

Economia de dados: conceitos e desafios para a medição

Governando a economia de dados

Por Marielza Oliveira¹

Introdução

Uma busca na Web pela frase “dados são o novo petróleo” retorna 2,12 bilhões de respostas em 0,58 segundos², demonstrando o quão profundamente ela se incorporou ao nosso léxico. Cunhada em 2006 pelo matemático britânico Clive Humby, é repetida com tanta frequência porque é evocativa, conjurando a imagem de uma misteriosa substância natural, escondida em algum lugar, esperando para ser desenterrada e refinada para que possa revelar o significado de tudo.

A analogia “dados-igual-a-petróleo” transmite sucintamente uma série de noções subjacentes. Para começar, a forma singular, “dado”, é menos usada, pois o dado é percebido como útil apenas em quantidade: enquanto não se pode fazer muito com uma única unidade, ter múltiplas dessas, coletadas ao longo do tempo, entre entidades ou

categorias, permite a aplicação de métodos analíticos para extrair *insights* das relações entre dados individuais. Assim, embutida no conceito de dados, está uma suposição de que “mais é melhor”.

Como na filosofia Ubuntu, que postula que o “si” só pode ser compreendido em relação à comunidade na qual está inserido, um dado é percebido como sem significado até ser extraído e colocado em contraste e comparação com outro dado. Esse ponto é ilustrado com humor por Douglas Adams, em seu romance *O guia do mochileiro das galáxias*, quando um computador, ao ser perguntado sobre o “significado da vida, do universo e de tudo”, oferece “42” como resposta, tornando-a inútil pela falta de contexto. A formação de conjuntos de dados por meio da coleta de quantidades crescentes de dado enriquece o número e a variedade de conexões das quais podem derivar valiosos *insights* (Morrell, 2021). Os dados, portanto, são considerados de natureza relacional por definição e produzem maior valor quando *insights* podem ser obtidos sobre toda a comunidade que eles representam.

Entende-se que os dados pertencem ao seu coletor, o qual decide quais perguntas investigar, quais informações coletar, como rotular e classificar

¹ Diretora de Parcerias e Programas do Setor de Comunicação e Informação da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO). Foi diretora da Representação da UNESCO em Pequim (2015-2020) e previamente atuou como gerente global de resultados do Programa das Nações Unidas para o Desenvolvimento (PNUD), em que também foi gerente de um portfólio de países latino-americanos (2001-2015). É doutora em Administração (1995) e mestre em Finanças (1990) pela Universidade de Illinois, nos Estados Unidos.

² Busca realizada usando a ferramenta Google em 8 de setembro de 2023.



Marielza Oliveira

Diretora de Parcerias e Programas do Setor de Comunicação e Informação da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO).

pontos de dados obtidos e como usar os bancos de dados resultantes. O corolário do ponto de vista de que o valor é criado por acúmulo, organização e comparação de pedacinhos de informações é que o valor derivado dos dados deve ser inteiramente atribuído àqueles que executam essas tarefas. O dado individual não merece nenhuma consideração nesse marco referencial, para além de ser (tal como a tradução desse substantivo latino indica) “uma coisa que é dada” para ser transmutada em dados e, a partir daí, em conhecimento.

Sob essa perspectiva, assume-se que cada dado está disponível gratuitamente e é intrinsecamente sem valor, e o ato de capturar e de aprisionar o maior número possível deles em (um nome apropriado) “células” de alguma tabela é o que gera dados valiosos. A própria palavra “in-formação” alude à importância atribuída a esse processo de organização. Por exemplo, a Wikipédia define economia de dados como o ecossistema digital no qual os chamados dados brutos são “coletados, organizados e trocados por uma rede de empresas, indivíduos e instituições para criar valor econômico” e obter uma recompensa; não é feita nenhuma referência a quaisquer tarefas necessárias para criar o dado e nenhuma consideração é feita em relação à remuneração dos criadores do dado por seu trabalho ou a quaisquer danos que possam sofrer no processo de extração e organização.

Do simples dado ao *Big Data*

Essas visões sobre o dado individual e os conjuntos de dados digitais resultam da forma como os primeiros sistemas informatizados de gestão de banco de dados operavam. A partir do final da década de 1960, as tecnologias da informação permitiram o rápido armazenamento e a recuperação de dados organizados em bancos de dados, e a localização e troca de um dado específico em grandes conjuntos de dados. Esses bancos de dados eram pré-projetados – com especificações claras em termos de tabelas, colunas, índices e outros parâmetros – para conter e processar rapidamente dados bem estruturados, gerados e coletados de acordo com métodos e protocolos de pesquisa expressos e em conformidade com padrões éticos que exigiam neutralidade, ausência de viés e nenhum dano aos provedores de cada dado.

O advento de dispositivos mais rápidos, muito mais potentes e cada vez mais interconectados – computadores de mesa, *tablets*, celulares, eletrodomésticos inteligentes, sensores, *smart TV*, rastreadores de atividade física, câmeras de rua, aparelhos inteligentes vestíveis, carros e outros – catalisou um aumento maciço nas quantidades e variedades de informações disponíveis em formato digital. De acordo com a Statista (2021), a quantidade total de dados criados globalmente atingiu 64,2 *zettabytes* em 2020, e deve crescer para além de 180 *zettabytes* até 2025.

Os recursos de armazenamento também estão crescendo. Por exemplo, a base da capacidade instalada de armazenamento atingiu 6,7 *zettabytes* em 2020 e, desde então, tem aumentado a uma taxa composta de crescimento anual (*compound annual growth rate* [CAGR]) de 19,2%. Entre 2017 e 2021, o valor de mercado cresceu a uma CAGR de 5,5%, chegando a US\$ 101 bilhões

em 2022, e estima-se que crescerá a uma CAGR de 7,5% entre 2022-2032 (Future Marketing Insights, 2022).

A maioria desses dados não é estruturada, consistindo em rastros digitais – texto, áudio, imagem, vídeo – deixados por interações humanas em mídias sociais, comércio eletrônico, economia compartilhada, *websites* de serviços públicos e outras plataformas de Internet. O dado individual das pessoas – trocas com amigos, padrões de pesquisa e navegação, carrinhos de compras *online*, “curtidas” e reclamações nas mídias sociais, fotos de família, geolocalização, rostos e palmas das mãos, batimentos cardíacos, passos e vozes e outros detalhes íntimos e reveladores – tem sido coletado, inclusive, por meio de tecnologias sofisticadas de vigilância, em conjuntos de *Big Data* para processamento em tempo real. Uma proporção significativa de dado, incluindo taxas de cliques, dados de localização específicos de protocolo de Internet (IP) e registros de pesquisa, é considerada como o mero “escape” da atividade *online* (Snyder & Castrounis, 2018), que seria desperdiçado se não fosse coletado e processado para ganho econômico.

Nesse sentido, o advento do *Big Data* alterou as visões anteriores sobre o dado individual e os conjuntos de dados.

MUITO MAIS É AINDA MELHOR

As recentes evoluções na análise preditiva e na aprendizagem de máquina tornaram os grandes bancos de dados capazes não apenas de gerar *insights* para a tomada de decisão, mas também possibilitar a criação de tecnologias inovadoras, como a Inteligência Artificial (IA): os usuários podem nem mesmo saber, no momento da coleta de dados, todas as maneiras que esses dados poderão ser usados no futuro. Empresas e governos estão cada vez mais dependentes de *Big Data* para obter *insights* que permitam respostas rápidas às mudanças nas condições socioeconômicas, políticas, culturais e ambientais, transformando as empresas que mais coletam e processam *Big Data* nas de maior valor. Sete das dez maiores empresas do mundo, medidas pela capitalização de mercado, são atores-chave na economia de dados (Statista, 2023b). Logo, o potencial multiuso de *Big Data* semiestruturado e não-estruturado o torna mais valioso do que os dados estruturados, conforme refletido no crescimento exponencial do valor de mercado global de análise de *Big Data*: atingiu US\$ 272 bilhões em 2022, deve chegar a US\$ 308 bilhões até o final de 2023 e superar US\$ 745 bilhões até 2030 (Fortune Business Insights, 2022).

A mesma lógica sustenta as decisões tomadas por desenvolvedores de IA generativa. Os conjuntos de *Big Data* sobre os quais a IA generativa é construída originam-se da raspagem de dados da Internet, com o objetivo de coletar o máximo possível de seu conteúdo (Washington Post, 2023). Com algumas exceções, até o momento uma tendência óbvia nesse nicho de tecnologia é o aumento no tamanho dos bancos de dados usados para treinar modelos fundacionais, na expectativa de que a adição de parâmetros aumente a versatilidade do modelo (ou seja, permita o ajuste fino para tarefas adicionais), bem como o “surgimento” de novas propriedades e capacidades. O Bidirectional Encoder Representations from Transformers (BERT), um modelo básico lançado em 2018, possuía 340 milhões de parâmetros treinados em um banco de dados

Empresas e governos estão cada vez mais dependentes de *Big Data* para obter *insights* que permitam respostas rápidas às mudanças nas condições socioeconômicas, políticas, culturais e ambientais, transformando as empresas que mais coletam e processam *Big Data* nas de maior valor.

As tecnologias e os incentivos econômicos estão, portanto, em vigor para acelerar a dataficação; por isso, as sociedades devem intensificar os esforços para governar esse processo a fim de obter ganhos sociais.

de 3,3 bilhões de *tokens* totalizando 16 GB, enquanto há rumores de que o GPT-4, lançado em 2023, possui 1,76 trilhão de parâmetros treinados em trilhões de *tokens* contidos em um banco de dados de 45 GB (Amazon, s.d.).

As tecnologias e os incentivos econômicos estão, portanto, em vigor para acelerar a dataficação; por isso, as sociedades devem intensificar os esforços para governar esse processo a fim de obter ganhos sociais.

O DADO TEM UM SIGNIFICADO INTRÍNSECO

A definição mais simples de *Big Data* caracteriza-o como tendo “3 V”: volume (tamanho medido pelo número de registros no banco de dados), velocidade (taxa na qual novas informações são adicionadas) e variedade de fontes e tipos de conteúdo do banco de dados (IT Chronicles, s.d.). Esses enormes bancos de dados são preenchidos com unidades de conteúdo criadas propositalmente para transmitir ideias específicas. Por exemplo, um artigo recente do Washington Post revelou o conteúdo do banco de dados *Colossal Clean Crawled Corpus* (C4) do Google (Goodwin, 2023), usado para treinar modelos fundacionais de IA generativa, como o T5, do Google, e o LLaMA, da Meta. Ele contém raspagens de 15,7 milhões de *websites*, particularmente de textos coletados de páginas dedicadas ao jornalismo, ciências, academia, *marketing*, patentes e outros, com material proveniente de vários locais: Wikipedia, Coursera, grandes jornais (como New York Times e Washington Post), *blogs* pessoais, postagens de mídias sociais (como Reddit e X, anteriormente conhecido como Twitter), livros piratas e romances inéditos, e milhões de outros conteúdos criados com a intenção expressa de entregar informações significativas para seus usuários. Diferentemente de como um dado em bancos de dados estruturados é interpretado, cada dado naqueles bancos de dados não requer associação com outro dado para ter significância intrínseca e transmitir informações. Isso altera o valor que deve ser atribuído ao dado, inclusive por sua criação, e como esse valor deve ser distribuído.

PROTEGER O DADO É TÃO IMPORTANTE QUANTO PROTEGER OS CONJUNTOS DE DADOS

O Departamento de Assuntos Econômicos e Sociais das Nações Unidas (UN DESA) afirma que “A cadeia de valor na economia de dados começa com a coleta de dados pessoais e não pessoais e os torna disponíveis para armazenamento e eventual análise” (UN DESA, 2019, p. 2). O dado pessoal, como nome, número de identificação, geolocalização, endereço de *email* ou de IP, identificação de *cookie*, biometria e qualquer outro dado que possa identificar uma pessoa, por si só ou em combinação com outros dados, é considerado o tipo mais valioso, pois permite que empresas e governos identifiquem e enderecem indivíduos específicos com soluções digitais personalizadas. Pela mesma razão, também é o mais sensível, sendo dotado de proteções especiais de privacidade de dados sob normativas gerais e setoriais (de saúde, por exemplo). Isso porque o dado pessoal está tão intimamente ligado à noção de individualidade que é quase indistinguível dela.

O Artigo 17 do Pacto Internacional sobre os Direitos Civis e Políticos (Organização das Nações Unidas, 1976) consagrou o direito à privacidade no direito internacional. Uma pesquisa da Conferência das Nações Unidas sobre

Comércio e Desenvolvimento (UNCTAD) mostra que 137 dos 194 países promulgaram legislação protegendo a privacidade e os dados pessoais, e outros 17 países têm projetos de legislação em análise (UNCTAD, 2021). Além disso, dada a capilaridade dos fluxos de dados transfronteiriços, órgãos internacionais e regionais, como a Organização para a Cooperação e Desenvolvimento Econômico (OCDE) e a União Europeia, emitiram diretrizes supranacionais e outros marcos normativos para oferecer as proteções de privacidade necessárias aos dados pessoais, como o influente Regulamento Geral sobre a Proteção de Dados (RGPD).

O Fórum Econômico Mundial (FEM) reconhece que as tecnologias da informação representam um desafio para a privacidade, embora as pessoas geralmente entendam por que o rastreamento e o compartilhamento de suas informações são essenciais para materializar os benefícios da conectividade (Schwab, 2016). Esse desafio fez o mercado de anonimização, mascaramento e segurança de dados crescer rapidamente, com um CAGR previsto maior que 13% no período entre 2021 e 2028 (Data Bridge Market Research, 2021). No entanto, com tantos dados coletados sobre os usuários da Internet, as ansiedades continuam aumentando sobre a proteção de dados pessoais contra exploração, violação e exposição. Uma pesquisa recente da PrivacyHawk mostrou que 45% dos usuários de Internet nos Estados Unidos (EUA) estão extremamente preocupados com sua privacidade *online* e 40% estão usando ativamente ferramentas para ocultar seus rastros digitais e proteger sua identidade (Business Wire, 2023).

Essas ansiedades têm fundamento. A Organização Internacional de Padronização (ISO) define anonimização de dados como o “processo pelo qual os dados pessoais são alterados de tal forma que um titular de dados não pode mais ser identificado direta ou indiretamente, seja pelo controlador de dados sozinho ou em colaboração com qualquer outra parte” (ISO, 2017, p. 7). A anonimização permanente é, no entanto, uma tarefa muito difícil. Rocher *et al.* (2019) desenvolveram um método para calcular a probabilidade de reidentificação de indivíduos cujos dados estão contidos em banco de dados considerados anonimizados e descobriram que, geralmente, é bastante alta. Além disso, novos avanços nas tecnologias digitais e nos mecanismos de coleta de dados provavelmente exacerbarão os riscos à privacidade. O trabalho desses autores indica, portanto, que os métodos atuais de anonimização são insuficientes para o desafio de proteger tecnicamente a privacidade, tornando mais urgente o fortalecimento de suas proteções legais.

As normas europeias sobre proteção de dados e da privacidade são fundamentadas nos Direitos Humanos, um marco normativo que enfatiza a proteção da dignidade e do valor das pessoas e se concentra na proteção do dado pessoal. As normas americanas, por outro lado, são fundamentadas no direito comercial e se concentram em regular como as organizações mantêm e usam os dados dos consumidores. Logo, um marco de referência holístico que considere como valorizar e proteger simultaneamente o dado e os dados é necessário para conciliar essas perspectivas.

Os riscos são maiores com o *Big Data*, que, além de auxiliar ou mesmo automatizar decisões, também serve para treinar modelos de IA. Por exemplo, a

No entanto, com tantos dados coletados sobre os usuários da Internet, as ansiedades continuam aumentando sobre a proteção de dados pessoais contra exploração, violação e exposição.

Grandes quantidades de dados não estruturados são mais úteis do que os dados estruturados para encontrar semelhanças entre os membros de uma população, pois as características a serem aplicadas para agrupar indivíduos podem ser determinadas de forma ilimitada e flexível (...).

IA generativa – que engloba Grandes Modelos de Linguagem (*Large Language Models* [LLM]) e Modelos Probabilísticos de Difusão de Redução de Ruído (*Denoising Diffusion Probabilistic Models* [DDPM]) – utiliza métodos de aprendizagem profunda para identificar padrões subjacentes em conjuntos de *Big Data* com base em uma distribuição de probabilidade e, quando solicitado pelos usuários, gera resultados que recriam a distribuição de dados original. Como os modelos de IA generativa são treinados em grandes quantidades de dados, é difícil determinar se o dado incluído na base de dados, ou em seus resultados gerados, viola as leis de direitos autorais ou de privacidade. A Recomendação da UNESCO sobre a Ética da Inteligência Artificial (2021), adotada por todos os seus Estados-Membros, inclui especificamente o direito à privacidade e à proteção de dados pessoais entre os princípios para o desenvolvimento e uso de tecnologias de IA. A recomendação estabelece que os marcos de referência de proteção de dados e os mecanismos de governança devem ser desenvolvidos com uma abordagem multissetorial e sob a proteção dos sistemas judiciais, a fim de permitir que os titulares dos dados exerçam plenamente seus direitos em relação aos dados pessoais durante todo o ciclo de vida da IA.

AS RELAÇÕES ENTRE CADA DADO EM BANCOS DE DADOS DE *BIG DATA* CRIAM MAIOR VALOR, MAS TAMBÉM MAIORES PREOCUPAÇÕES

A análise de *Big Data* é projetada para gerar informações sobre grupos que podem ser aplicadas a todos os indivíduos que compartilham certas características, padrões de comportamento ou preferências. Grandes quantidades de dados não estruturados são mais úteis do que os dados estruturados para encontrar semelhanças entre os membros de uma população, pois as características a serem aplicadas para agrupar indivíduos podem ser determinadas de forma ilimitada e flexível por meio da análise, em vez de serem limitadas pela amostragem da população.

A identidade grupal e as relações entre seus membros revelam informações sobre os indivíduos que eles podem não ter fornecido diretamente, o que, por sua vez, facilita a tomada de decisão sobre eles. Essa capacidade do *Big Data* levou Kitchin (2014) a incluir “identificação indexical” e “natureza relacional” entre suas características definidoras, e Viljoen (2021) a afirmar que “esse aspecto relacional da produção de dados impulsiona grande parte do valor social e dos danos da coleta e do uso de dados em uma economia digital” (p. 573). A destilação de *insights* sobre indivíduos enquanto membros de grupos chegou a tal ponto que um “vazamento” recente revelou 650 mil maneiras pelas quais as pessoas são segmentadas e rotuladas por empresas de publicidade digital (AdTech) (Keegan & Eastwood, 2023).

As primeiras referências ao *Big Data* descreviam a coleta e o processamento de dados comportamentais, transacionais e demográficos das pessoas em plataformas de mídias sociais e de comércio eletrônico, com a intenção de obter *insights* para prever ou influenciar o comportamento daquele usuário específico, mas, mais importante, de outros com características e/ou preferências semelhantes. Serviços valiosos poderiam ser desenvolvidos, como sistemas de recomendação que ajudam os clientes a encontrarem produtos interessantes por meio de ofertas como “clientes que compraram este item também compraram...”,

e algoritmos de detecção de fraude que reduzem o risco de serviços financeiros ao detectar padrões anômalos que se desviam das atividades de usuários típicos. No entanto, o mesmo poderia acontecer com sistemas nocivos, incluindo a distribuição algorítmica de desinformação cujo alvo são pessoas que procuram conteúdo inofensivo nas redes sociais, e sistemas de policiamento preditivo que sinalizam pessoas inocentes como potenciais criminosos com base em características compartilhadas com os condenados pelos crimes em questão. Uma vez que o *Big Data* permite a destilação de *insights* de nível populacional a partir do modo como as pessoas se relacionam umas com as outras, os indivíduos têm interesse em cada dado sobre elas e sobre como esse dado foi obtido. Assim, interesses jurídicos e econômicos supraindividuais também devem ser considerados nos marcos normativos éticos e legais sobre dados.

As preocupações sobre a dataficação de experiências humanas e os danos causados devido a extração, análise e uso de dados estão aumentando. A coleta indiscriminada de conteúdo da Internet traz não apenas informações, mas também conteúdo nocivo, como informações falsas, desinformação e discurso de ódio, para dentro de bancos de dados. As brechas digitais no acesso à infraestrutura e ao conteúdo da Internet geram *Big Data* desequilibrado em termos de representação de vários grupos sociais, o que, por sua vez, leva a algoritmos, serviços e outros resultados tendenciosos e discriminatórios. Para minimizar o potencial de danos, é essencial que os padrões éticos sejam aplicados em toda a cadeia de valor dos dados, desde sua criação até sua coleta, rotulagem, análise, uso, compartilhamento e descarte.

EMBORA O DADO TENHA VALOR ECONÔMICO, QUEM O GERA NÃO GANHA UMA PARCELA JUSTA

O advento da IA generativa evidenciou a distinção entre quem cria dados, e quem apenas coleta dados em conjuntos de *Big Data*. Que esta distinção é problemática tornou-se óbvio à medida que notícias se acumulam sobre a angústia sentida por artistas, cientistas, jornalistas e outros criadores de conteúdo, por terem seus esforços intelectuais coletados em bancos de dados colossais criados com base em qualquer dado disponível publicamente, muitas vezes sem seu consentimento explícito ou mesmo seu conhecimento, com a justificativa de que isso cria valor socioeconômico quando esses conjuntos de dados são usados para inovação. A sub-remuneração pelo trabalho criativo também é galopante; por exemplo, artistas recentemente relataram que lhes solicitaram escanear seus corpos e vozes para gerar réplicas digitais de propriedade de terceiros, que poderão usá-las “para a eternidade” sem nenhum pagamento adicional além dos 15 minutos de trabalho realizado durante a digitalização (Allyn, 2023).

Discussões acaloradas ocorreram em todas as plataformas da Internet, com desenvolvedores de IA afirmando que seria (um uso?) justo um criador doar seu dado “em benefício da humanidade”, ao passo que criadores apontam a hipocrisia dos desenvolvedores de IA por pedirem que eles cedam suas criações “de graça” enquanto esperam lucrar com os modelos de IA resultantes. Uma série de ações coletivas foram movidas sobre o uso de materiais protegidos por direitos autorais encontrados em bancos de dados de treinamento de IA generativa. Os Termos de Uso de notícias digitais, revistas científicas e plataformas de

As brechas digitais no acesso à infraestrutura e ao conteúdo da Internet geram *Big Data* desequilibrado em termos de representação de vários grupos sociais, o que, por sua vez, leva a algoritmos, serviços e outros resultados tendenciosos e discriminatórios.

mídia social vem sendo atualizados para proibir a raspagem, e muitos desses *websites*, que até recentemente eram abertos, instituíram *paywalls* para limitar o acesso a seu conteúdo, reduzindo o acesso à informação, especialmente para pessoas que não podem arcar com esses custos adicionais.

O dado, incluindo o dado pessoal, é cada vez mais considerado um ativo sobre o qual os criadores devem poder exercer direitos de privacidade, propriedade e outros. Arrieta-Ibarra *et al.* (2018) chamam a atenção para o papel negligenciado dos usuários na criação de unidades de dados e propõem que essa tarefa seja reconhecida como trabalho para que as pessoas possam ser adequadamente remuneradas por ela. No entanto, alertam que essa abordagem “pode ir contra os interesses de curto prazo dos principais monopônios de dados que se beneficiaram do fato de os dados serem tratados como ‘gratuitos’” (p. 38).

No centro dos conflitos entre quem gera e quem usa o dado, estão conceitos de valor antagônicos: intrínseco *versus* instrumental, atribuído a conjuntos de dados sem referência ao dado individual, valor de troca *versus* valor de uso. Esses elementos devem ser conciliados.

Do *Big Data* ao valor econômico

O estudo da criação e da apropriação de valor é de alçada da Economia, que examina como uma sociedade utiliza recursos escassos para produção, alocação e consumo de bens e serviços. Um conceito central é a função de produção, que descreve como os insumos econômicos, chamados fatores de produção, podem ser combinados para gerar uma determinada quantidade de produtos específicos. A função clássica de produção contempla ativos tais como equipamentos (“capital”), trabalho realizado por humanos (“trabalho”) e recursos naturais (representados genericamente como “terra”), como seus fatores. Os agentes econômicos que fornecem esses fatores ao processo produtivo são remunerados, respectivamente, por meio de lucros, salários e aluguel, enquanto as empresas empregam a função de produção para determinar a quantidade de produto que devem produzir a cada preço e que combinação de insumos devem usar, considerando seus custos e disponibilidade.

As funções de produção dependem da tecnologia utilizada. As revoluções industriais representam grandes decréscimos na proporção de trabalho necessária para produzir uma determinada quantidade de produtos. As máquinas a vapor aumentaram a força humana via uso desses equipamentos e, assim, iniciaram a primeira Revolução Industrial; eletricidade e correias transportadoras alimentaram a segunda; a automação do trabalho físico trouxe a terceira. Máquinas capazes de aprender com grandes quantidades de dados estenderam a automação ao trabalho cognitivo, desencadeando a 4ª Revolução Industrial.

A analogia dos dados como petróleo consagra os dados, especialmente o *Big Data*, como a nova força que move o motor econômico global para criar prosperidade. Dados aparecem nas funções de produção de duas maneiras: como fator de produção, quando os agentes econômicos usam o conhecimento obtido por meio do processamento e da análise de dados para melhorar tomadas de decisões e ações, ou como um produto, quando os agentes econômicos tratam os dados como o principal ativo valioso e comercializável na economia de dados.

Máquinas capazes de aprender com grandes quantidades de dados estenderam a automação ao trabalho cognitivo, desencadeando a 4ª Revolução Industrial.

Assim, os dados são reconhecidos não apenas como um novo fator de produção (inclusive no planejamento econômico nacional formal, como feito pela China desde 2020), mas também o mais importante na Era da Informação, tão essenciais para a economia digital quanto o petróleo é para a economia industrial. De acordo com o Banco Mundial, durante a última década, a economia digital contribuiu anualmente com mais de 15% do produto interno bruto (PIB) global, além de crescer 2,5 vezes mais rápido do que a economia física (Hayat, 2022). Os ativos físicos não representam mais a maior parte do valor do negócio: bens intangíveis, incluindo dados, respondem por 90% do valor total de mercado das 500 maiores empresas do mundo atualmente. No entanto, as discussões sobre dados em Economia ainda são relativamente incipientes.

Uma das razões é que os dados, ao contrário do petróleo e de outros bens físicos, não são um recurso esgotável, tornando difícil determinar seu valor econômico e ainda mais difícil determinar quem deve ser remunerado por qualquer valor criado a partir deles. Os dados pertencem a uma categoria de bens denominada “não rival”, ou seja, bens que podem ser simultaneamente armazenados, compartilhados e (re)utilizados por múltiplos usuários, sem reduzir sua qualidade ou a quantidade disponível para outros. Em termos econômicos, os dados têm um custo de oportunidade igual a zero, ou seja, os agentes econômicos não precisam escolher entre usos alternativos: todos os dados disponíveis podem ser empregados em todas as oportunidades produtivas de uma só vez. Na verdade, quanto mais usos forem encontrados para os mesmos dados, mais rápido os custos de produção e armazenamento de um conjunto de dados são depreciados, levando a menores custos de produção de cada bem ou serviço para o qual eles são um fator de produção. Além disso, os dados geram efeitos de escala: quanto mais criadores de dados estão na economia, mais dados a partir dos quais um usuário de dados pode obter *insights* estão disponíveis, o que aumenta sua própria produtividade. Entretanto, todos os usuários se beneficiam da mesma forma, levando a um aumento geral na eficiência econômica e na produtividade. Assim, as sociedades usufruem da mais ampla utilização possível dos dados, limitada apenas por sua disponibilidade e qualidade.

A disponibilidade do *Big Data* depende da capacidade de as pessoas criarem e trocarem informações *online* com segurança, o que, por sua vez, depende de vários elementos: o acesso das pessoas a conexões de Internet e dispositivos digitais confiáveis e acessíveis financeiramente, a capacidade de as plataformas de Internet acomodarem criadores com deficiência ou que utilizam idiomas diferentes, e as habilidades de as pessoas buscarem e transmitirem informações em ecossistemas digitais. A criação de *Big Data* de alta qualidade, por outro lado, depende do grau de proteção dos Direitos Humanos (incluindo privacidade, liberdade de expressão e associação, acesso à informação, não discriminação e outros) na Internet. Quando as pessoas se sentem seguras *online*, quando são capazes de identificar e desconsiderar conteúdos nocivos, quando não há barreiras (como jardins murados ou fragmentação) para amplas trocas de informações, e quando elas confiam que ecossistemas de informações salvaguardam seus dados e seus direitos, a qualidade dos dados gerados por essas trocas aumenta.

(...) os dados, ao contrário do petróleo e de outros bens físicos, não são um recurso esgotável, tornando difícil determinar seu valor econômico e ainda mais difícil determinar quem deve ser remunerado por qualquer valor criado a partir deles.

Liberar ainda mais o poder dos dados para maximizar o bem-estar socioeconômico requer modelos econômicos e regimes de governança de dados que empoderem e recompensem equitativamente criadores de unidades e de conjuntos de dados, usuários, e a sociedade, evitando a privatização ou o acúmulo de dados de maneiras socioeconomicamente ineficientes.

Os elementos de criação e troca de informações estão refletidos no marco de referência de Universalidade da Internet da UNESCO, lançado em 2018, que estabeleceu cinco princípios fundamentais para o desenvolvimento da Internet, resumidos no acrônimo “DAAM-X”: a Internet deve ser baseada nos Direitos Humanos, Aberta a todos, Acessível a todos, governada por meio da participação de Múltiplas partes interessadas e manter elementos transversais, como segurança *online* e igualdade de gênero (UNESCO, 2018). Os princípios DAAM foram implementados em mais de 40 países para diagnosticar possíveis melhorias nos ecossistemas digitais que podem levar ao acesso inclusivo à informação para todos e à construção de verdadeiras sociedades do conhecimento (UNESCO, s.d. a).

A disponibilidade depende ainda do grau de abertura dos dados, ou seja, de quem os controla, quem tem o direito legal de usá-los e em que condições. Padrões técnicos para interoperabilidade de dados também são importantes, uma vez que muitos dados disponíveis nunca são utilizados simplesmente porque são mantidos em formatos e estruturas incapazes de serem processadas por sistemas digitais, inclusive em formatos legados desatualizados, como papel, fitas e CD-ROM. Por fim, os dados também devem ser fáceis de encontrar, ou seja, devem ser “detectáveis” por aqueles que desejam coletá-los, analisá-los e utilizá-los. Esses requisitos estão perfeitamente encapsulados nos princípios “FAIR”, que norteiam os dados científicos (GO FAIR, s.d.) – Facilidade de localização, Acessibilidade, Interoperabilidade e Reutilização.

Do valor econômico ao valor social

Pessoas, empresas e governos têm usado dados para reduzir os custos de busca e transação, e obter *insights* sobre os quais podem tomar decisões informadas. Os dados estão tornando as sociedades mais eficientes e as economias mais produtivas, ao mesmo tempo em que melhoram a eficácia das políticas públicas, a prestação de serviços públicos, a transparência e a *accountability*. Liberar ainda mais o poder dos dados para maximizar o bem-estar socioeconômico requer modelos econômicos e regimes de governança de dados que empoderem e recompensem equitativamente criadores de unidades e de conjuntos de dados, usuários, e a sociedade, evitando a privatização ou o acúmulo de dados de maneiras socioeconomicamente ineficientes.

Quando os dados são capturados e privatizados, surgem assimetrias na capacidade de extrair conhecimento e inovar, aumentando a desigualdade. Portanto, uma questão fundamental para as economias modernas é como são definidos os direitos de propriedade, uso ou lucro sobre dados. Em seu tratado sobre O problema do custo social (1960), o economista Ronald Coase propõe que os usuários em potencial de um bem não-rival devem reunir seus recursos, aumentando, assim, a quantidade disponível para todos, distribuindo os custos de manutenção de sua qualidade e facilitando a capacidade de serem encontrados. A UNESCO promove dados abertos (UNESCO, s.d. b), particularmente governamentais e de pesquisa, para que possam desempenhar seu papel na melhoria da qualidade de vida de todos.

PROMOVER UMA CULTURA DE DADOS POR MEIO DO EMPODERAMENTO DE CRIADORES DO DADO, DE PRODUTORES DE CONJUNTOS DE DADOS, E DE SEUS USUÁRIOS

Romper os silos de dados e torná-los mais amplamente acessíveis permite que eles permeiem toda a economia e produzam maiores ganhos socioeconômicos. Uma função nova e importante para os governos é fomentar *pools* de dados, por meio da promoção de plataformas nacionais de governo aberto/dados abertos que abordem desafios nacionais e locais, bem como da co-operação internacional para o *pool* transfronteiriço de recursos de dados para enfrentar desafios globais. Os governos também podem incentivar a participação do setor privado em *pools* de dados setoriais pelos quais os dados proprietários são compartilhados, aumentando a produtividade setorial e as oportunidades de inovação para todos os participantes do mercado, além de nivelar o campo de atuação das *startups*. Por último, mas não menos importante, os governos podem capacitar as pessoas com controle sobre seu dado para que se tornem agentes ativos na economia de dados, promovendo mecanismos de licenciamento, como os *data commons* para causas sociais e os *data trust* de dados para a comercialização de dados. Mecanismos apropriados também devem ser promulgados para coibir o parasitismo de dados, que ocorre quando os usuários de um *data pool* não pagam por ele ou o subvaloram, o que leva a uma quantidade e/ou qualidade menor, prejudicando todos os usuários de dados.

ELIMINAR AS DESIGUALDADES DE DADOS DIGITAIS QUE INIBEM A PARTICIPAÇÃO NA ECONOMIA DE DADOS

Na era digital, aumentar o acesso a oportunidades de participação justa nas cadeias de valor de dados é fundamental para promover o desenvolvimento socioeconômico. Em nível nacional, devem ser envidados maiores esforços para expandir a conectividade significativa, desenvolver habilidades e competências digitais de pessoas, instituições governamentais e empresas, e regular adequadamente a privacidade e os direitos de propriedade de dados individuais e de conjuntos. Em nível internacional, devem ser intensificados os esforços para apoiar os países em desenvolvimento no uso mais eficaz de dados a fim de combater a pobreza e melhorar o bem-estar, incluindo o compartilhamento de boas práticas em mecanismos de regulação para remunerar a criação de dados e proteger os dados pessoais e a privacidade, promulgadas em apenas 48% dos países menos desenvolvidos (Chakraborty, 2022).

INCENTIVAR ATUALIZAÇÕES EM SISTEMAS DE DADOS LEGADOS

Uma quantidade significativa de dados permanece em formatos legados, como papel, filme, fita e outras mídias analógicas. A digitalização de arquivos legados garante o acesso a informações que, de outra forma, poderiam ser perdidas devido a deterioração, desastres naturais, conflitos e outras causas. A melhoria da infraestrutura técnica aumenta a capacidade de o agente econômico criar, coletar, armazenar e analisar conjuntos de dados de rápido crescimento, de forma segura e econômica. O desbloqueio do valor dos dados também exige a atualização de mecanismos legais para proteger os direitos sobre os dados, facilitar as transações de dados e fornecer mecanismos para a resolução de disputas.

Na era digital, aumentar o acesso a oportunidades de participação justa nas cadeias de valor de dados é fundamental para promover o desenvolvimento socioeconômico.

(...) os economistas precisam de melhores ferramentas para medir e modelar o impacto da economia de dados na criação de valor, incluindo fatores de produção do dado individual, insumos intermediários e produtos finais.

PROMOVER A SUSTENTABILIDADE AMBIENTAL DAS CADEIAS DE VALOR DOS DADOS DIGITAIS

Assim como a perfuração de petróleo, a “mineração” e o processamento de dados causam danos ao ambiente natural e à exploração de recursos naturais finitos. Até agora, a Economia tem tratado esses impactos negativos como externalidades, mas novas ideias sobre como contabilizar os impactos ambientais surgem. Com os *data centers* se tornando rapidamente alguns dos consumidores mais vorazes de recursos naturais – água, energia, minerais raros necessários para a produção de *chips*, bem como dispositivos digitais antigos entupindo aterros sanitários e vazando substâncias tóxicas –, é imperativo que as sociedades considerem os impactos ambientais dos sistemas de dados, inclusive por meio das medidas regulatórias necessárias.

AMPLIAR OS PADRÕES E AS ABORDAGENS CONTÁBEIS NACIONAIS

Os modelos econômicos concentram-se nos domicílios e nos indivíduos como consumidores ou fornecedores de mão de obra e não os reconhecem como produtores que usam seu próprio trabalho e capital. As tecnologias digitais têm tornado mais urgente o exame das maneiras pelas quais o uso da mão de obra está mudando devido à automação, com muitas tarefas (executadas, por exemplo, por meio de caixas de supermercado de autoatendimento, caixas eletrônicos e compras pela Internet) redirecionadas das empresas para os indivíduos, bem como dos governos para as pessoas (por meio de serviços públicos digitais) e, inversamente, das pessoas para as empresas (por exemplo, por meio da terceirização da produção doméstica de alimentos para plataformas digitais, como o iFood). Nesse sentido, os economistas precisam de melhores ferramentas para medir e modelar o impacto da economia de dados na criação de valor, incluindo fatores de produção do dado individual, insumos intermediários e produtos finais. Excelentes pesquisas sobre domicílios, como a TIC Domicílios, do Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), departamento do Núcleo de Informação e Coordenação do Ponto BR (NIC.br), podem ser uma fonte-modelo sobre os equipamentos (dispositivos e infraestrutura) usados e a mão de obra empregada na criação do dado (NIC.br, s.d.). Estimativas globais, como as realizadas pelo Statista, também podem contribuir para a análise: por exemplo, os números da pesquisa mais recente mostram que, no quarto trimestre de 2022, o tempo médio gasto na Internet por pessoa foi de 395 minutos (seis horas e 35 minutos) por dia (Oberlo, 2023), ao passo que as redes sociais, sozinhas, contabilizavam 151 minutos por dia (Statista, 2023a). Isso também poderia melhorar as estatísticas oficiais de trabalho, que ignoram as contribuições da economia doméstica à economia de mercado, e sub-representam o *gig work* e outras tarefas realizadas na economia de dados.

Referências

- Allyn, B. (2023, Agosto). Movie extras worry they'll be replaced by AI. Hollywood is already doing body scans. *NPR*. <https://www.npr.org/2023/08/02/1190605685/movie-extras-worry-theyll-be-replaced-by-ai-hollywood-is-already-doing-body-scan>
- Amazon. (s.d.). *What are Foundation Models?* <https://aws.amazon.com/pt/what-is/foundation-models/>
- Arrieta-Ibarra, I., Goff, L., Jiménez-Hernández, D., Lanier, J., & Glen Weyl, E. (2018). (2018). Should we treat data as labor? Moving beyond “free”. *American Economic Association Papers and Proceedings*, 108, 38-42.
- Business Wire. (2023). *PrivacyHawk releases 2023 personal data, privacy and AI report highlighting consumer alarm and sentiment about privacy and AI*. <https://www.businesswire.com/news/home/20230822244117/en/PrivacyHawk-releases-2023-personal-data-privacy-and-AI-report-highlighting-consumer-alarm-and-sentiment-about-privacy-and-AI>
- Chakraborty, T. (2022). Is data the new oil? *Yubi*. <https://www.go-yubi.com/blog/data-new-oil/>
- Conferência das Nações Unidas sobre Comércio e Desenvolvimento. (2021). *Protection and privacy legislation worldwide*. <https://unctad.org/page/data-protection-and-privacy-legislation-worldwide>
- Data Bridge Market Research. (2021). *Global data masking market – Industry trends and forecast to 2028*. <https://www.databridgemarketresearch.com/reports/global-data-masking-market>
- Departamento das Nações Unidas para Assuntos Econômicos e Sociais. (2019). Data economy: Radical transformation or dystopia? *Frontier Technology Quarterly*. https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/FTQ_1_Jan_2019.pdf
- Fortune Business Insights. (2022). *Market research report*. <https://www.fortunebusinessinsights.com/big-data-analytics-market-106179>
- Future Marketing Insights. (2022). *Data Center Market report*. <https://www.futuremarketinsights.com/reports/data-center-market>
- GO FAIR. (s.d.). *FAIR Principles*. <https://www.go-fair.org/fair-principles/>
- Goodwin, D. (2023). Search the 15.7 million websites in Google's C4 dataset. *Search Engine Land*. <https://searchengineland.com/search-websites-google-c4-dataset-395820>
- Hayat, Z. (2022, August). Digital trust: How to unleash the trillion-dollar opportunity for our global economy. *World Economic Forum*. <https://www.weforum.org/agenda/2022/08/digital-trust-how-to-unleash-the-trillion-dollar-opportunity-for-our-global-economy>
- IT Chronicles. (s.d.). *What is Big Data*. <https://itchronicles.com/what-is-big-data/>
- Keegan, J., & Eastwood, J. (2023, June). From “heavy purchasers” of pregnancy tests to the depression-prone: We found 650,000 ways advertisers label you. *The Markup*. <https://themarkup.org/privacy/2023/06/08/from-heavy-purchasers-of-pregnancy-tests-to-the-depression-prone-we-found-650000-ways-advertisers-label-you>
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1). <https://doi.org/10.1177/2053951714528481>
- Morrell, J. (2021). Does more data equal better analytics? *Datameer*. <https://www.datameer.com/blog/does-more-data-equal-better-analytics/>
- Núcleo de Informação e Coordenação do Ponto BR. (s.d.). *Pesquisa sobre o uso das Tecnologias de Informação e Comunicação nos domicílios brasileiros*. <https://cetic.br/pt/pesquisa/domicilios/>
- Oberlo. (2023). *How much time do people spend online?* <https://www.oberlo.com/statistics/how-much-time-does-the-average-person-spend-on-the-internet>

- Organização das Nações Unidas. (1976). *Pacto Internacional dos Direitos Civis e Políticos*.
- Organização das Nações Unidas para a Educação, a Ciência e a Cultura. (s.d. a). *Internet Universality Indicators*. <https://www.unesco.org/en/internet-universality-indicators>
- Organização das Nações Unidas para a Educação, a Ciência e a Cultura. (s.d. b). *Open Data*. <https://www.unesco.org/en/open-solutions/open-data>
- Organização das Nações Unidas para a Educação, a Ciência e a Cultura. (2018). *UNESCO's Internet Universality Indicators: A framework for assessing Internet development – draft for the consideration of the Intergovernmental Council of IPDC*. <https://unesdoc.unesco.org/ark:/48223/pf0000265830>
- Organização Internacional de Padronização. (2017). *ISO 25237:2017 Health informatics: Pseudonymization*. <https://www.iso.org/standard/63553.html>
- Rocher, L., Hendrickx, J. M., & Montjoye, Y. A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, *10*, 3069. <https://doi.org/10.1038/s41467-019-10933-3>
- Schwab, K. (2016, January). The Fourth Industrial Revolution: What it means, how to respond. *World Economic Forum*. <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>
- Snyder, S., & Castrounis, A. (2018). How to turn 'data exhaust' into a competitive edge. *Knowledge at Wharton*. <https://knowledge.wharton.upenn.edu/article/turn-iot-data-exhaust-next-competitive-advantage/>
- Statista. (2021). *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025*. <https://www.statista.com/statistics/871513/worldwide-data-created/>
- Statista. (2023a). *Daily time spent on social networking by Internet users worldwide from 2012 to 2023*. <https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/>
- Statista. (2023b). *The 100 largest companies in the world by market capitalization in 2023*. <https://www.statista.com/statistics/263264/top-companies-in-the-world-by-market-capitalization/>
- Viljoen, S. (2021). A relational theory of data governance. *Yale Law*. <https://heinonline.org/HOL/LandingPage?handle=hein.journals/ylr131&div=12&id=&page=>
- Washington Post. (2023). *Inside the secret list of websites that make AI like ChatGPT*. <https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/>

Entrevista I

A economia de dados no Brasil

Marcos Dantas é professor titular da Escola de Comunicação da Universidade Federal do Rio de Janeiro (UFRJ) e doutor em Engenharia de Produção pela mesma instituição. Integra o Conselho de Administração do Núcleo de Informação e Coordenação do Ponto BR (NIC.br) e o Comitê Gestor da Internet no Brasil (CGI.br), e é autor (com D. Moura, L. Ormay e G. Raulino) de *O valor da informação: de como o capital se apropria do trabalho social na era do espetáculo e da Internet* (Boitempo, 2022). Nesta entrevista, ele discute a importância do tema “economia de dados” no contexto atual e os possíveis caminhos para a distribuição do valor obtido com os dados entre a sociedade que os produz; por fim, aborda os desafios para a mensuração da economia de dados no Brasil.

Panorama Setorial da Internet (P.S.I.)_ Qual a importância do tema “economia de dados” no contexto de transformação digital em que vivemos atualmente?

Marcos Dantas (M.D.)_ Uma edição da revista *The Economist*, de 2017, dizia que os “dados são o petróleo do século XXI”. Sabemos que a economia do século XX foi movida a petróleo. A gente não faz nada sem petróleo: não apenas como fonte de energia, mas contido em quase todo tipo de utensílios que usamos no nosso dia a dia. Além de economicamente determinante – por isso mesmo –, o petróleo também foi e ainda é motivo de guerras, golpes de Estado, até assassinatos de lideranças políticas, ou seja, petróleo também significa poder e luta pelo poder. Nesse sentido de recurso econômico básico e de instrumento de poder político e geopolítico, podemos admitir ser válida a analogia feita pela revista. No século XXI, em torno dos dados se moverá a economia e, também, a luta política. Basta ver o tamanho que já atingiram algumas corporações, como Alphabet ou Amazon, para se confirmar essa dimensão econômica. Basta atentar para as denúncias de Edward Snowden e Julian Assange, ou para as atividades eleitorais da Cambridge Analytica, para termos um deslumbre – sublinhe-se “um deslumbre” – de como o acesso e a manipulação de dados já podem estar substituindo, de modo muito mais eficaz, com a vantagem de parecer indolor, antigas práticas de golpes violentos e assassinatos, comuns nos tempos da hegemonia do petróleo. No entanto, apesar dessa importância decisiva, pouco ainda sabemos sobre essa nova economia. Até mesmo a definição de “dados” está em disputa. Essa ignorância contribui para que também não tenhamos ainda alguma sólida política pública para fomentar, além de regular, a economia de dados, diferentemente da que temos, há muitas décadas, no Brasil e no mundo, em relação à economia do petróleo e dos recursos energéticos, em geral.



Foto: Arquivo pessoal

Marcos Dantas

Professor Titular da Universidade Federal do Rio de Janeiro (UFRJ) e conselheiro do Comitê Gestor da Internet no Brasil (CGI.br).

“Se chegamos a uma etapa histórica que tornou os dados sociais digitalizados um recurso econômico, o primeiro ponto a abordarmos seria como fazer essa riqueza gerar benefícios para toda a sociedade.”

P.S.I._ *Considerando que a economia de dados tem como parte importante de sua estrutura as plataformas digitais, quais os caminhos possíveis para se repensar a monetização dos dados produzidos pelos usuários?*

M.D._ As plataformas sociodigitais inventaram um modelo de negócios baseado na oferta para os anunciantes, em leilões, dos dados de seus usuários. Os usuários, ingenuamente, mantêm-se em intensa atividade, fornecendo esses dados de graça, para as plataformas lucrarem com eles. Os dados são registros, em formato eletroeletrônico, das atividades sociais dos usuários. Assim como o conteúdo de um livro é o registro, em tinta e papel, da atividade de um escritor, uma atividade sempre social (nenhum escritor é um Robinson Crusoe), os dados são produzidos por bilhões de “escritores”, vamos dizer assim, cujos “leitores” são os anunciantes. Por ser produto de uma atividade social, pois sempre é produzido nas interações interpessoais ou nas necessidades sociais dos indivíduos, os dados são, antes de mais nada, um recurso social – como o petróleo ou a água. Se chegamos a uma etapa histórica que tornou os dados sociais digitalizados um recurso econômico, o primeiro ponto a abordarmos seria como fazer essa riqueza gerar benefícios para toda a sociedade. Por enquanto, só beneficia um pequeno grupo de grandes acionistas daquelas plataformas...

Porém, em seu processo de produção ou, podemos dizer com todas as letras, de trabalho individual e social, os reais produtores trabalham de graça. Uma parte da receita remunera o trabalho de cientistas, engenheiros, técnicos e empregados pelas plataformas, que desenvolvem os sistemas e os algoritmos de mineração dos dados. Em alguns modelos de negócios, parte da receita pode também remunerar os produtores mais bem sucedidos de dados, os tais “influenciadores”. Mas eles são bem-sucedidos exatamente porque conseguem atrair o trabalho não pago de milhões, até bilhões, de fornecedores de dados. Eles ajudam (e como ajudam!) as plataformas a minerar os dados socialmente disponíveis nos corpos e nos atos dos indivíduos.

Numa economia capitalista, as plataformas sociodigitais cumprem um papel essencial: anular o espaço pelo tempo e, assim, multiplicar exponencialmente a rotação do capital. Quanto menor o tempo de realização de um investimento (e quanto mais vezes consegue girar um mesmo capital), mais cresce o lucro de um produtor qualquer. Por isso, uma vez que elas tenham surgido, a economia capitalista não mais aceitará viver sem elas. E nós teremos de saber conviver com isso. O problema é *como*? Como fazer com que aquela dimensão praticamente incalculável de valor contida nos dados sociais por elas apropriados possa, de algum modo, ser também compartilhada com a sociedade que os produz? Assim estou entendendo a sua pergunta.

A resposta mais óbvia talvez seja remunerando os produtores de dados. A Meta informa, em seus relatórios, que sua receita média por usuário é de US\$ 40. O usuário não paga nada para usar as plataformas da Meta, mas gera, para ela, US\$ 40. Ou seja, o tempo que esse usuário empregou a serviço da Meta, dedilhando uma tela de celular ou o *mouse* do computador,

vale, para ela, US\$ 40. Por que não remunerar o usuário por esse tempo? Por que não dividir com ele essa receita? Por que só Jeff Bezos, Elon Musk, Black Rock Inc., Vanguard Group Inc., State Street e outros assemelhados podem amealhar os lucros de toda essa riqueza?

P.S.I._ Em sua opinião, quais os desafios para a mensuração da economia de dados no Brasil?

M.D._ Os desafios são comuns ao Brasil e ao mundo. Talvez no Brasil, como nos demais países da periferia capitalista, aos desafios comuns somam-se os específicos da condição subordinada e das nossas grandes desigualdades sociais. Mensurar dados, em princípio, não deveria ser difícil, já que estamos falando em grandezas físicas. No momento em que escrevo, sei que tenho 245 *gigabytes* de dados registrados no disco rígido do meu computador. O problema é que a sociedade desconhece realmente o volume dos dados que circulam e são registrados nos servidores das grandes plataformas. Podemos até ter uma noção geral do volume total de tráfego. O NIC.br, por exemplo, fornece para nós essa estatística relativa ao Brasil. Mas, os poderes públicos não têm instrumentos para controlar o fluxo de dados, como pode, por exemplo, fiscalizar o fluxo de petróleo que sai ou entra no país. Banco Central e Receita Federal podem identificar operações financeiras suspeitas porque os bancos são obrigados a informar todas as operações financeiras realizadas às autoridades públicas. Não temos, porém, nem no Brasil nem no mundo, autoridades públicas com poderes e meios de controlar o fluxo de dados, um recurso porém, como vimos, que está merecendo ser tratado como se trata o petróleo. À medida que pudermos melhor compreender essa economia, podemos esperar que a sociedade brasileira reconhecerá que “os dados são nossos”, assim como, décadas atrás, a geração de nossos pais (ou avós) entendeu que “o petróleo é nosso!”.

“Mensurar dados, em princípio, não deveria ser difícil, já que estamos falando em grandezas físicas. (...) Mas, os poderes públicos não têm instrumentos para controlar o fluxo de dados, como pode, por exemplo, fiscalizar o fluxo de petróleo que sai ou entra no país.”

Artigo II

Opções políticas para a economia de dados: uma revisão da literatura³

Por Ramy El-Dardiry⁴, Milena Dinkova⁵ e Bastiaan Overvest⁶

Introdução

Neste artigo, discutimos a literatura sobre economia de dados. A crescente importância dos dados em nossa economia e sociedade tem motivado mais pesquisas sobre esses temas nos últimos anos. Os dados e a decorrente digitalização de nossa sociedade apresentam enormes oportunidades e desafios. Por um lado, a digitalização pode introduzir e sustentar um novo período de crescimento econômico e ajudar a superar os desafios sociais, permitindo educação personalizada ou medicina preventiva, por exemplo. Por outro lado, há incertezas sobre o futuro da privacidade e de nossa democracia, e preocupações com o poder de um punhado de empresas de tecnologia. Desse modo, novas pesquisas permitem compreender melhor a forma de pensar sobre estas oportunidades e desafios.

Dado o amplo e profundo impacto da digitalização, talvez não seja surpreendente que a pesquisa sobre dados envolva muitas disciplinas. Recentemente, economistas começaram a compreender melhor de que maneira os dados funcionam como um fator de produção; juristas têm aproveitado conceitos da economia para estudar externalidades associadas aos dados e à privacidade; filósofos, sociólogos e cientistas políticos têm se preocupado com o novo equilíbrio de poder que emerge da era digital. Enquanto isso, os cientistas da computação estão constantemente inventando novas maneiras de proteger melhor a privacidade ou viabilizar a troca de dados.

Neste artigo, definimos e caracterizamos o que se entende por dados. Também revisamos as propriedades econômicas dos dados, como não rivalidade e baixos custos de replicação, e as estimativas do valor dos dados em nossa economia.

Por um lado, a digitalização pode introduzir e sustentar um novo período de crescimento econômico e ajudar a superar os desafios sociais (...). Por outro lado, há incertezas sobre o futuro da privacidade e de nossa democracia, e preocupações com o poder de um punhado de empresas tecnológicas.

³ Versão editada da obra homônima publicada pelo Gabinete de Análise de Política Econômica dos Países Baixos. Disponível em: <https://www.cpb.nl/en/brave-new-data-policy-pathways-for-the-data-economy-in-an-imperfect-world#docid-160570>

⁴ Gabinete de Análise de Política Econômica dos Países Baixos.

⁵ Gabinete de Análise de Política Econômica dos Países Baixos.

⁶ Gabinete de Análise de Política Econômica dos Países Baixos.

Características dos dados

Os dados são apresentados em muitas formas diferentes e usados de várias maneiras. Entender essas diferenças é importante para conceber políticas que equilibrem oportunidades e desafios. Por exemplo, a utilização de declarações de rendimentos anonimizadas para um trabalho acadêmico sobre o desempenho financeiro de pequenas e médias empresas levanta problemas diferentes da utilização do perfil de uma pessoa nas redes sociais para direcionar publicidade. Ao mesmo tempo, algumas propriedades econômicas dos dados (digitais) são independentes do tipo de dados. Neste artigo, nosso primeiro objetivo é compreender melhor os dados, definindo-os, categorizando suas diferenças e identificando seus denominadores comuns. Em seguida, estudamos a economia de dados com mais detalhes e discutimos o valor dos dados.

DEFINIÇÕES E CATEGORIZAÇÃO

Carrière-Swallow e Haksar (2019) definem “dados” como uma “representação factual de uma característica, ação ou ocorrência natural” (p. 7); além disso, fazem uma distinção entre dados qualitativos e quantitativos, e a forma como são armazenados (digital versus analógico). Hilbert e López (2011) mostram como os dados têm se tornado cada vez mais digitalizados nas últimas décadas. Atualmente, os dados são predominantemente armazenados em formato digital⁷.

Os dados diferem das ideias: ambos são formas de informação, mas servem a propósitos diferentes. Segundo Jones e Tonetti (2020), “uma ideia é uma função de produção, enquanto os dados são um fator de produção” (p. 2821). Em termos concretos, isso significa que as ideias são pedaços de informação que fornecem instruções sobre como criar resultados (*output*) a partir de um determinado conjunto de entradas (*input*) (Romer, 1990). Os dados, por outro lado, são usados no processo de produção, seja para criar produtos ou serviços ou novas ideias.

Vários esquemas de classificação de dados surgiram na literatura (Wdowin & Diepeveen, 2020). Crémer *et al.* (2019) fazem uma distinção entre dados pessoais e não pessoais, classificando-os como voluntários, observados ou inferidos com base no canal pelo qual os dados foram adquiridos. Além disso, propõem a distinção entre quatro categorias de casos de uso: as aplicações e análises podem usar dados de nível individual, dados de nível individual agrupados, dados de nível agregado ou dados contextuais. Os dados de nível individual referem-se a dados de um usuário ou máquina específico. Quando esses dados são combinados para criar recomendações de filmes ou músicas, por exemplo, utilizam o termo “dados de nível individual agrupados”. Sem informações adicionais, não é possível vincular dados agregados de volta ao nível individual. Os exemplos incluem tabelas de frequência que mostram a distribuição dos níveis de competências digitais de um grupo populacional ou demonstrativos de lucros e prejuízos. Os dados contextuais não são derivados dos dados de nível individual. Exemplos típicos são dados de satélite, dados cartográficos ou dados de terremotos.

Os dados diferem das ideias: ambos são formas de informação, mas servem a propósitos diferentes. (...) as ideias são pedaços de informação que fornecem instruções sobre como criar resultados (...). Os dados, por outro lado, são usados no processo de produção, seja para criar produtos ou serviços ou novas ideias.

⁷ Para algumas formas antigas e fascinantes de armazenar dados, ver, por exemplo, artigo da BBC sobre os primeiros contadores do mundo (Harford, 2017).

Uma das características mais distintas dos dados é a não rivalidade. Um bem econômico é não-rival quando pode ser usado por vários consumidores ou empresas ao mesmo tempo, sem diminuir sua quantidade ou qualidade.

A Statistics Canada (2019) sugere que os dados sejam organizados de acordo com o que eles são ou o que representam, por exemplo, dados meteorológicos, dados esportivos ou dados econômicos. Em um relatório sobre transferências internacionais de dados, a Junta Comercial Nacional da Suécia (2015) classifica os dados com base na forma como eles são usados no processo de produção das empresas. Os exemplos incluem dados sobre emprego, dados sobre a qualidade e dados sobre clientes.

PROPRIEDADES ECONÔMICAS

Nesta seção, estudamos as propriedades econômicas dos dados. Primeiro, focamos na não rivalidade e exclusibilidade parcial. Em segundo lugar, discutimos o impacto dos dados sobre os custos econômicos e seus rendimentos marginais.

NÃO RIVALIDADE

Uma das características mais distintas dos dados é a não rivalidade. Um bem econômico é não-rival quando pode ser usado por vários consumidores ou empresas ao mesmo tempo, sem diminuir sua quantidade ou qualidade. Jones e Tonetti (2020) usam uma analogia ilustrativa com bens rivais para explicar o que significa o fato de que, no “nível tecnológico, os dados são infinitamente utilizáveis” (p. 2819). Em razão da rivalidade, os trabalhadores normalmente precisam de sua própria mesa e computador, e todo depósito depende de sua própria e exclusiva coleção de empilhadeiras. No entanto, se assumirmos que esse capital é não-rival, todos os trabalhadores poderiam usar todas as mesas e computadores ao mesmo tempo, e todos os depósitos poderiam usar qualquer empilhadeira do setor. É o caso dos dados. Devido à não rivalidade, todos os dados poderiam teoricamente ser usados por todas as empresas ao mesmo tempo, o que implica que os ganhos econômicos permaneceriam inexplorados enquanto essa não rivalidade não for explorada. Carrière-Swallow e Haksar (2019) observam que políticas e interesses privados determinam se os dados serão não-rivais na prática.

Goldfarb e Tucker (2019) generalizam a não rivalidade de dados para produtos e serviços, comparando bens feitos de átomos e bens feitos de *bits*. Ao contrário dos bens feitos de átomos, os *bits* não são rivais, porque os custos de replicação de informações digitais são quase zero – você pode copiar e colar o código do *software*, mas não uma Ferrari.

EXCLUSIBILIDADE PARCIAL

Alguns tipos de dados são excludentes, ou seja, negar o acesso a outros não é proibitivamente dispendioso. Quando os coletores de dados excluem outros, os dados assumem as características de um bem de clube (Buchanan, 1965). Quando outras pessoas não podem ser impedidas de acessar dados, os dados não são excludentes e podem ser considerados um bem público.

Coyle *et al.* (2020) fornecem um breve panorama da exclusibilidade de diferentes tipos de dados. Por exemplo, dados administrativos (como declarações fiscais ou registros de pacientes) ou planejados (como cronogramas de trabalho

ou orçamentos) são tipos de dados dos quais outros podem ser facilmente excluídos. Em contrapartida, os dados ambientais, como os pluviométricos ou os geoespaciais, são acessíveis a qualquer pessoa, uma vez que todos podem recolher seus próprios dados de fenômenos observáveis publicamente – embora os custos privados da medição possam ser muito altos para realmente fazê-lo. Uma forma comum de tornar os dados excludentes é colocá-los atrás de um acesso pago (*paywall*) – frequentemente vinculado ao registo de uma conta. Pense em artigos de jornal ou bancos de dados para pesquisadores. O armazenamento *offline* é provavelmente a maneira mais fácil de limitar o acesso – só a entrada física no dispositivo ou no espaço onde os dados são armazenados pode remover o bloqueio.

Os coletores de dados e os processadores de dados enfrentam diferentes incentivos ao decidir o nível de acesso aos dados. Podem, por exemplo, restringir o acesso para garantir sua vantagem competitiva e manter sua posição atual no mercado (Carrière-Swallow & Haksar, 2019). A legislação de privacidade pode ser outra razão para uma organização excluir o acesso de outras pessoas.

O IMPACTO DOS DADOS NOS CUSTOS ECONÔMICOS

Goldfarb e Tucker (2019) descrevem como a digitalização reduz cinco custos econômicos (custos de pesquisa, replicação, transporte, rastreamento e verificação), os quais se relacionam às propriedades dos dados digitalizados.

Primeiro, a digitalização diminui os custos de pesquisa. Os mecanismos de busca, por exemplo, tornaram muito mais fácil encontrar informações relevantes, sejam elas sobre produtos, conhecimento ou dados em si. Segundo, os custos de replicação de produtos digitais são próximos de zero: em outras palavras, os custos marginais são negligenciáveis. Além disso, como vimos, a reprodução de dados não impacta outros dados devido à sua natureza não-rival. Embora os custos marginais quase desapareçam, o lançamento de produtos digitais bem-sucedidos geralmente requer um investimento inicial significativo – por exemplo, para estabelecer uma rede suficientemente grande ou construir uma infraestrutura de dados sólida. Terceiro, os dados estão associados a custos de transporte quase nulos e podem ser transferidos para todo o mundo sem muito esforço; conseqüentemente, os modelos de negócio digitais tornaram-se cada vez mais globais, enquanto as empresas podem escalonar a um ritmo mais rápido. Quarto, os custos de rastreamento são reduzidos: os dados digitais facilitam o controle de transações, pessoas e empresas. A digitalização levou, portanto, a níveis crescentes de personalização. Exemplos disso são a discriminação de preços e anúncios personalizados, os quais têm o potencial de facilitar a correspondência entre oferta e demanda. Quinto, os custos de rastreamento mais baixos permitiram uma redução nos custos de verificação. Os produtos e serviços digitais facilitaram a verificação de identidades e a criação de sistemas de reputação. Plataformas digitais, como Uber e Airbnb, aproveitaram extensivamente a redução na verificação para construir confiança em seus mercados bilaterais.

Os produtos e serviços digitais facilitaram a verificação de identidades e a criação de sistemas de reputação. Plataformas digitais, como Uber e Airbnb, aproveitaram extensivamente a redução na verificação para construir confiança em seus mercados bilaterais.

(...) na economia de dados, em que os algoritmos de aprendizagem de máquina desempenham um papel cada vez mais importante, os valores marginais dos dados podem estar aumentando. (...) Normalmente, são necessários mais dados quanto mais complexo for o problema.

RETORNOS CRESCENTES E DECRESCENTES

Em seu livro *Radical markets* (2018), Eric Posner e Glen Weyl discutem o valor marginal dos dados em profundidade. O valor marginal de um ponto de dados extra pode ser decrescente ou crescente em relação ao número de pontos de dados coletados, dependendo do contexto.

Para entender como isso funciona, primeiro considere um problema de estatística padrão. Digamos, por exemplo, que você esteja interessado em determinar a poupança média das famílias. Ainda que a incerteza sobre a poupança média das famílias diminua com o número de pontos de dados coletados, o declínio marginal torna-se cada vez menor à medida que mais pontos de dados são adicionados. Assim, os dados perdem seu valor devido ao volume e à variedade. Além disso, há sempre algum grau de incerteza que é suficiente para a aplicação em questão. A coleta de mais dados quando esse nível de incerteza é atingido é ineficiente.

Posner e Weyl (2018) explicam como na economia de dados, em que os algoritmos de aprendizagem de máquina desempenham um papel cada vez mais importante, os valores marginais dos dados podem estar aumentando. A razão subjacente para esse aumento de retornos dos dados é que diferentes algoritmos exigem diferentes quantidades de dados. Normalmente, são necessários mais dados quanto mais complexo for o problema. Para um único problema de aprendizagem, os dados novamente exibem um valor de retorno decrescente, mas a coleta de mais dados agora pode permitir a solução de novos problemas, causando um salto no valor dos dados coletados. Se os dados têm um valor de retorno crescente ou decrescente é, então, determinado pelo valor dos diferentes problemas. Quando a maior parte do valor reside no problema mais complexo, é provável que os dados tenham um valor de retorno crescente. Em contrapartida, quando a maioria do valor reside nos problemas mais simples, os dados provavelmente terão um valor de retorno decrescente.

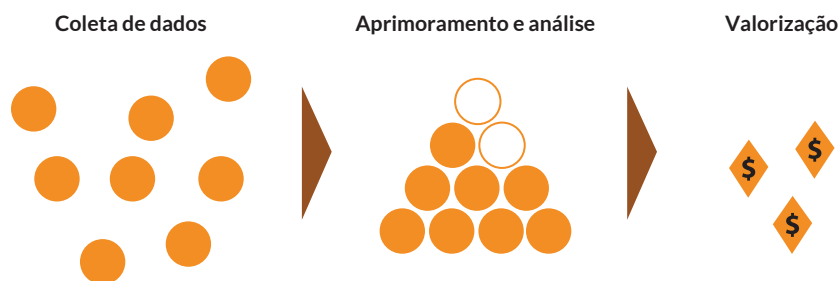
O VALOR DOS DADOS NA ECONOMIA

EXTRAINDO VALOR DOS DADOS: A CADEIA DE VALOR DOS DADOS

Os dados são um *input* nos processos de produção. A cadeia de valor dos dados descreve como esses contribuem para a produção. Em nosso artigo, dividimos a cadeia de valor dos dados em três (Figura 1). Em primeiro lugar, os dados precisam ser coletados e armazenados. Em segundo lugar, são analisados e combinados para criar conhecimentos (*insights*). Terceiro, os conhecimentos traduzem-se em produtos e serviços.

As empresas e os institutos da economia de dados concentram-se em uma parte da cadeia ou controlam toda a cadeia de valor de seus negócios. Serviços em nuvem e consultores de *Big Data* são exemplos de empresas especializadas em oferecer produtos para uma determinada parte da cadeia de valor. As atividades de empresas de tecnologia, por exemplo, Alphabet, Amazon e Apple, abrangem toda a cadeia de valor. Na literatura, aparecem diferentes versões da cadeia de valor, as quais têm origem no número de cadeias ou numa terminologia ligeiramente diferente.

Figura 1 – CADEIA DE VALOR DOS DADOS



Fonte: Elaboração própria.

Muitas vezes, os agentes econômicos que desempenham um papel na cadeia de valor são referidos como titulares de dados, coletores de dados e processadores de dados (Carrière-Swallow & Haksar, 2019). No caso de dados pessoais, a pessoa, cujos detalhes de informação foram registrados, é referida como o titular dos dados. Um coletor de dados coleta e armazena dados, consequentemente incorre em custos. No lado da demanda, o processador de dados usa os dados, os agrega e analisa. Na prática, o coletor e o processador de dados poderiam ser a mesma organização.

A ECONOMIA DE DADOS

Cada vez mais atividades econômicas ocorrem na cadeia de valor de dados. Essas atividades e as cadeias de suprimentos conectadas têm, assim, se tornado partes mais importantes da economia geral. Para monitorar o impacto da economia dos dados, a Comissão Europeia utiliza a seguinte definição:

A economia de dados mede os impactos globais do mercado de dados na economia como um todo. Envolve a geração, coleta, armazenamento, processamento, distribuição, elaboração de análises, entrega e exploração de dados possibilitados por tecnologias digitais. A economia de dados também inclui os efeitos diretos, indiretos e induzidos do mercado de dados sobre a economia. (Comissão Europeia, 2017, p. 2)

Usando esta definição, a dimensão da economia de dados em 2019 foi estimada em 2,6% do produto interno bruto (PIB) da União Europeia (325 bilhões de euros, excluindo o Reino Unido). Além disso, a economia de dados está se expandindo rapidamente. Num cenário conservador, prevê-se que a economia de dados cresça para 430 bilhões de euros em 2025 (3,3% do PIB), enquanto na perspectiva mais agressiva a sua dimensão deverá atingir 827 bilhões de euros até 2025 (5,9% do PIB) (Comissão Europeia, 2020). Em um recente esforço complementar para definir o tamanho da economia digital, a Organização para a Cooperação e o Desenvolvimento Econômico (OCDE) destaca que “continua havendo alguma subjetividade ou ‘nebulosidade’ em transformar definições em números” (OCDE, 2020). Assim, os números absolutos dessas estimativas dependem do modo como a definição é traduzida na prática; portanto, são um tanto arbitrários.

Cada vez mais atividades econômicas ocorrem na cadeia de valor de dados. Essas atividades e as cadeias de suprimentos conectadas têm, assim, se tornado partes mais importantes da economia geral.

Referências

- Buchanan, J. M. (1965). An economic theory of clubs. *Economica*, 32(125), 1-14.
- Carrière-Swallow, M. Y., & Haksar, M. V. (2019). The economics and implications of data: An integrated perspective. *International Monetary Fund*.
- Crémer, J., Montjoye, Y. A., & Schweitzer, H. (2019). Competition policy for the digital era. *Report for the European Commission*.
- Comissão Europeia. (2017). *Communication from the commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Building a European data economy*. <https://digital-strategy.ec.europa.eu/en/library/communication-building-european-data-economy>
- Comissão Europeia. (2020). *The European data market monitoring tool: Key facts & figures, first policy conclusions, data landscape and quantified stories: d2.9 final study report*. Publications Office. <https://data.europa.eu/doi/10.2759/72084>
- Coyle, D., Diepeveen, S., Wdowin, J., Kay, L., & Tennison, J. (2020). *The value of data – policy implications report*. Bennet Institute for Public Policy, Cambridge and Open Data Institute.
- Goldfarb, A., & Tucker, C. (2019). Digital Economics. *Journal of Economic Literature*, 57(1), 3-43.
- Harford, T. (2017). *How the world's first accountants counted on cuneiform*. BBC World Service, 50 Things That Made the Modern Economy. <https://www.bbc.com/news/business-39870485>
- Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science*, 332(0), 6025, 60-65.
- Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review*, 110(9), 2819-2858.
- Junta Comercial Nacional da Suécia. (2015). *No transfer, no production – a report on cross-border data transfers, global value chains, and the production of goods*. <https://ec.europa.eu/futurium/en/system/files/ged/publ-no-transfer-no-production.pdf>
- Organização para a Cooperação e o Desenvolvimento Econômico. (2020). *A roadmap toward a common framework for measuring the digital economy*. OECD publishing.
- Posner, E. A., & Weyl, E. G. (2018). *Radical markets – uprooting capitalism and democracy for a just society*. Princeton University Press.
- Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5), S71-S102.
- Statistics Canada. (2019). Measuring investment in data, databases and data science: Conceptual framework. *Series: Latest Developments in the Canadian Economic Accounts*. <https://www150.statcan.gc.ca/n1/en/catalogue/13-605-X201900100008>.
- Wdowin, J., & Diepeveen, S. (2020). *The value of data – accompanying literature review*. Bennet Institute for Public Policy, Cambridge and Open Data Institute.

Entrevista II

Os dados e a gratuidade da Internet

Nesta entrevista, Michael Kende, especialista em Políticas da Internet e presidente do Conselho da Datasphere Initiative, discute alguns dos temas debatidos em seu livro *O outro lado da gratuidade: entendendo a economia da Internet* (*The flip side of free: Understanding the economics of the Internet*) (The MIT Press, 2021), entre eles: os motivos da “gratuidade” da Internet e as implicações econômicas decorrentes disso, os riscos para os usuários de Internet no contexto da utilização e da reutilização de dados pessoais, as políticas para amenizá-los e as agendas de pesquisa para compreensão da economia da Internet.

Panorama Setorial da Internet (P.S.I.) *Em seu livro **The flip side of free: Understanding the economics of the Internet**, algumas questões centrais em sua análise se referem a por que a Internet é gratuita e quais as implicações econômicas decorrentes disso. Por que essas questões são tão importantes hoje em dia?*

Michael Kende (M.K.) O livro destaca três maneiras pelas quais a Internet é gratuita. Primeiro, muitos padrões e *software* subjacentes à Internet são abertos e estão disponíveis gratuitamente para os desenvolvedores. Em segundo lugar, cada vez mais os pacotes de banda larga são ilimitados, o que significa que o aumento do uso é gratuito após o pagamento da tarifa mensal e, muitas vezes, o Wi-Fi público pode ser usado de forma totalmente gratuita. E, por fim, diversos serviços de Internet estão disponíveis gratuitamente.

Isso é único em muitos aspectos, quando comparado aos serviços mais tradicionais, e tem impulsionado muitos dos benefícios econômicos e sociais da Internet desfrutados pelos usuários atualmente – os quais talvez tenham sido mais evidentes durante a pandemia, quando nossa dependência da Internet atingiu o auge. No entanto, como indica o título do livro, há um outro lado que precisa ser abordado.

Padrões abertos e *software* de código aberto podem não receber atenção suficiente por parte de desenvolvedores, devido à falta de retorno financeiro. Como resultado, por exemplo, há um caso famoso em que a biblioteca de *software* OpenSSL, usada para proteger transações, era vulnerável ao que ficou conhecido como o *bug* Heartbleed. Foram necessários US\$ 500 milhões para corrigir a falha. Embora seja verdade que *software* comerciais também são vulneráveis a *bugs*, tornou-se evidente que havia muito poucos recursos disponíveis para desenvolver o *software* OpenSSL. À medida que nossa dependência *online* continua a aumentar, o mesmo deve acontecer com os recursos usados para minimizar as vulnerabilidades que podem ser endereçadas.



Foto: Manon Volland

Michael Kende

Especialista em Políticas da Internet e presidente do Conselho da Datasphere Initiative.

“Mesmo se soubéssemos todos os dados que disponibilizamos para provedores de serviços e lêssemos atentamente os termos de serviço, não poderíamos ter pleno conhecimento de como ou por quem nossos dados têm sido usados.”

A utilização de pacotes de dados ilimitados tem impulsionado uma variedade cada vez maior de usos voltados para trabalho, saúde, educação e entretenimento. No entanto, recentemente uma desvantagem foi revelada, visto que os grandes provedores de banda larga, na Europa e em outros países, têm reclamado para os formuladores de políticas sobre o volume de tráfego transportado para os usuários finais e, em especial, sobre o tráfego de vídeo. Os provedores relutam em aumentar as taxas cobradas ou impor tarifas com base no uso de dados: em vez disso, reivindicam que os maiores provedores de conteúdo deveriam pagar de “forma proporcional” os custos que atribuem ao tráfego. Isso ameaçaria fundamentalmente os acordos de interconexão voluntária usados hoje para criar a “rede de redes” que constitui a Internet.

Por fim, os serviços gratuitos oferecem benefícios significativos aos usuários, como comunicações, mídias sociais e entretenimento. “Grátis” é um preço especial, que oferece um valor significativo para usuários e desenvolvedores que não precisam considerar o custo em sua decisão de adotar serviços e padrões. Entretanto, o “outro lado” deve ser abordado de maneira que não altere fundamentalmente a natureza da Internet.

P.S.I._ Uma das conclusões do seu livro é que muitos serviços de Internet estão disponíveis para nós sem nenhum custo, em troca de dados pessoais, que podem ser usados e reutilizados. Quais os principais riscos para os usuários de Internet nesse cenário? Que políticas poderiam efetivamente lidar com esses riscos, preservando as vantagens dos serviços de Internet para usuários?

M.K._ Os serviços gratuitos são pagos por meio de publicidade direcionada com base nos dados pessoais de usuários. Por sua vez, esses dados pessoais geram pontos de atenção sobre a privacidade e estão sujeitos ao risco de exposição por meio de violações de dados. O resultado é um déficit de confiança digital, a qual deve ser considerada para potencializar os benefícios do impacto da Internet e dos novos serviços que continuam a surgir. Não existe uma analogia perfeita para os dados, pois eles possuem propriedades únicas. Uma analogia comum para dados é o petróleo; embora seja verdade que os dados estão impulsionando a revolução digital da mesma forma que o petróleo impulsionou a Revolução Industrial, as semelhanças terminam por aí. O petróleo não é renovável, seu uso inevitavelmente gera efeitos climáticos negativos e é algo tangível. Por outro lado, os dados são generativos, podendo ser combinados e utilizados simultaneamente, e um uso cuidadoso não precisa resultar em efeitos negativos. Além disso, os dados são virtuais e intangíveis.

Essas propriedades geram riscos para os usuários de Internet. Mesmo se soubéssemos todos os dados que disponibilizamos para provedores de serviços e lêssemos atentamente os termos de serviço, não poderíamos ter pleno conhecimento de como ou por quem nossos dados têm sido usados.

Isso exige regulamentações de privacidade que restrinjam razoavelmente o uso de dados e coloquem mais controle nas mãos dos usuários, bem como termos de serviço dos provedores de serviços mais compreensíveis e que ofereçam mais controles aos usuários. Nenhuma parte interessada pode enfrentar sozinha as questões de privacidade.

Também existem desafios significativos em relação à cibersegurança, que resultam no uso não autorizado e ilegal de nossos dados pessoais. Embora algumas violações de dados sejam inevitáveis devido a vulnerabilidades desconhecidas ou ações internas, muitas poderiam ser evitadas. A questão que se coloca aqui é: por que as empresas não conseguem proteger adequadamente os dados de seus próprios usuários? Em termos econômicos, existem duas razões para isso. Em primeiro lugar, é difícil as empresas demonstrarem que implementaram medidas de cibersegurança rigorosas, por isso há pouco incentivo para investir na proteção de dados. Segundo, as empresas têm responsabilidade limitada caso ocorra uma violação de dados, o que significa que há pouca desvantagem em não investir nessa área. A situação é análoga à segurança automotiva na década de 1960, quando os carros podiam ser vendidos com um mínimo de proteção. Desde então, os governos começaram a impor alguns regulamentos, como proteção contra colisões, e os carros passaram a ser testados e classificados quanto à segurança, o que gerou uma demanda por níveis de segurança além do mínimo necessário. O mesmo processo deve ocorrer em relação à cibersegurança, com a introdução de novas regulamentações e normas que criem conscientização e demanda por maior proteção e responsabilidade das empresas que não atendam aos padrões estabelecidos.

P.S.I._ Como o debate proposto em seu livro avançou desde sua publicação? Em sua opinião, quais são os principais temas que devem ser investigados para entender melhor a economia da Internet?

M.K._ Questões relacionadas à privacidade e à proteção de dados pessoais têm afetado os níveis de confiança digital, as quais mereceriam mais investigações. Em particular, existem paradoxos, uma vez que os usuários, cada vez mais cientes das questões de privacidade, ainda fornecem dados pessoais *online* e, mesmo após a uma violação de confiança se tornar conhecida, não reduzem o uso dos serviços existentes. Esse fato pode ser mais bem entendido pelas lentes da economia comportamental, que busca analisar ações que não parecem “racionais” no sentido econômico, mas são comuns.

Esses paradoxos podem surgir por uma série de razões. Em primeiro lugar, pode haver uma falta de conscientização em relação aos verdadeiros riscos de privacidade dos dados pessoais, o que pode resultar em um aparente uso excessivo de serviços relevantes. Em segundo lugar, mesmo quando há essa conscientização, os usuários podem superestimar sua própria capacidade de evitar problemas ou subestimar a probabilidade de que eles ocorram. Outra questão comum na economia comportamental é a preferência

“Em particular, existem paradoxos, uma vez que os usuários, cada vez mais cientes das questões de privacidade, ainda fornecem dados pessoais *online* e, mesmo após a uma violação de confiança se tornar conhecida, não reduzem o uso dos serviços existentes.”

“Compreender a sensibilização dos usuários em relação às preocupações com a privacidade e a proteção de dados, e a maneira como esses fatores influenciam o comportamento, ajudará a abordar essas questões e aumentar o nível geral de confiança.”

pelos benefícios imediatos dos serviços de Internet em detrimento de quaisquer custos futuros, mesmo quando os riscos são bem compreendidos.

Há também outra questão relacionada à circunstância. Parece improvável que os usuários desistam dos serviços existentes, mesmo que tenham enfrentado uma quebra significativa de confiança. Por exemplo, o *website* Ashley Madison – criado com o objetivo explícito de permitir o adultério – foi alvo de um ataque: os dados de todos os seus 36 milhões de usuários foram expostos, causando prejuízos pessoais significativos para eles. No entanto, o *website* não apenas sobreviveu, como também afirma ter agora o dobro de usuários em comparação com antes do incidente. Se muito, a publicidade em torno da violação parece ter gerado demanda por seus serviços.

Portanto, os usuários relutam em desistir dos serviços existentes. Por outro lado, novos serviços – que ainda não oferecem benefícios – podem ser facilmente deixados de lado em prol da privacidade. Os aplicativos de rastreamento de contatos durante a pandemia COVID-19 – cujo uso generalizado poderia ter limitado a propagação do vírus – não alcançou mais de 30% da população na maioria dos países. Um dos motivos foi a preocupação com a privacidade, já que os aplicativos rastreavam a localização dos usuários para determinar se eles haviam sido expostos a alguém que tivesse testado positivo. Embora os aplicativos tendessem a preservar a privacidade, mantendo muito menos dados do que muitos outros *websites* com indiscutivelmente menos benefícios, eles não conseguiram superar a relutância.

Consequentemente, a adoção e o uso de serviços *online* podem não ser racionais, mas têm impactos significativos. Compreender a sensibilização dos usuários em relação às preocupações com a privacidade e a proteção de dados, e a maneira como esses fatores influenciam o comportamento, ajudará a abordar essas questões e aumentar o nível geral de confiança.

Relatório de Domínios

A dinâmica dos registros de domínios no Brasil e no mundo

O Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), departamento do Núcleo de Informação e Coordenação do Ponto BR (NIC.br), monitora mensalmente o número de nomes de domínios de topo de código de país (*country code Top-Level Domain* [ccTLD]) registrados entre os países que compõem a Organização para a Cooperação e Desenvolvimento Econômico (OCDE) e o G20⁸. Considerados os membros de ambos os blocos, as 20 nações com maior atividade somam mais 90,76 milhões de registros. Em agosto de 2023, os domínios registrados sob .de (Alemanha) chegaram a 17,59 milhões. Em seguida, aparecem Reino Unido (.uk), China (.cn) e Países Baixos (.nl), com, respectivamente, 9,51 milhões, 7,37 milhões e 6,33 milhões de registros. O Brasil teve 5,22 milhões de registros sob .br, ocupando a quinta posição na lista, como mostra a Tabela 1⁹.

⁸ Grupo composto pelas 19 maiores economias mundiais e a União Europeia. Saiba mais: <https://g20.org/>

⁹ A tabela apresenta a contagem de domínios ccTLD segundo as fontes indicadas. Os valores correspondem ao registro publicado por cada país, tomando como base os membros da OCDE e do G20. Para países que não disponibilizam uma estatística oficial fornecida pela autoridade de registro de nomes de domínios, a contagem foi obtida em: <https://research.domaintools.com/statistics/tld-counts>. É importante destacar que há variação no período de referência, embora seja sempre o mais atualizado para cada localidade. A análise comparativa de desempenho de nomes de domínios deve considerar ainda os diferentes modelos de gestão de registros ccTLD. Assim, ao observar o ranking, é preciso atentar para a diversidade de modelos de negócio existentes.

/Panorama Setorial da Internet

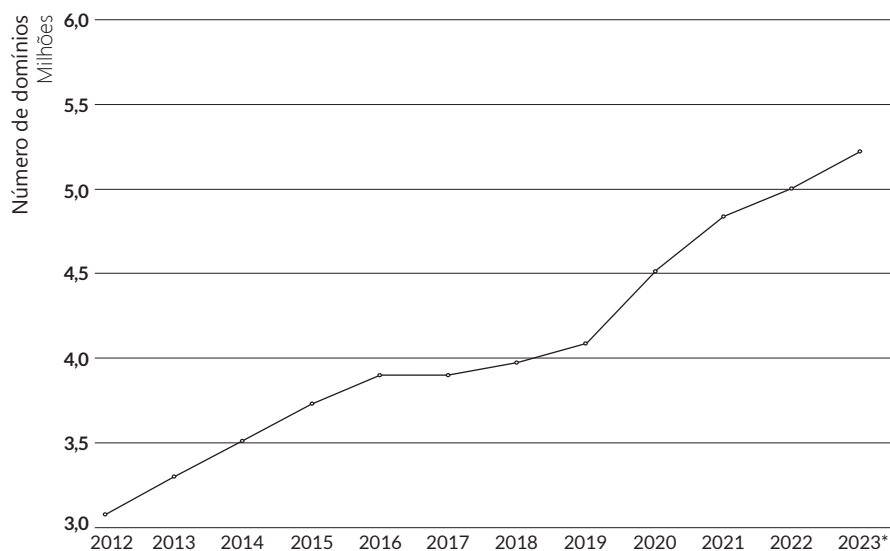
Tabela 1 – TOTAL DE REGISTROS DE NOMES DE DOMÍNIOS ENTRE OS PAÍSES DA OCDE E DO G20

Posição	País	Número de domínios	Data de referência	Fonte (website)
1	Alemanha (.de)	17.598.905	01/09/2023	https://www.denic.de
2	Reino Unido (.uk)	9.511.231	31/07/2023	https://www.nominet.uk/news/reports-statistics/uk-register-statistics-2023
3	China (.cn)	7.370.599	01/09/2023	https://research.domaintools.com/statistics/tld-counts
4	Países Baixos (.nl)	6.333.731	01/09/2023	https://stats.sidnlabs.nl/en/registration.html
5	Brasil (.br)	5.223.034	31/08/2023	https://registro.br/dominio/estatisticas
6	Rússia (.ru)	5.073.407	01/09/2023	https://cctld.ru
7	Austrália (.au)	4.263.841	01/09/2023	https://www.auda.org.au
8	França (.fr)	4.081.492	31/08/2023	https://www.afnic.fr/en/observatory-and-resources/statistics
9	União Europeia (.eu)	3.679.990	01/09/2023	https://research.domaintools.com/statistics/tld-counts
10	Itália (.it)	3.482.097	01/09/2023	http://nic.it
11	Colômbia (.co)	3.465.280	01/09/2023	https://research.domaintools.com/statistics/tld-counts
12	Canadá (.ca)	3.364.108	01/09/2023	https://www.cira.ca
13	Índia (.in)	2.954.328	01/09/2023	https://research.domaintools.com/statistics/tld-counts
14	Suíça (.ch)	2.554.894	15/08/2023	https://www.nic.ch/statistics/domains
15	Polônia (.pl)	2.525.714	01/09/2023	https://www.dns.pl/en
16	Espanha (.es)	2.069.267	09/08/2023	https://www.dominios.es/dominios/en
17	Estados Unidos da América (.us)	1.986.914	01/09/2023	https://research.domaintools.com/statistics/tld-counts
18	Japão (.jp)	1.748.824	01/09/2023	https://jprs.co.jp/en/stat
19	Bélgica (.be)	1.739.915	01/09/2023	https://www.dnsbelgium.be/en
20	Portugal (.pt)	1.735.197	01/09/2023	https://www.dns.pt/en/statistics

Data de coleta: 01 de setembro de 2023.

O Gráfico 1 apresenta o desempenho do .br desde o ano de 2012.

Gráfico 1 – TOTAL DE REGISTROS DE DOMÍNIOS DO .BR – 2012 a 2023*



*Data de coleta: 31 de agosto de 2023.

Fonte: Registro.br

Recuperado de: <https://registro.br/dominio/estatisticas/>

Em agosto de 2023, os cinco principais domínios genéricos (*generic Top-Level Domain* [gTLD]) totalizaram mais de 189,77 milhões de registros. Com 159,14 milhões de registros, destaca-se o .com, conforme apontado na Tabela 2.

Tabela 2 – TOTAL DE REGISTROS DE DOMÍNIOS DOS PRINCIPAIS gTLD

Posição	gTLD	Número de domínios
1	.com	159.149.727
2	.net	12.796.114
3	.org	10.780.717
4	.info	3.753.027
5	.xyz	3.293.663

Data de coleta: 01 de setembro de 2023.

Fonte: DomainTools.com

Recuperado de: research.domaintools.com/statistics/tld-counts

Marcadores da Internet no Brasil

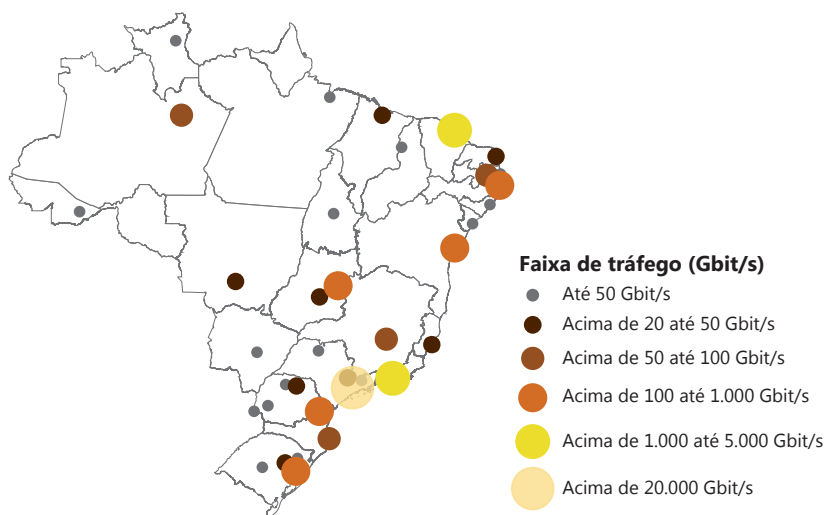
IX.br: dados sobre os Pontos de Troca de Tráfego

O IX.br (Brasil Internet Exchange) é uma iniciativa do Núcleo de Informação e Coordenação do Ponto BR (NIC.br), apoiada pelo Comitê Gestor da Internet no Brasil (CGI.br), que promove e implementa os Pontos de Troca de Tráfego (PTT – *Internet Exchange Points [IXP]*), infraestrutura necessária para a interconexão direta entre as redes, também conhecidas como Sistemas Autônomos (*Autonomous Systems [AS]*), que compõem a Internet no Brasil.

A interligação de diversos AS em um IXP simplifica o trânsito da Internet, estabelecendo um tráfego mais direto até um determinado destino. Isso melhora a qualidade, reduz custos e aumenta a resiliência da rede.

Atualmente, a iniciativa abrange 36 PTT independentes, distribuídos em todo o Brasil (Figura 1), e é um dos mais importantes conjuntos de PTT do mundo. O Gráfico 1 revela o crescimento contínuo do tráfego para o conjunto de PTT que compõe o IX.br nos últimos cinco anos.

Figura 1 – PONTOS DE TROCA DE TRÁFEGO (PTT) NO BRASIL, POR FAIXA DE TRÁFEGO

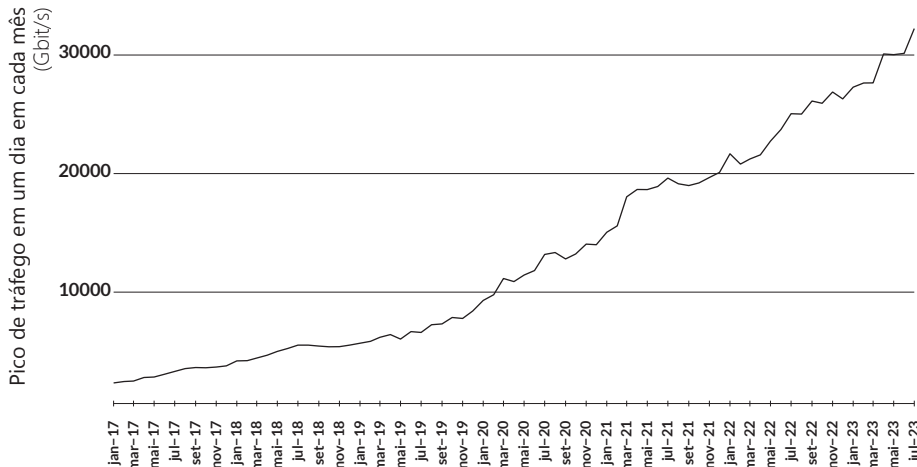


Período de referência: julho de 2023.

Fonte: IX.br | NIC.br

Recuperado de: <https://ix.br/trafego/agregado/>

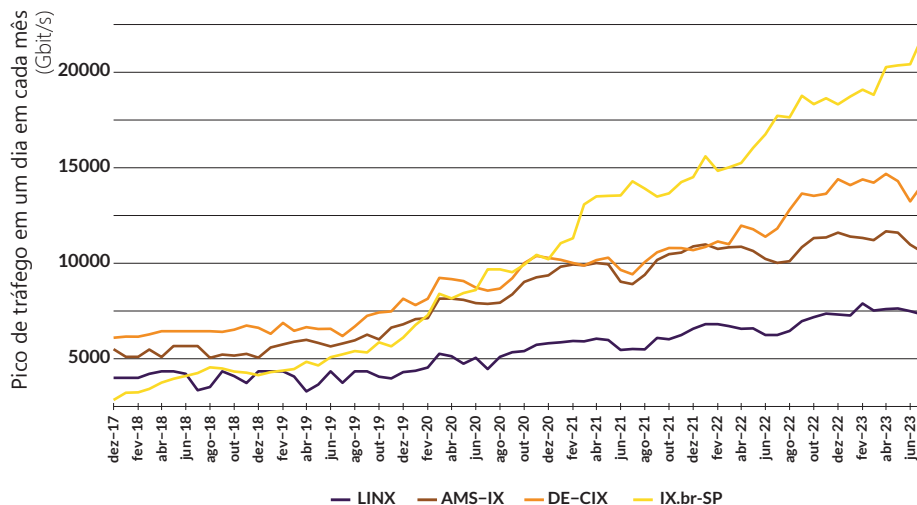
Gráfico 1 – PICO DE TRÁFEGO PARA O CONJUNTO DE PTT DO IX.br – 2017 a 2023



Fonte: IX.br|NIC.br
 Recuperado de: <https://ix.br/agregado/>

O Gráfico 2 compara o pico do tráfego do PPT de São Paulo, o maior do Brasil, com os três maiores da Europa: LINX (Londres, Inglaterra), AMS-IX (Amsterdã, Países Baixos) e DE-CIX (Frankfurt, Alemanha), entre 2017 e 2023.

Gráfico 2 – PTT DE LONDRES (LINX), AMSTERDÃ (AMS-IX), FRANKFURT (DE-CIX) E SÃO PAULO (IX.br-SP), POR PICO DE TRÁFEGO – 2017 a 2023



Fonte: IX.br|NIC.br
 Recuperado de: <https://www.de-cix.net/en/locations/frankfurt/statistics>;
<https://www.ams-ix.net/ams/documentation/total-stats>;
<https://portal.linx.net/services/lans-snmp>; <https://ix.br/trafego/agregado/>

Aqui você encontra mais informações sobre a atuação e estatísticas do IX.br.

/Tire suas dúvidas



Atividades online

e as migalhas digitais

Os dados são um ativo fundamental no contexto da economia de dados, contribuindo para os processos de tomada de decisão das organizações, como também servem de base para novos modelos de negócio.

As atividades e os serviços realizados online deixam uma série de “migalhas” digitais ou de dados¹⁰. Assim, a presença crescente de

plataformas digitais no cotidiano de milhões de pessoas resulta na produção de uma enorme quantidade de dados.

Em 2022, 149 milhões (81%)¹¹ de pessoas no Brasil com 10 anos ou mais eram usuárias de Internet¹². Os dados a seguir retratam algumas atividades e serviços¹³ realizados na Internet por essa população:

Atividades



93% mandou mensagens instantâneas.



80% usou redes sociais.



80% assistiu a vídeos, programas, filmes ou séries pela Internet.



69% compartilhou conteúdo na Internet.

Serviços

40% pediu táxis ou motoristas em aplicativos como, por exemplo, Uber ou 99.



38% pagou por serviços de filmes ou séries pela Internet como, por exemplo, Netflix ou Globoplay.



33% fez pedidos de refeições em sites ou aplicativos como, por exemplo, iFood ou Rappi.



¹⁰ Letouzé, E. (2018). *Big Data e desenvolvimento: uma visão geral*. <https://cetic.br/pt/publicacao/ano-x-n-1-big-data-para-o-desenvolvimento>

¹¹ Dados da pesquisa TIC Domicílios 2022, do Cetic.br|NIC.br. Disponível em: <https://cetic.br/pt/pesquisa/domicilios/>

¹² Considera-se “usuário” aquele que utilizou a Internet há menos de três meses em relação ao momento da entrevista, conforme definição da União Internacional de Telecomunicações (UIT).

¹³ Outras atividades e serviços realizados na Internet por usuários de Internet, coletadas pela pesquisa TIC Domicílios 2022, podem ser encontradas em: <https://cetic.br/pt/tics/domicilios/2022/individuos/>

/Créditos

REDAÇÃO

RELATÓRIO DE DOMÍNIOS

Thiago Meireles (Cetic.br | NIC.br)

MARCADORES DA INTERNET NO BRASIL

Julio Sirota (IX.br | NIC.br) e Milton Kaoru Kashiwakura (NIC.br)

PROJETO GRÁFICO E INFOGRAFIA

Giuliano Galves, Larissa Paschoal e Maricy Rabelo (Comunicação | NIC.br)

DIAGRAMAÇÃO

Grappa Marketing Editorial

EDIÇÃO DE TEXTO EM PORTUGUÊS

Érica Santos Soares de Freitas

TRADUÇÃO INGLÊS-PORTUGUÊS

Ana Zuleika Pinheiro Machado

COORDENAÇÃO EDITORIAL

Alexandre F. Barbosa, Graziela Castello, Javiera F. M. Macaya e Mariana Galhardo Oliveira (Cetic.br | NIC.br)

AGRADECIMENTOS

Marielza Oliveira (UNESCO)

Ramy El-Dardiry, Milena Dinkova e Bastiaan Overvest (Gabinete de Análise de Política Econômica dos Países Baixos)

Marcos Dantas (UFRJ e CGI.br)

Michael Kende (The Datasphere Initiative)

SOBRE O CETIC.br

O Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação – Cetic.br (<https://www.cetic.br/>), departamento do NIC.br, é responsável pela produção de estudos e estatísticas sobre o acesso e o uso da Internet no Brasil, divulgando análises e informações periódicas sobre o desenvolvimento da rede no país. O Cetic.br atua sob os auspícios da UNESCO.

SOBRE O NIC.br

O Núcleo de Informação e Coordenação do Ponto BR – NIC.br (<https://nic.br/>) é uma entidade civil de direito privado e sem fins de lucro, encarregada da operação do domínio .br, bem como da distribuição de números IP e do registro de Sistemas Autônomos no país. Conduz ações e projetos que trazem benefícios à infraestrutura da Internet no Brasil.

SOBRE O CGI.br

O Comitê Gestor da Internet no Brasil – CGI.br (<https://cgi.br/>), responsável por estabelecer diretrizes estratégicas relacionadas ao uso e desenvolvimento da Internet no Brasil, coordena e integra todas as iniciativas de serviços Internet no país, promovendo a qualidade técnica, a inovação e a disseminação dos serviços ofertados.

*As ideias e opiniões expressas nos textos dessa publicação são as dos respectivos autores e não refletem necessariamente as do NIC.br e do CGI.br.



unesco

Centro
sob os auspícios
da UNESCO

cetic.br

Centro Regional
de Estudos para o
Desenvolvimento
da Sociedade
da Informação

nic.br

Núcleo de Informação
e Coordenação do
Ponto BR

cgi.br

Comitê Gestor da
Internet no Brasil

CREATIVE COMMONS

Atribuição
Não Comercial
(by-nc)



ISSN - 2965-2642



POR UMA INTERNET CADA VEZ MELHOR NO BRASIL

CGI.BR, MODELO DE GOVERNANÇA MULTISSETORIAL

<https://cgi.br>

nic.br cgi.br