado, foram identificadas três fontes de coleta de dados pessoais: i) fornecidos pelo usuário ao detentor (são dados necessários para o cadastro, para pagamentos e relacionados às mídias audiovisuais), ii) coletados automaticamente pelo detentor (dados gerados a partir do uso dos serviços e dados técnicos relacionados a identificadores de *hardware/software*) e iii) dados pessoais recebidos pelo detentor via Agentes externos.

Neste último, as políticas de privacidade não apresentaram as características dos dados que recebem por terceiros, isto é, não havia elementos suficientes para identificar a entidade e os atributos dos dados recebidos via Agentes externos. Por esse motivo, foi possível apenas identificar os Agentes externos que compartilham dados com o detentor

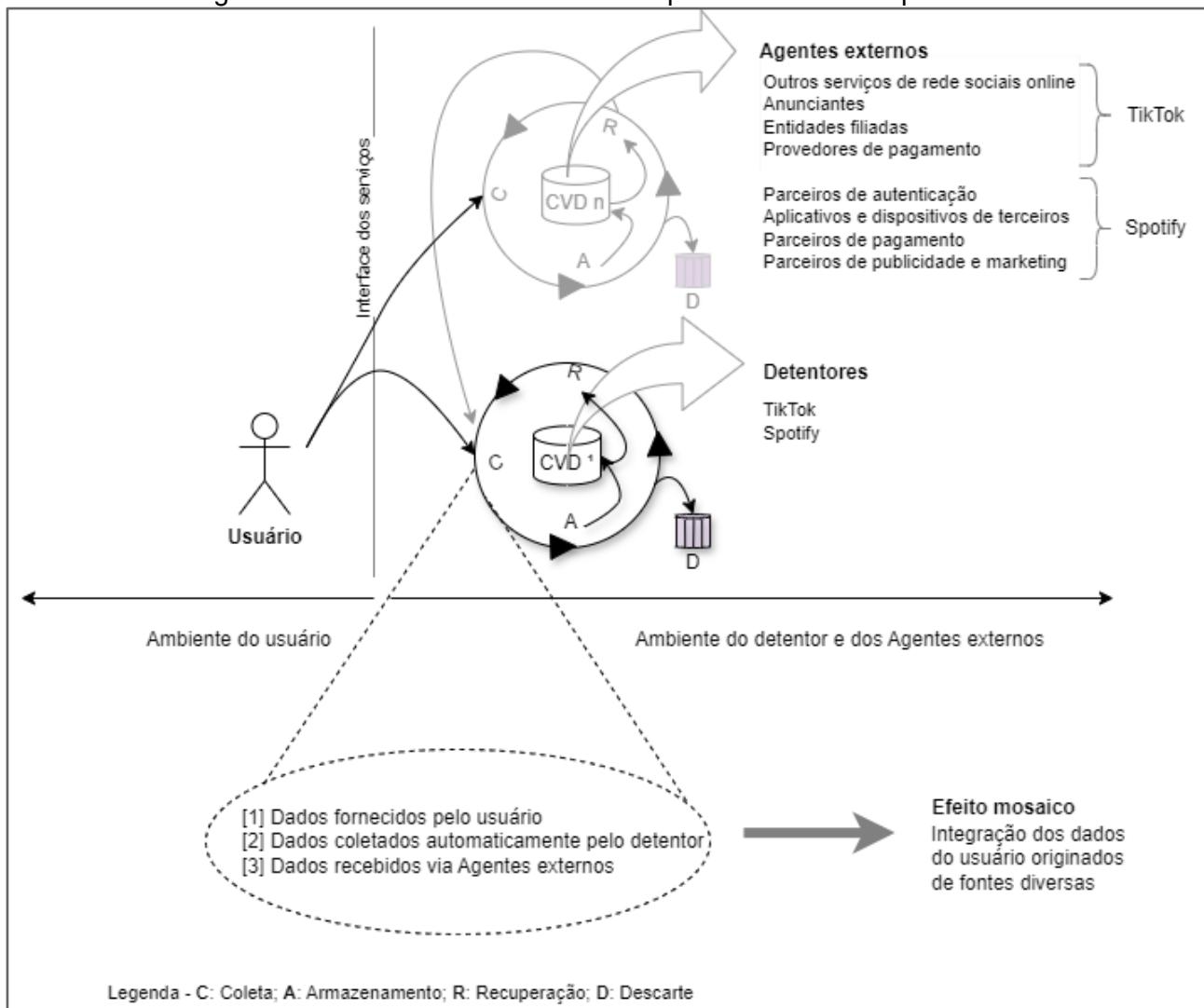


do serviço *online* e a descrição do papel desempenhado.

Os dados das três fontes quando contextualizados podem revelar informações pessoais não previstas inicialmente na fase de coleta. Por exemplo, a integração de dados como *e-mail*, endereço de faturamento, histórico de navegação e pesquisa aliado às informações dos dispositivos que utiliza para acessar os serviços *online* podem apresentar indícios sobre intenções de adquirir produtos ou inferir sobre a sua renda, podendo direcionar anúncios a partir dessas análises.

Para explicitação das três fontes de coleta de dados pessoais foi elaborada a Figura 2, que ilustra o processo de coleta de dados realizado pelos detentores analisados (TikTok e Spotify) que contribui para a caracterização do efeito mosaico.

Figura 2: Processo de coleta de dados pessoais realizada pelos detentores



Fonte: Elaborado pelos autores (2024).



Os dados pessoais podem ser coletados de duas formas pelo detentor: dados fornecidos pelo usuário e dados coletados automaticamente, a partir do uso (na interface dos serviços *online*). Além disso, o usuário também pode fornecer seus dados pessoais a outros Agentes externos (CVD_n), estes poderão disponibilizar os dados do usuário para o detentor (TikTok e Spotify) do serviço (CVD₁) através fase de Recuperação.

Os dados recebidos pelo detentor via Agentes externos (CVD_n) possuem maior potencial de identificação de usuários e implicações para a privacidade, pois os usuários podem ser inscientes sobre o compartilhamento de seus dados pessoais entre os Agentes externos e os detentores. Observou-se que existem ainda diferentes agentes que podem disponibilizar dados sobre o usuário que contribuem para a integração de dados.

Para a caracterização do efeito mosaico, a identificação das fontes de dados é necessária para entender a origem e como ocorre a reconstituição de dados fragmentados para recompor um novo todo. Sendo assim, observou-se as fontes de dados coletadas pelo detentor e as possíveis transformações que podem caracterizar o efeito mosaico na identificação de usuários, em cada política de privacidade.

4.1 Política de Privacidade do Serviço de Rede Social *Online* TikTok

O efeito mosaico pode ser caracterizado a partir da combinação dos dados fornecidos pelo usuário e associados aos dados coletados automaticamente pelo detentor (TikTok). O Quadro 2 apresenta os dados identificados na fase de Coleta do detentor, seguido da descrição deles.

Quadro 2: Dados identificados na fase de Coleta do TikTok

Fonte	Categoria de dados	Descrição dos atributos identificados na política
Fornecido pelo usuário ao detentor	Conta	nome de usuário; senha; data de nascimento; e-mail; número de telefone.
	Pagamento	moedas TikTok; dados sobre a compra; número do cartão de pagamento; faturamento.
	Conteúdo	fotografias, áudios, vídeos carregados ou criados; comentários; hashtags; feedback; avaliações e transmissões ao vivo; quando escreve, envia ou recebe mensagens.
Coletados automaticamente pelo detentor	Uso do serviço <i>online</i>	acesso a Plataforma; anúncios; os vídeos que vê; o histórico de navegação e pesquisa; o conteúdo que gosta; o conteúdo salvo em “Meus Favoritos”; seguidores; seguidores mútuos; inferência dos atributos; interesses, gênero e faixa etária.
	Técnicos	endereço IP; agente do usuário; operadora de celular; configurações de fuso horário; modelo do dispositivo; sistema do dispositivo; tipo de rede; IDs do dispositivo;



		resolução de tela e sistema operacional; nomes; tipos de aplicativos; arquivos, padrões ou ritmos de digitação; estado da bateria, configurações de áudio e dispositivos de áudio conectados; login a partir de vários dispositivos; localização aproximada; localização com base no cartão SIM e/ou endereço IP; cookies e outras tecnologias semelhantes
--	--	--

Fonte: Dados de pesquisa.

Os dados apresentados no Quadro 2 podem ser combinados de modo a buscar novas interpretações em novos contextos. Por exemplo, os dados do usuário como o nome, a data de nascimento e o número de telefone servem como identificadores. A partir disso, se for associado aos dados sobre as compras realizadas e avaliações feitas no Serviço de Rede Social *Online* sobre um produto é possível mensurar as chances de adquiri-lo quando também se associa a faixa etária, além de atribuir nessa análise dados sobre a localização de outros usuários com o mesmo perfil socioeconômico. O modo de análise sobre a aderência ou não de um produto, a partir da compilação desses dados, também pode ser utilizado para direcionar propagandas que podem gerar constrangimentos aos usuários, reforçar preconceitos, direcionar, por exemplo, conteúdos políticos, religiosos ou de quaisquer segmentações sociais, entre outras possibilidades.

A extensão da coleta de dados realizada pelo detentor contribui para o efeito mosaico quando esta passa a receber dados dos usuários via Agentes externos. O Quadro 3 apresenta os Agentes externos identificados na política de privacidade do Serviço de Rede Social *Online* TikTok que podem enviar dados pessoais do usuário.

Quadro 3: Agentes externos identificados no serviço de rede social *online*

Agentes externos	Descrição do papel desempenhado pelo agente externo no compartilhamento dos dados identificados na política
Outros serviços de sociais <i>online</i>	Registrar ou usar a Plataforma usando informações de uma conta de rede social <i>online</i> de terceiros.
Anunciantes	Ações tomadas fora da Plataforma, como atividades em outros sites e aplicativos ou em lojas, incluindo os produtos ou serviços que comprou on-line ou pessoalmente.
Entidades filiadas	Dentro do grupo corporativo, inclusive sobre atividades em suas plataformas.
Provedores de pagamento	Informações de confirmação de pagamento e informações sobre a entrega de produtos que comprou por meio de recursos de compras.

Fonte: Dados de pesquisa.

Desse modo, o TikTok recebe dados do usuário quando se registra ou usa a plataforma de outros Serviços de Redes Sociais *Online*, além de coletar dados sobre as ações realizadas em outros sites e aplicativos, sobretudo na aquisição de produtos e



serviços. As entidades filiadas também compartilham dados entre as filiadas do grupo corporativo sobre as atividades do usuário. O Serviço de Rede Social *Online* também recebe dados sobre confirmação de pagamento e sobre a entrega dos produtos por meio dos provedores de pagamento.

4.2 Política de Privacidade do *Streaming* de música Spotify

O Spotify coleta dados fornecidos pelo usuário e dados coletados automaticamente pelo detentor quando usa a interface do serviço. O Quadro 4 apresenta os dados identificados na fase de Coleta do detentor a partir da política de privacidade, seguido da descrição deles.

Quadro 4: Dados identificados na fase de Coleta do Spotify

Fonte	Categoria de dados	Descrição dos atributos identificados na política
Fornecido pelo usuário ao detentor	Conta	Nome do perfil; e-mail; senha; número de telefone; data de nascimento; gênero; endereço; país e instituição de ensino superior (no spotify premium universitário).
	Pagamento	Forma de pagamento (cartões de crédito ou débito), o tipo de cartão, a data de validade, certos dígitos do número do cartão; CEP; detalhes do histórico de compras e pagamentos.
	Conteúdo	Imagens, áudio, texto, nomes, descrições e outros tipos de conteúdo.
Coletados automaticamente pelo detentor	Uso do serviço <i>online</i>	Consultas de buscas; histórico do streaming; playlists e biblioteca criada; histórico de navegação; configurações da conta; interações com outros usuários do Spotify; uso de serviços; Inferências de interesses e preferências com base no uso.
	Técnicos	Informações de <i>Uniform Resource Locator</i> (URL) Identificadores online, tais como dados de cookies e endereços IP; informações sobre os dispositivos usados, como: IDs dos dispositivos; IP de conexão de rede, por exemplo, Wi-Fi, 4G, Long-Term Evolution (LTE) e <i>Bluetooth</i> ; provedor; desempenho de rede e dispositivo; tipo de navegador; idioma; informações que permitem gestão de direitos digitais; sistema operacional; versão do aplicativo Spotify; dados de sensores de dispositivos gerados por movimento ou orientação.

Fonte: Dados de pesquisa.

A partir da identificação das fontes de dados, o efeito mosaico pode ser caracterizado quando os atributos do usuário são combinados com os atributos relacionados ao uso do serviço aliado às características técnicas. Por exemplo, as informações cadastrais do perfil



(nome, número de telefone e endereço) podem ser agregadas ao histórico de navegação e inferências com base nesse uso. Ainda, os dados técnicos de localização permitem identificar as regiões onde o usuário mais frequenta, traçar inclusive uma rota diária a partir do tempo de conexão nas redes acessadas. O resultado desta análise poderia revelar a frequência e a permanência do usuário em um determinado local, caracterizando uma ação de vigilância.

Contribuindo para a reidentificação do usuário, o compartilhamento de dados pessoais via Agentes externos favorece o efeito mosaico quando são recebidos dados pessoais do usuário. O Quadro 5 apresenta os Agentes externos e a descrição do papel desempenhado pelo agente no compartilhamento dos dados.

Quadro 5: Agentes externos identificados no Spotify.

Agentes externos	Descrição do papel desempenhado pelo Agente externo no compartilhamento dos dados
Parceiros de autenticação	Se você se cadastrar ou acessar o Serviço Spotify usando outro serviço, receberemos suas informações desse terceiro.
Aplicativos e dispositivos de terceiros	Se você conectar sua conta Spotify a um aplicativo, serviço ou dispositivo de terceiros, poderemos coletar e usar as informações deles. Isso serve para tornar a integração possível.
Parceiros de serviços técnicos	Trabalhamos com parceiros de serviços técnicos que nos fornecem certos dados. Isso inclui mapeamento de endereços de <i>Internet Protocol</i> (IP) para dados de localização não precisos (por exemplo, país ou região, cidade, estado).
Parceiros de pagamento	Se você escolher pagar por um terceiro (por exemplo, operadoras de telecomunicações) ou por fatura, podemos receber os dados dos nossos parceiros de pagamento.
Parceiros de publicidade e <i>marketing</i>	Recebemos inferências de certos parceiros de publicidade ou marketing. Essas inferências são o entendimento dos parceiros de seus interesses e preferências.

Fonte: Dados de pesquisa.

O efeito mosaico pode ocorrer na reconstituição dos dados já obtidos aplicando uma nova interpretação no novo contexto. Entre esses dados estão aqueles relacionados ao cadastramento em outros serviços envidados por meio de parceiros de autenticação. A partir disso, esses dados podem ser combinados com os recebidos de parceiros de serviços técnicos, que fornecem dados de mapeamento de endereço *Internet Protocol* (IP) para identificar a região onde o usuário se encontra, podendo ser integrado com dados de pagamento para direcionar publicidade e *marketing* a partir de inferências socioeconômicas.



4.3 Limitações do estudo e indicação de pesquisas futuras

A percepção do efeito mosaico está na camada de abstração dos sistemas de informação que estruturam o ambiente informacional necessário para viabilizar os serviços *online*. Para a sua caracterização foi necessário a análise minuciosa da política de privacidade e apesar da disponibilidade aos usuários, a leitura e a compreensão das condições exigem conhecimentos prévios para uma análise criteriosa, como, por exemplo, conhecimentos de como ocorre a integração de dados. Isso pode ser uma barreira para que os usuários sejam conscientes sobre o tratamento dos dados coletados por detentores e Agentes externos, aumentando a assimetria informacional.

Para alterar esse cenário é importante implementar elementos da privacidade desde o design como incentivar o desenvolvimento de serviços *online* que incorporem a privacidade como princípio fundamental desde a concepção, minimizando a coleta e o compartilhamento de dados pessoais. As técnicas de *User Experience* (UX) também podem ser utilizadas para informar ao usuário sobre a coleta de dados feita pelos detentores dos serviços, inclusive expondo quais os elementos constituem os dados pessoais, como as entidades e os atributos descritos nas políticas de privacidade, além da indicação de quais dados pessoais não podem ser compartilhados com Agentes externos. Pesquisas de UX com fins acadêmicos, comerciais e científicos se tornaram essenciais para projetar e avaliar essas experiências com foco nos usuários (Barros, et. al., 2023).

Outros fatores que contribuíram para as limitações do estudo, foram a ausência de elementos que possibilitassem a análise das fases posteriores à coleta de dados no CVD, como as fases de Armazenamento e Descarte. O acesso a essas fases poderia contribuir com o aprofundamento da discussão sobre as soluções para mitigar os riscos do efeito mosaico, explorando alternativas como a anonimização de dados mais robustos que dificultem a reidentificação dos usuários a partir da combinação de diferentes conjuntos de dados. Nesse sentido, é possível convergências interdisciplinares entre Ciência da Informação e Ciência de Dados para contribuir no desenvolvimento de suportes tecnológicos cada vez mais eficazes (Moutinho, et. al. 2024).

Como indicação de pesquisas futuras há necessidade de comparar as políticas de privacidade com o objetivo de verificar se as práticas de coleta e uso de dados descritas nas políticas dos serviços *online* TikTok e Spotify se refletem de fato. Essa análise poderia ser realizada por meio de testes e estudos de caso, complementando a análise documental.



5 CONSIDERAÇÕES FINAIS

As considerações finais apontam para a necessidade de reflexões sobre o alcance da coleta e o controle dos dados pessoais pelo detentor. O aumento do repertório de dados coletados sobre os usuários a partir da utilização de serviços *online* pode agravar a assimetria informacional entre os titulares dos dados (usuários) e o controlador (serviços oferecidos pelas plataformas digitais) quando há ausência de transparência do tratamento dos dados.

Essa assimetria ocorre quando as camadas de abstrações inerentes aos sistemas de informação distanciam o usuário de conhecer o tratamento dos seus dados pessoais, podendo ocasionar implicações para a privacidade. Aliado a uma visualização textual extensa de diversas informações distribuídas nas políticas de privacidade que contém terminologias jurídicas, dificultando a explicitação dos dados coletados e armazenados após o tratamento pelos detentores.

Os fenômenos decorrentes da reconstituição de dados fragmentados para recompor um novo todo, efeito mosaico, tem potencial de revelar informações pessoais ainda que resultante de diferentes fontes de dados, a princípio inócuos em relação à privacidade quando tratados individualmente. Este efeito tem potencial para ser objeto de análise da Ciência da Informação, pois reúne aspectos operacionais do funcionamento de bancos de dados e informacionais quanto à interpretação resultante da análise do tratamento dos dados para tomada de decisão. Sendo esta pesquisa uma contribuição de natureza teórica para estudos futuros sobre o efeito mosaico.

Em síntese, esta pesquisa teve como objetivo caracterizar o efeito mosaico na coleta de dados pessoais em serviços *online*, sob a perspectiva da Ciência da Informação, apresentando implicações para a privacidade dos usuários. Destaca-se que a necessidade de investigação do efeito mosaico é uma das consequências ocasionadas pelo aumento da complexidade e do volume de dados disponíveis no ambiente digital capaz de gerar novas interpretações com potencial de identificação de usuários e, por consequência, implicações para a sua privacidade.

Dados pessoais têm sido utilizados para direcionar conteúdos que estão além de preferências consumeristas. A insciência dos usuários sobre o que é feito com seus dados os distancia de uma análise crítica sobre os potenciais efeitos para a privacidade e sociedade, pois esses dados podem ser usados para tomada de decisão que afetam a coletividade.



REFERÊNCIAS

AFFONSO, E. P.; SANT'ANA, R. C. G. Privacy awareness issues in user data collection by digital libraries. **IFLA Journal**, , v. 44, n. 3, p. 170–182, 1 out. 2018. Disponível em: <https://journals.sagepub.com/doi/10.1177/0340035218777275>. Acesso em: 9 jun. 2024.

BARDIN, L. **Análise de conteúdo**. São Paulo: Edições 70, 2011.

BARROS, S.T; VIDOTTI, S. A. B. G.; RODAS, C. M.; SANT'ANA, R. C. G. O ciclo de vida dos dados no contexto da pesquisa em User eXperience. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 28, p. 1–24, 2023. DOI: 10.5007/1518-2924.2023.e83434. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/83434> . Acesso em: 25 nov. 2024.

BIONE, B. R. Compreendendo o conceito de anonimização e dado anonimizado. **Direito Digital e proteção de dados pessoais**, São Paulo, v. 21, n. 53, p. 191–201, 2020. Disponível em:

https://www.mpsp.mp.br/portal/page/portal/documentacao_e_divulgacao/doc_biblioteca/bibli_servicos_produtos/bibli_boletim/bibli_bol_2006/Cad-Juridicos_n.53.pdf. Acesso em: 17 jun. 2024.

BRASIL. **Lei Geral de Proteção de Dados Pessoais**. Lei 13.709 de agosto de 2018. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/ato2015-2018/lei/l13709.htm. Acesso em: 17 jun. 2024.

CASTELLS, M. **Sociedade em Rede**. São Paulo: Terra e Paz, 2000.

CONEGLIAN, C. S.; SANTAREM SEGUNDO, J. E.; SANT'ANA, R. C. G. Big Data: fatores potencialmente discriminatórios em análise de dados. **Em Questão**, Rio Grande do Sul, , p. 62–86, 1 jan. 2017. Disponível em: <https://doi.org/10.19132/1808-5245231.62-86> . Acesso em: 9 jun. 2024.

DIAS, G. A.; VIEIRA, A. A. N. Big data: questões éticas e legais emergentes. **Ciência da Informação**, Brasília, v. 42, n. 2, 2013. Disponível em: <https://revista.ibict.br/ciinf/article/view/1380>. Acesso em: 17 jun. 2024.

GONZÁLEZ DE GÓMEZ, M. N. As ações de informação e seus contextos: aportes da pragmática ao campo investigativo da ciência da informação. João Pessoa, **Informação & Sociedade**, v. 30, n. 4, p. 1-20, out./dez. 2020. Disponível em: <https://doi.org/10.22478/ufpb.1809-4783.2020v30n4.57788>. Acesso em: 17 jun. 2024.

MIGLIANI, R. O panorama mundial do digital em 2024. WeAreSocial e MeltWater [Resumo e Relatório Completo]. 5 fev. 2024. **Amper**. Disponível em: <https://www.amper.ag/post/panorama-digital-2024-insights-global-report>. Acesso em: 17 jun. 2024.

MOUTINHO, S. O. M.; MARTINS, P. G. M.; ALENCAR, D. F; CONEGLIAN, C. S. Ciência da Informação e Ciência de Dados: convergências interdisciplinares. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 29, p. 1–26, 2024. DOI: 10.5007/1518-2924.2024.e99127. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/99127> . Acesso em: 25 nov. 2024.



OECD. Organisation for Economic Co-operation and Development. Exploring the economics of personal data: a survey of methodologies for measuring monetary value. **OECD Digital Economy Papers**, n. 220, 2013. Disponível em: <http://dx.doi.org/10.1787/5k486qtxldmq-en>. Acesso em: 17 jun 2024.

POZEN, D. E. **The mosaic theory, national security, and the freedom of information act**. Yale LJ, v. 115, p. 628, 2005.

RODRIGUES, F. A. **Estruturas de dados em serviços de redes sociais online: uma abordagem metodológica de análise**. Marília/São Paulo: Oficina Universitária e Cultura Acadêmica, 2024. ISBN: 978-65-5954-469-1. 312 p.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Informação & Informação**, Londrina, v. 21, n. 2, p. 116, 20 dez. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940>. Acesso em: 17 jun. 2024.

SANT'ANA, R. C. G. A algoritmização a partir dos atores do acesso a dados. In: Souza, E. A.; Broens, M. C.; Gonzalez, M. E. Q., **Big data: Implicações Epistemológicas e Éticas**. Campinas: CLE, 2020. Disponível em: <https://www.cle.unicamp.br/ebooks/index.php/publicacoes/catalog/book/8>. Acesso em: 02 set. 2024.

SANT'ANA, R. C. G. A transdução nos processos de mediação e a informação biodisponível. In: SMIT, J. W.; GUIMARAES, J. A. C.; SEGUNDO, J. E. S.; ROLLO, M. F.; SANT'ANA, R. C. G.; HADI, W. M. E. **Humanidades digitais, Big data e pesquisa científica**. São Paulo: Fundação Fernando Henrique Cardoso (FFHC), 2021. Disponível em: <https://fundacaofhc.org.br/iniciativas/publicacoes/humanidades-digitais-big-data-e-pesquisa-cientifica>. Acesso em: 02 set. 2024.

SANTOS, P. L. V. A. C.; SANT'ANA, R. C. G. Dado e Granularidade na perspectiva da Informação e Tecnologia: uma interpretação pela Ciência da Informação. **Ciência da Informação**, v. 42, n. 2, 2013. Disponível em: <https://revista.ibict.br/ciinf/article/view/1382>. Acesso em: 17 jun. 2024.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, v. 1, n. 1, 1996. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/22308>. Acesso em: 17 jun. 2024.

SIMIONATO, A. C.; SANT'ANA, R. C. G.; SANTOS, P. L. V. A. C. Privacidade e os simulacros digitais gerados pelos dados pessoais. **Anais... Encontro Internacional Dados, Tecnologia e Informação**, 2013, Marília/SP.

SILVEIRA, S. A.; AVELINO, R.; SOUZA, J. A privacidade e o mercado de dados pessoais. **Liinc em Revista**, v. 12, n. 2, 30 nov. 2016. Disponível em: <http://revista.ibict.br/liinc/article/view/3719>. Acesso em: 17 jun. 2024.

TAURION, C. **Big data**. Brasport, 2013.



NOTAS

CONTRIBUIÇÃO DE AUTORIA

Concepção e elaboração do manuscrito: A. G. Gomes e R. C. G. Sant'Ana

Coleta de dados: A. G. Gomes

Análise de dados: A. G. Gomes; R. C. G. Sant'Ana e F. A. Rodrigues

Discussão dos resultados: A. G. Gomes; R. C. G. Sant'Ana e F. A. Rodrigues

Revisão e aprovação: A. G. Gomes; R. C. G. Sant'Ana e F. A. Rodrigues

FINANCIAMENTO

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, nº processo 88887.831188/2023-00

CONSENTIMENTO DE USO DE IMAGEM

Não se aplica

APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA

Não se aplica.

CONFLITO DE INTERESSES

Não se aplica.

LICENÇA DE USO

Os autores cedem à **Encontros Bibl** os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a [Licença Creative Commons Attribution](#) (CC BY) 4.0 International. Esta licença permite que **terceiros** remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os **autores** têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

PUBLISHER

Universidade Federal de Santa Catarina. Programa de Pós-graduação em Ciência da Informação. Publicação no [Portal de Periódicos UFSC](#). As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES

Edgar Bisset Alvarez, Genilson Geraldo, Jônatas Edison da Silva, Mayara Madeira Trevisol, Edna Karina da Silva Lira e Luan Soares Silva.

HISTÓRICO

Recebido em: 18-02-2024 – Aprovado em: 04-09-2024 – Publicado em: 14-03-2025

